Essays in Finance and Econometrics

Essays in Finance and Econometrics

A dissertation presented

by

William Diamond

to

The Committee for the PhD in Business Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

June 2017

Dissertation Advisor:
Professor David S. Scharfstein

Author:
William Diamond

Essays in Finance and Econometrics

Abstract

This thesis presents two essays studying the role of banks in financial markets and one which studies statistical inference in matching markets. The first chapter presents a new theory of the role of banks, providing an explanation for the role of publicly available securities on bank balance sheets. The model provides a unified framework for studying asset prices, portfolio choices, capital structure, and macroeconomic policies such as quantitative easing. Relative to existing models of banking, the paper emphasizes the demand for deposits rather than the expertise of bankers in making loans. The second chapter expands on the research agenda presented in the first by studying why traders might demand bank deposits. It formalizes the idea that deposits function as a form of money, because they are safe assets that avoid adverse selection problems in trade. The model presents a fundamental tension between banks creating large quantities of money-like assets and being vulnerable to financial panics. The third chapter studies identification and estimation in two sided matching markets where the desirability of matching with an agent can be summarized by a latent index. The paper first studies identification, showing that a many-to-one matching market allows for the estimation of parameters that cannot be estimated in a one-to-one matching market. It then studies the limiting distribution of a class of estimators and develops novel methods for proving such limit theorems.

Contents

List of Figures

Acknowledgments

The writing of this thesis relied on outstanding help and support from friends, colleagues, and family members. My dissertation committee, David Scharfstein, Jeremy Stein, Sam Hanson, and Adi Sunderam, put an incredible amount of effort into teaching me how to do research and communicate effectively. My coauthor and friend Nikhil Agarwal taught me many things, such as mathematical tools, good work habits, and how to maintain a healthy outlook on research and life. Finally, my parents and sister (Douglas, Elizabeth, and Rebecca Diamond) were incredibly supportive emotionally and intellectually throughout the process of writing my thesis in both the best and worst of times. This thesis could not have been written alone, and there are many debts I will try my best to repay over the years.

Introduction

The first chapter develops a theory of financial intermediation in public securities markets. Riskless securities earn a convenience yield, and all firms face agency costs of equity financing. Intermediaries endogenously emerge to buy a low risk, diversified portfolio of debt securities, allowing intermediaries to issue many riskless deposits and little equity. The model explains the credit spread puzzle in bonds and low risk anomaly in stocks, why intermediary leverage is high and corporate leverage is low, why intermediaries own debt and households own equity, how safe asset demand fueled the subprime boom, and how quantitative easing effects output and financial stability.

The second chapter presents a model of how banks are structured to create deposits that function as a form of money and the resulting exposure of banks to runs. The paper begins with a model of anonymous, bilateral transactions where agents who sell goods are unable to verify the quality of any good provided to pay for the goods sold. As a result, goods whose quality is most easily verified naturally circulate as a form of money. In particular, if there is one good "gold" whose quality is commonly known and others that can be costlessly counterfitted, all transactions use gold as a medium of exchange. The scarcity of gold induces agents to inefficiently overproduce it relative to other goods that can be counterfitted. By creating deposits that can be exchanged on demand for gold, which are only partially backed by gold reserves, banks reduce the scarcity of gold. The resulting inflation incentivizes a socially beneficial substitution towards producing non-gold goods. Due to this fractional reserve banking, banks are endogenously exposed to the risk of runs. The model provides a unified framework in which the demand for money naturally leads to

a run-prone banking system.

The third chapter studies a large class of two-sided matching models that include both transferable and non-transferable utility, resulting in positive assortative matching along a latent index. Data from matching markets, however, may not exhibit perfect assortativity due to the presence of unobserved characteristics. This paper studies the identification and estimation of such models. We show that the distribution of the latent index is not identified when data from one-to-one matches are observed. Remarkably, the model is non-parametrically identified using data in a single large market when each agent on one side has at least two matched partners. The additional empirical content in many-to-one matches can be illustrated using simulations and stylized examples. We then derive asymptotic properties of a minimum distance estimator as the size of the market increases, allowing estimation using dependent data from a single large matching market. The nature of the dependence requires modification of existing empirical process techniques to obtain a limit theorem.

Chapter 1

Safety Transformation and the Structure of the Financial System

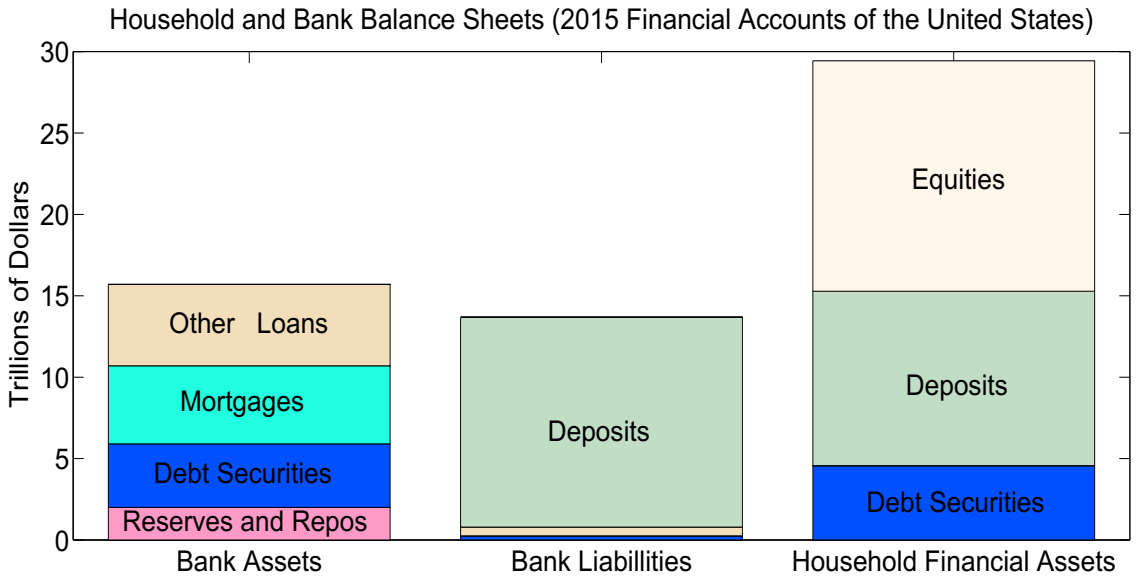

Figure 1.1: *Household and Bank Balance Sheets.*

An important role of financial intermediaries is to issue safe, money-like assets, such as bank deposits and money market fund shares. As an empirical literature has documented Krishnamurthy and Vissing-Jorgensen (2012); Nagel (2016); Sunderam (2015), these assets have a low rate of return, strictly below the risk-free rate they would earn without providing monetary services. Agents who can issue these assets therefore raise financing on attractive terms, capturing the "demand for safe assets" that pushes their cost of borrowing below that of others. As shown in Gorton and Pennacchi (1990), any firm that can issue riskless securities meets the demand for safe, money-like assets. This raises the question of why financial intermediaries almost uniquely can issue such assets.

The assets owned by money-creating financial institutions are primarily loans and debt securities issued by firms, households, and governments. Of the $17.3 trillion of assets owned by depository institutions in the USA in 2015, $4.8 trillion were mortgages, $3.9 were debt securities including $2.1 trillion of agency and GSE backed securities, $5.0 trillion were non-mortgage loans to firms and households, and $2.0 trillion were reserves, while only $100 billion were equities which are held primarily by households. While money creation

in the "shadow banking" system is harder to measure, money market funds, securitization vehicles, and broker dealers that play a role here also invest significantly in debt.[1] The role of publicly traded debt and readily securitized mortgages in the asset portfolios of banks and shadow banks is not consistent with many existing models that imply intermediaries hold special assets that are unavailable to other investors.[2]

This paper develops a general equilibrium model in which financial intermediaries emerge endogenously, buying a portfolio of publicly available debt securities to most effectively create safe, money-like assets. The model explains (i) why money-creating financial intermediaries invest in debt while households invest in equity, (ii) why intermediaries are highly levered while non-financial firms are not, and (iii) why risk is priced more expensively in the debt market than the equity market, consistent with the "credit spread puzzle" in bonds and "low risk anomaly" in stocks. In addition to its implications for the structure of the financial system, the model provides a framework for understanding the general equilibrium effect of changes in the supply and demand for safe assets. An increased demand for safe assets replicates many features of the subprime boom, with intermediaries expanding and taking more risk while the non-financial sector increases its leverage. Quantitative easing policies increase the supply of safe assets, decrease the price of risk in debt markets, reduce intermediary risk taking, and increase output at the zero lower bound.

Two basic ingredients are at the core of the model. First, households obtain utility directly from holding riskless assets, which captures the demand for money-like assets without modelling the frictions that make money essential Stein (2012b). The idea that only safe assets function as money goes back at least to Gorton and Pennacchi (1990), who show

[1] Another financial institution that can be said to issue long duration safe assets is a life insurance company, since life insurance contracts promise fixed dollar values in the future. The portfolios in the general account of life insurers which back insurance contracts are also composed almost entirely of debt.

[2] Household portfolio holdings are based on the assumption that their mutual funds are 70% equity and 30% debt, consistent with data from the Investment Company Institute's Investment Company Fact Book. 37% of households' direct holdings of debt securities are municipal bonds where they face a tax advantage over other investors.

that risky assets are subject to a lemons problem when informed and uninformed agents trade. Second, all firms face an agency problem in financing risky investment. Each firm's management privately observes its output and reports this output to outside investors. If management underreports, it can divert some fraction of the difference between the true and reported output. This costly state falsification problem is due to Lacker and Weinburg (1989) and implies that riskier investments face more severe agency frictions. The optimal strategy of a financial intermediary is to choose a low risk portfolio that backs as many riskless assets as possible while minimizing the agency costs due to the risk in its asset portfolio. High risk assets that would cause too severe of an agency problem for the intermediary are bought by households instead.

The model provides a new theory of the connection between a bank's assets and liabilities that is consistent with the role of publicly available securities on bank balance sheets. Existing theories that explain both the assets and liabilities of financial intermediaries imply that bank assets are too illiquid to ever sell to outsiders. Diamond (1984); Diamond and Rajan (2001) argue that banks acquire information that makes their assets illiquid, while Dang *et al.* (2017) requires banks to conceal information so that their assets cannot trade at a market price.[3] In my framework, banks have the same investment opportunities and information as households and face the same frictions in raising outside financing as other firms. The key connection between the assets and liabilities of banks in this paper is that a bank's asset portfolio should be low risk in order to back many riskless deposits with a minimum of agency costs. This explanation for the role of intermediaries in public securities markets connects financial intermediation theory with a literature on the role of intermediaries in the pricing of public securities Krishnamurthy and He (2013); Adrian *et al.* (2014) that has had some empirical success. While banks own some assets unavailable to households, this paper bridges the gap between financial intermediation theory and the large holdings of publicly available securities on intermediary balance sheets by studying a

[3]The branch of this literature that assumes bankers monitor borrowers implies that public equities are too informationally sensitive to be sold, while empirically non-expert households have large holdings of equity.

framework in which all financial assets are publicly available.[4]

The liquidity of bank balance sheets has increased over time due to the development of securitization and syndication, suggesting that this paper is most relevant for understanding the modern financial system. Loutskina (2011); Loutskina and Strahan (2009) show a large secular increase in the liquidity of bank assets as they become easier to securitize and show that this mitigates their financial constraints. Barnish *et al.* (1997) argues that the rise of syndication has made the bank loan market more liquid. In addition, the role of securitized assets and other public securities in the shadow banking system seems to be particularly in tension with models that emphasize illiquid relationship lending. While existing literature DeMarzo and Duffie (1999); DeMarzo (2005) studies the degree to which informed originators are able to sell securitizations to outsiders, these models do not explain why the stakes sold to outsiders are bought primarily by levered financial institutions who may not have private information.

In the model, a continuum of projects with exogenous output (Lucas trees) provide all resources and must be managed by firms.[5] Firms choose whether to buy a single tree or act as a financial intermediary who can invest in securities. Each tree-owning non-financial firm sells securities whose payoffs must be increasing in its own cashflows and chooses to issue a low risk debt security and a high risk equity security.[6] These securities are exposed to both aggregate and tree-specific idiosyncratic risk, and this idiosyncratic risk ensures that non-financial debt cannot directly meet households' demand for riskless assets. This provides a role for intermediaries, who buy a diversified portfolio of non-financial debt which is safe enough to back a large quantity of riskless deposits with a small buffer of loss-bearing capital. Intermediaries do not buy riskier equities because the agency costs of doing so pushes their willingness to pay below that of households. As is true

[4]A natural extension is to study a model in which assets are publicly available but may still be illiquid.

[5]As noted later, the model can be interpreted to also include trees that represent houses, which households can use as collateral to borrow from banks.

[6]In practice, conglomerate firms such as Berkshire-Hathaway and General Electric do exist and are sometimes thought to play a role as financial intermediaries. A firm that could hold a diversified tree portfolio at a cost could also create safe assets in my model and compete with other intermediaries.

empirically, the balance sheet of an intermediary is composed of a pool of debt which it then tranches into a riskless deposit and risky equity. The fact that non-financial debt has low systematic risk allows the intermediary to be highly levered, consistent with Berg and Gider (forthcoming)'s empirical finding that the low asset risk of banks explains their high leverage.

The fact that intermediaries are willing to pay more than households for low systematic risk assets but less for high systematic risk assets implies that asset prices are segmented. The pricing kernel of assets owned by the intermediary features a low risk-free rate, since riskless assets can back deposits without any loss-bearing capital, but a high price of systematic risk, reflecting the intermediary's agency costs of holding a risky portfolio. As in models with leverage constraints Frazzini and Pedersen (2014); Black (1972), less systematic assets therefore earn a higher risk-adjusted return than more systematic assets. The intermediary's ability to raise deposit financing gives it a low borrowing cost, so it exploits this segmentation by holding a low risk portfolio on a highly levered balance sheet.

This endogenous market segmentation is arbitraged by non-financial firms when they choose their capital structure, resulting in segmentation between debt and equity markets. Each firm chooses its leverage so that its debt is sufficiently low risk to sell to intermediaries and its equity is sufficiently high risk to sell to households. The firm's total market value is therefore strictly higher than any agent would be willing to pay for all of the firm's cashflows. When each firm chooses its capital structure optimally, all debt is low enough risk to be priced by the intermediary's pricing kernel and all equity is high enough risk to be priced by the household's pricing kernel. Thus, the segmentation between low and high risk assets is endogenously segmentation between the debt and equity markets. This is consistent with the "credit spread puzzle" (Huang and Huang 2012) that structural credit models that infer credit spreads assuming the debt and equity markets are integrated tend to imply smaller spreads than empirically observed. It also explains the "low risk anomaly" (Black Jensen Scholes 1972, Baker Bradley Taliaferro 2014, Bansal Coleman 1996), which finds that the price of risk in the stock market is too low for simple measures of risk to be

consistent with the empirically observed high return on the stock index and low risk-free rate.

Because the model endogenously determines intermediary and household balance sheets, financial and non-financial capital structure, and segmented pricing of debt and equity securities, it provides a rich framework for studying the financial system's response to changes in the supply and demand for safe assets. I use it to study the effects of a growing demand for safe assets, which a macroeconomic literature (Bernanke et. al. 2011, Caballero Farhi 2017) argues is a feature of the global economy in recent decades, and to understanding the effects of the quantitative easing policies that involved purchasing publicly available bonds. The model implies that an increased demand for safe assets induces the financial system to expand and invest in riskier debt, decreasing the borrowing costs of the non-financial sector, and induces the non-financial sector to increase its leverage. This is consistent with the subprime boom of the 2000s.

The model is a natural framework for studying how quantitative easing policies impact intermediary risk taking and non-financial leverage decisions. The fact that intermediaries hold public securities in my model allows it to speak to the effects of government purchases of public securities.[7] By swapping intermediaries' risky assets for riskless assets, quantitative easing reduces intermediary risk taking, compresses risk premia in debt markets, increases the supply of safe assets, and stimulates aggregate demand at the zero lower bound. The model also can be used to understand the policy speech (Stein 2012b) which argues that the reduced borrowing costs caused by quantitative easing leads firms to issue debt that weakens its effects. Away from the zero lower bound, a rise in the natural rate due to quantitative easing can increase borrowing costs. At the zero lower bound, borrowing costs decrease, but the increase in consumption also boosts the price of equities owned by households, consistent with event studies (Neely 2011, Chodorow-Reich 2014). Firms may delever in response to quantitative easing, since the cost of equity financing decreases.

[7]There do exist models that simply assume assets purchased in quantitative easing can only be held by intermediaries. My model reconciles this literature with models where intermediaries appear endogenously.

1.1 Baseline Model

I summarize the model's agents, timing, and frictions. Next, I solve the portfolio choice problems of the representative household and intermediary in partial equilibrium, taking as given a set of securities available for purchase. I use these portfolio choice results to show that the market for low risk assets (which the intermediary buys) are segmented from the market for high risk assets (which the household buys). I then show how non-financial firms choose the securities they issue to take advantage of this segmented capital market. After characterizing the model's unique equilibrium, I use the model as a framework for showing how the financial system responds to changes in the supply and demand for safe assets and to quantitative easing policies.

Setup The model has two periods ($t = 1, 2$). Goods C_1 are available at time 1 which cannot be stored. Output at time 2 is produced by a continuum of trees indexed by $i \in [0, 1]$, where tree i produces f_i. At time 2, a binary aggregate shock is realized to be "good" or "bad" with probability $\frac{1}{2}$, and the output of the trees are conditionally independent given this aggregate shock. These aggregate and idiosyncratic shocks to each tree's output are the only sources of risk.

There are two classes of agents: households and firms. Households are endowed with wealth W_H which they invest in order to consume. The household maximizes its expected utility

$$u(c_1) + E[u(c_2)] + v(d). \tag{1.1}$$

which depends on its consumption (c_1, c_2) at times 1 and directly on its holding d of riskless assets that pay out at time 2. Households can invest in securities issued by firms, but trees must be held by firms.[8]

Firms can choose either to be an "intermediary" or a "non-financial firm." Each non-

[8] Allowing some trees to be held by households (representing houses rather than corporate assets) would allow the model to have homeowners getting mortgages from banks with little added complexity as explained later on.

financial firm can invest in one tree i and sell securities backed by the tree. Firms are not able to invest in diversified pools, motivated by the idea that conglomerate firms can be difficult to manage. Intermediaries cannot invest in trees but can invest in the same financial securities available to households and can issue securities backed by their portfolio. Unlike non-financial firms, intermediaries can hold a diversified portfolio. An intermediary can invest in a diversified portfolio like a household and issue securities like a firm, allowing it to issue riskless assets backed by a pool of securities, which other agents cannot do.

The output of firms is not verifiable and must be reported by its management to outside investors. Management can underreport output to divert resources. If a firm has payoffs δ_{firm} at period 2 and its management reports $\delta'_{firm} < \delta_{firm}$ in the support of the firm's output distribution, management can divert resources $C\left(\delta_{firm} - \delta'_{firm}\right)$, where $C'(0) = 0$, $C'' > 0$, and $\sup_e C'(e) < 1$. $C'(e) < 1$ implies that resources are destroyed when management diverts. The owners of the firm can provide the management with output-contingent compensation, and it is optimal to incentivize management not to divert. This agency problem is equivalent to the costly state falsification model of (Lacker Weinburg 1986). The problem makes it costly for a firm to own risky assets, since more asset risk increases the amount management can divert. This problem incentivizes the intermediary to choose a low risk portfolio, while it is an unavoidable cost for non-financial firms since the riskiness of each tree's output f_i is exogenous.[9]

Once management has reported the firm's output, the equityholders who control the firm can choose to either destroy output or raise additional funding.[10] Equityholders will destroy output if their residual claim is decreasing in the firms output and will raise additional funding if their residual claim increases more than one for one in the firm's output. Following (Innes 1990), each firm will choose to issue securities that are increasing in its own cashflows so equityholders will not manipulate the firm's output. In addition,

[9] At the end of time 2, households can transfer utility directly to management to buy the consumption goods paid to them, preserving the tractability of an endowment economy.

[10] If the firm's owners raise hidden funding, they do so at time 2 and also pay back the loan at time 2 so that the market interest rate is 0 consistent with (Innes 1990).

firms cannot issue securities whose payoffs depend on the uncontractible aggregate state or the output of other firms. Given these constraints, all firms optimally issue only debt and equity, so for simplicity the paper can be understood taking these securities as given and ignoring this second agency problem.

Financial securities are indexed by $s \in [0,1]$. Each security s has payoff δ_s at time 2 and is sold for a price p_s at time 1. These securities $s \in [0,1]$ are issued by the firms owning trees $i \in [0,1]$. To relate the indexing of trees and securities, let $s = \frac{i}{2}$ refer to the debt of the firm owning tree $i \in [0,1]$ and $s = \frac{1}{2} + \frac{i}{2}$ refer to that firm's equity. All assets can be purchased by either the household or the intermediary.[11]

In this model, securities cannot be broken into Arrow-Debreu claims or be sold short. The expected payment of each security is positive in both states of the world. The ratio $\frac{E_{good}\delta_s}{E_{bad}\delta_s}$ determines the exposure of security s to systematic risk, and agents can buy high or low systematic risk securities. However, it is impossible for an agent who wants only bad state payoffs to avoid buying good state claims as well. If agents were able to form long/short portfolios, they could go long assets for which $\frac{E_{good}\delta_s}{E_{bad}\delta_s}$ is low and short assets for which $\frac{E_{good}\delta_s}{E_{bad}\delta_s}$ is high to isolate bad state payoffs, so this is forbidden.

Household's problem The household faces a standard intertemporal consumption problem, except that it obtains utility directly from holding riskless assets. The household may either consume or invest in securities. Risky securities owned by the household are priced by the marginal utility of consumption they provide. The risk-free rate lies strictly below the rate implied by the household's consumption preferences, reflecting the extra utility benefit of holding riskless assets. An arbitrage trade which exploits this low risk-free rate is to buy a portfolio of assets and sell a riskless senior tranche and risky junior tranche backed by the portfolio, which is precisely the role played by intermediaries.

---

[11]The continuum law of large numbers is assumed to hold. A portfolio of $m(s)$ units of asset s pays $\int_0^1 [E_{bad}\delta_s]\, m(s)\, ds$ in the bad state and $\int_0^1 \left[E_{good}\delta_s\right] m(s)\, ds$ in the good state. A sufficient condition if $\|m\|_\infty < \infty$ as required by the resource constraint is $\sup_s \max\left(Var_{good}\delta_s, Var_{bad}\delta_s\right) < \infty$ which follows from $\sup_i \max\left(E_{bad}f_i^2, E_{good}f_i^2\right) < \infty$.

The household maximizes expected utility in expression 1.1 over period 1 consumption c_1, period 2 consumption $c_{2,}$ and "deposits" d, which are riskless securities owned by the household. u and v are strictly increasing, strictly concave, twice continuously differentiable, and satisfy Inada conditions. The household's only choice is how to invest or consume its initial wealth W_H. It may purchase either riskless assets, which yield the direct benefit $v(d)$ as well as a riskless cashflow at period 2, or other securities issued by the intermediary or non-financial firms. It cannot sell short or borrow to invest.

The household's problem is to maximize its expected utility given a deposit rate i_d and prices p_s of securities s which pay stochastic cashflows δ_s in period 2. Given the rate i_d, the price of one deposit at time 1 is $\frac{1}{1+i_d}$. Consumption at period 2 is the sum of payoffs from deposits and securities $c_2 = \int \delta_s q_H(s)\, ds + d$, where $q_H(s)$ is the quantity of security s purchased by the household. $q_H(s)$ cannot be negative, since short selling is not allowed. The household's problem can be written as

$$\begin{aligned} &\max_{d, q_H(.), c_1} u(c_1) + E\left[u\left(\int_0^1 \delta_s q_H(s)\, ds + d\right)\right] + v(d) \qquad (1.2)\\ &\text{subject to } c_1 + \frac{d}{1+i_d} + \int_0^1 p_s q_H(s)\, ds = W_H \text{ (budget constraint)},\\ &q_H(.) \geq 0 \text{ (short sale constraint)} \end{aligned}$$

The first order conditions for deposits d (which has an interior solution since $v'(0) = \infty$) and for the quantity $q_H(s)$ to purchase of security s are

$$u'(c_1) = (1+i_d)\left(E\left[u'(c_2)\right] + v'(d)\right) \qquad (1.3)$$

$$p_s \geq E\left[\frac{u'(c_2)}{u'(c_1)}\delta_s\right] \qquad (1.4)$$

where inequality 1.4 must be an equality if $q_H(s) > 0$.

Two features of the household's optimal investments are notable. First, inequality 1.4

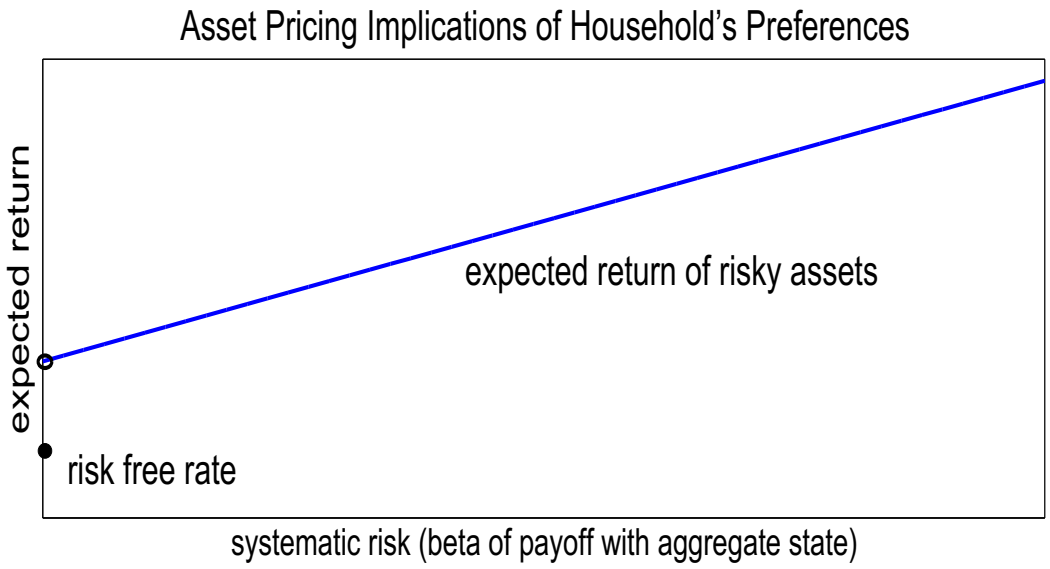

Figure 1.2: *Safety demand and a low risk-free rate.*

implies that only securities owned by the household must satisfy the consumption Euler equation. If other agents (such as an intermediary) are willing to pay more for an asset than the household, the price will not reflect the household's preferences. This is because the household is constrained from shorting assets it considers overvalued. Second, the extra marginal utility $v'(d)$, reflecting the "safe asset premium" households are willing to pay for riskless securities, depresses the risk-free rate. The interest rate $i_d = \frac{u'(c_1)}{(v'(d)+Eu'(c_2))} - 1$ for safe assets would equal the strictly higher rate $\frac{u'(c_1)}{Eu'(c_2)} - 1$ if $v'(d) = 0$. Safe asset demand leads to a low risk-free rate relative to the pricing of other assets owned by the household, as (Krishnamurthy Vissing-Jorgensen 2012) shows empirically in the pricing of treasury securities. This is illustrated below.

If all asset prices reflected the household's willingness to pay, the gap between the risk-free rate and the pricing of risky assets could be exploited by an arbitrage trade. Suppose that a financial intermediary buys a diversified portfolio $q_I(.)$ of risky assets that pays $\int \delta_s q_I(s)\, ds = \delta_p$ equal to $\delta_{p,good}$ in the good state and $\delta_{p,bad} < \delta_{p,good}$ in the bad state. The price of this portfolio is $E\left[\frac{u'(c_2)}{u'(c_1)}\delta_p\right]$. If the intermediary sells a riskless security backed by its portfolio paying $\delta_{p,bad}$ and a residual claim paying $\delta_{p,good} - \delta_{p,bad}$ in the good state, the

household would be willing to pay $E\left[\frac{u'(c_2)}{u'(c_1)}\delta_p\right] + \frac{v'(d)}{u'(c_1)}\delta_{p,bad}$ to buy both securities issued by the intermediary. This yields an arbitrage profit of $\frac{v'(d)}{u'(c_1)}\delta_{p,bad}$, equal to the quantity $\delta_{p,bad}$ of riskless assets produced by the arbitrage trade times the "safety premium" $\frac{v'(d)}{u'(c_1)}$ that households will pay for a riskless asset. This arbitrage trade, selling safe and risky tranches backed by a diversified portfolio of risky assets, is precisely what I refer to as safety transformation. The next section develops a model of how intermediaries exploit this arbitrage opportunity and the frictions they face when doing so.

Intermediary's problem The intermediary is a publicly traded firm that maximizes the value of its equity subject to an agency problem faced by its management. Unlike the household, the intermediary is able to issue securities backed by its asset portfolio, allowing it to increase the supply of riskless assets. It can raise funds either by issuing equity or other possible securities, and in equilibrium all securities it issues must be sold to the household. Riskless securities issued by the intermediary trade at the risk-free rate (reflecting the household's safety demand), while risky securities are priced by the consumption Euler equation. The cashflows $(\delta_{I,1}, \delta_{I,2})$ paid by the intermediary at $t = 1, 2$ in risky securities are valued as

$$E\left[\frac{u'(c_2)}{u'(c_1)}\delta_{I,2}\right] + \delta_{I,1}. \tag{1.5}$$

Because this value does not depend on how the intermediary divides its risky cashflows (i.e. into a risky debt security as well as equity), the intermediary can be assumed to issue only equity and riskless debt without loss of generality.

The management of the intermediary faces an agency problem because the assets on its balance sheet have payoffs that are observable only to its management.[12] As a result, the intermediary's management is able to misreport the payoff of its asset portfolio and

[12]As noted above, the intermediary (and non-financial firms) also faces a second agency problem between its owners and other investors, where owners can instruct management to divert resources or raise additional funding to manipulate security payoffs. Because this agency problem has no effect when a firm issues only debt and equity, the analysis in this section ignores it since these are the only securities the intermediary issues.

divert part of the difference between the true and reported payoff. If the true portfolio payoff is $\delta_{P,true}$ and the intermediary reports $\delta_{P,reported} < \delta_{P,true}$, the management can divert $C\left(\delta_{P,true} - \delta_{P,reported}\right) < \delta_{P,true} - \delta_{P,reported}$. Management must therefore be given some profit sharing to incentivize for truthful reporting. Because the intermediary's portfolio is not exposed to idiosyncratic risk, its payoff at time 2 depends only on the binary aggregate state. Management's payment cannot explicitly depend on the uncontractible aggregate state or the output of other firms but only on the intermediary's cashflows that management reports. The intermediary's management therefore needs only a payment $C\left(\delta_{P,good} - \delta_{P,bad}\right)$ in the good state to ensure the truthful reporting of its asset payoff, where $\delta_{P,s}$ is the payoff of its portfolio in state s. Because management diverts less than the total amount of output it destroys, it is optimal to induce management not to divert funds. Since this risky payoff cannot be used to back deposits and therefore must be sold as part of the intermediary's equity, the agency problem faced by the intermediary can be interpreted as a cost of raising equity capital. The cost $C\left(\delta_{P,good} - \delta_{P,bad}\right)$ can also be interpreted as a reduced form cost of paying dividends to the intermediary's equityholders, since $\delta_{P,bad}$ is the amount of riskless deposits it can issue.

At time 1, the equity e_1 raised by the intermediary is a negative payout $\delta_{I,1} = -e_1$. At time 2, the intermediary's payout is the total cashflows from its security portfolio minus the promised payments to depositors and management $\delta_{I,2} = \int_0^1 \delta_s q_I(s)\, ds - d - C\left(\int_0^1 \left(\delta_s - E_{bad}\delta_s\right) q_I(s)\, ds\right)$, where $q_I(s)$ is the quantity of security s purchased by the intermediary. The intermediary's problem can be written as

$$\max_{e_1,d,q_I(.)} E\left[\frac{u'(c_2)}{u'(c_1)}\left(\overbrace{\int_0^1 \delta_s q_I(s)\,ds}^{\text{cashflows}} - \overbrace{d - C\left(\int_0^1 (\delta_s - E_{bad}\delta_s)\,q_I(s)\,ds\right)}^{\text{payments to depositors and management}}\right)\right] - \overbrace{e_1}^{\text{equity issued}} \quad (1.6)$$

subject to: $e_1 + \frac{d}{1+i_d} = \int_0^1 p_s q_I(s)\,ds$ (budget constraint)

$\left(\int_0^1 \delta_s q_I(s)\,ds - d\right) \geq 0$ in all states of the world (solvency constraint)

$q_I(.) \geq 0$ (short sale constraint).

To simplify this problem, note that the budget constraint implies $e_1 = \int_0^1 p_s q_I(s)\,ds - \frac{d}{1+i_d}$. In addition, because of the safety premium, deposits are a cheaper source of funding for the intermediary than equity. The intermediary should therefore enough deposits to make its solvency constraint bind. This implies $d = \int (E_{bad}\delta_s)\,q_I(s)\,ds$, since $E_{bad}\delta_s \leq E_{good}\delta_s$ so the solvency constraint binds in the bad state.

The intermediary's problem reduces to

$$\max_{q_I(.)\geq 0} E\left[\begin{array}{c} \frac{u'(c_2)}{u'(c_1)}\left(\int_0^1 \delta_s q_I(s)\,ds - C\left(\int_0^1 (\delta_s - E_{bad}\delta_s)\,q_I(s)\,ds\right)\right) \\ -\int_0^1 p_s q_I(s)\,ds \;+\; \frac{v'(d)}{u'(c_1)}\int_0^1 (E_{bad}\delta_s)\,q_I(s)\,ds \end{array}\right] \quad (1.7)$$

which has the first order condition for each $q_I(s)$

$$p_s \geq \overbrace{E\frac{u'(c_2)}{u'(c_1)}\delta_s}^{\text{household's willingness to pay}} + \overbrace{\frac{v'\left(\int (E_{bad}\delta_s)\,q_I(s)\,ds\right)}{u'(c_1)}}^{\text{safety premium}} \overbrace{E_{bad}\delta_s}^{\text{deposits backed by asset}} - \quad (1.8)$$

$$\overbrace{C'\left(\int_0^1 (E_{good}\delta_s - E_{bad}\delta_s)\,q_I(s)\,ds\right)}^{\text{agency cost of equity}} \overbrace{\frac{1}{2}\frac{u'\left(c_2^{good}\right)}{u'(c_1)}(E_{good}\delta_s - E_{bad}\delta_s)}^{\text{equity required to buy asset}}$$

with equality whenever $q_I(s) > 0$. This expression uses the fact that $C'(.) \neq 0$ only in the good state, since management must be paid only then.

The intermediary's willingness to pay for asset s depends only on $E_{good}\delta_s$ and$E_{bad}\delta_s$, since the intermediary's portfolio is diversified. The distribution each asset's idiosyncratic returns given the aggregate state is irrelevant. By pooling and then tranching a portfolio of assets, the intermediary diversifies away its exposure to idiosyncratic risk. The intermediary can therefore back more riskless assets than would be possible by selling junior and senior tranches backed by individual assets. This is related to "risk diversification effect" of (DeMarzo 2005), who finds that pooling and tranching is an optimal strategy for issuing safe, informationally insensitive assets in the presence of asymmetric information.

The intermediary's required return for exposure to aggregate risk reflects its cost of equity financing and cheapness of deposit financing. As part of a diversified portfolio, a quantity $E_{bad}\delta_s$ of riskless securities can be backed by asset s, while the remaining good state payoff $E_{good}\delta_s - E_{bad}\delta_s$ increases the agency costs of equity. Because deposits earn the safety premium reflected in a low risk-free rate, the intermediary is willing to pay more than the household for assets that back large quantities of deposits. However, any systematic risk in an asset owned by the intermediary increases the intermediary's agency cost of equity financing. This makes the intermediary effectively more risk averse than the household.

Asset prices and portfolio choices The investment decisions of the household and intermediary described above can be used to solve for asset prices and determine which assets are owned by which investor. Assets owned by the intermediary imply a strictly lower risk-free rate and higher price of systematic risk than assets owned by the household. This segmentation in asset prices reflects the intermediary's ability to back riskless deposits with its asset portfolio and its agency cost of bearing risk. Low systematic risk assets are held by the intermediary and high systematic risk assets are held by the household, allowing the intermediary to issue many deposits while minimizing the agency costs it faces.

An expression for asset prices follows directly from the consumer's and intermediary's optimal investment decisions 1.4 and 1.8. Since every asset must be owned by some agent, at least one of these inequalities must hold with equality. If the household and intermediary are willing to pay different amounts for an asset, the agent willing to pay the most buys its

entire supply. This yields the following result.

Proposition 1.1 *(segmented asset prices) For any asset s in positive supply with payoffs δ_s at time 2, its price at time 1 is the maximum of the willingness to pay of the two agents*

$$p_s = \overbrace{E\left[\frac{u'(c_2)}{u'(c_1)}\delta_s\right]}^{\text{household's willingness to pay for asset}} + \max[0, \overbrace{E_{bad}\delta_s}^{\text{safe debt backed by asset}} \overbrace{\frac{v'\left(\int_0^1 (E_{bad}\delta_s)\, q_I(s)\, ds\right)}{u'(c_1)}}^{\text{safety premium}} - \overbrace{\frac{u'\left(c_2^{good}\right)\left[(E_{good}\delta_s - E_{bad}\delta_s)\right]}{2u'(c_1)}}^{\text{equity required to purchase asset}} \overbrace{C'\left(\int_0^1 (E_{good}\delta_s - E_{bad}\delta_s)\, q_I(s)\, ds\right)}^{\text{agency cost of equity}}]. \quad (1.9)$$

If the household and intermediary are willing to pay different prices for asset s, the entire supply of the asset is bought by the agent willing to pay more.

The pricing kernel of assets owned by the intermediary implies a risk-free rate

$$\left[E\frac{u'(c_2) + v'\left(\int_0^1 (E_{bad}\delta_s)\, q_I(s)\, ds\right)}{u'(c_1)}\right]^{-1} - 1 \quad (1.10)$$

strictly below the risk-free rate $\left(E\frac{u'(c_2)}{u'(c_1)}\right)^{-1} - 1$ implied by the pricing kernel of risky assets owned by the household. This is because the intermediary can use riskless payoffs to back deposits and meet the household's safety demand, while the household is unable to pool and tranche to create riskless assets.

Assets owned by the intermediary reflect a strictly higher price of systematic risk than assets owned by the household. A unit of consumption in the good state is worth $\frac{1}{2}\frac{u'\left(c_2^{good}\right)}{u'(c_1)}$ to the household but only $\frac{1}{2}\frac{u'\left(c_2^{good}\right)}{u'(c_1)}\left(1 - C'\left(\int_0^1 (E_{good}\delta_s - E_{bad}\delta_s)\, q_I(s)\, ds\right)\right)$ to the intermediary. The multiplicative factor $1 - C'(.)$ reflects the fact that good state payoffs increase the intermediary's agency costs, making these payoffs less valuable. This agency cost implies that the intermediary requires greater compensation for being exposed to systematic risk than the household.

This asset pricing result also characterizes the portfolios of the household and interme-

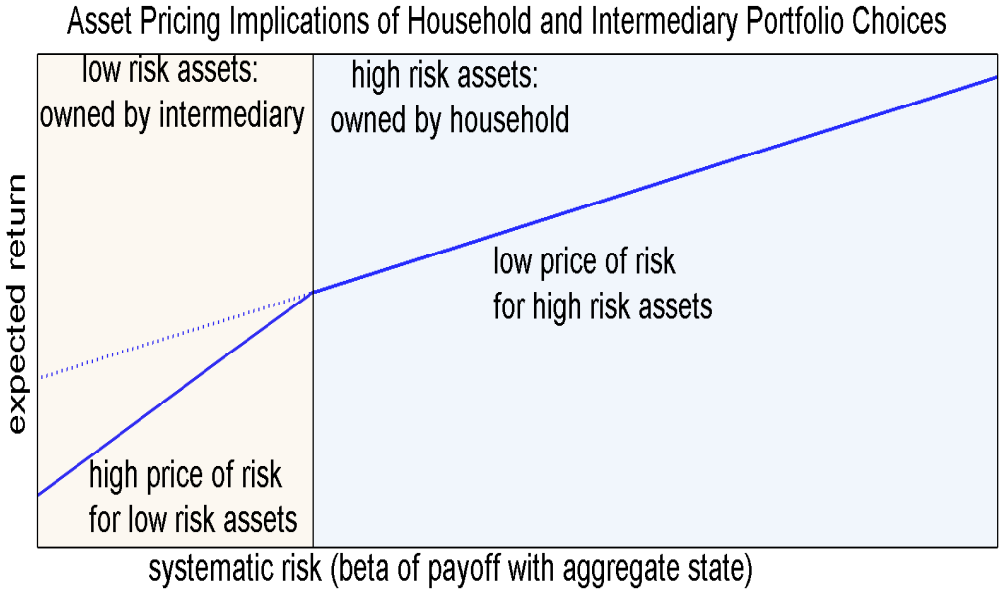

Figure 1.3: *Segmentation of low and high risk asset markets*

diary. The difference between these two agents' willingness to pay for asset s is

$$\frac{v'\left(\int_0^1 \left(E_{bad}\delta_s\right) q_I\left(s\right) ds\right)}{u'\left(c_1\right)} E_{bad}\delta_s - \tag{1.11}$$
$$C'\left(\int_0^1 \left(\left[E_{good} - E_{bad}\right]\delta_s\right) q_I\left(s\right) ds\right) \frac{u'\left(c_2^{good}\right)}{2u'\left(c_1\right)} \left(\left[E_{good} - E_{bad}\right]\delta_s\right).$$

The intermediary buys assets for which expression 1.11 is positive, while the household buys assets for which it is negative. The sign of the expression is determined by the ratio $\frac{E_{good}\delta_s}{E_{bad}\delta_s}$, yielding the following corollary.

Corollary 1.1 *(intermediary owns low systematic risk assets)*

Let $k^* = 1 + \frac{2v'\left(\int \left(E_{bad}\delta_s\right) q_I(s) ds\right)}{u'\left(c_2^{good}\right) C'\left(\int_0^1 \left(E_{good}\delta_s - E_{bad}\delta_s\right) q_I(s) ds\right)}$. *The intermediary buys all assets who cashflows* δ_s *satisfy* $\frac{E_{good}\delta_s}{E_{bad}\delta_s} < k^*$, *and the household buys all assets with* $\frac{E_{good}\delta_s}{E_{bad}\delta_s} > k^*$. *The pricing kernel for riskier assets owned by the household implies a strictly higher risk-free rate and strictly lower price of systematic risk than the pricing kernel for less risky assets owned by the intermediary.*

These asset pricing and portfolio choice results can be summarized by the "kinked"

securities market line above. Low risk assets owned by the intermediary earn a higher risk-adjusted return that high risk assets owned by the household. This segmentation occurs because intermediaries obtain cheap financing by meeting the household's demand for safe assets. In models with leverage constraints (e.g. Frazzini Pedersen 2014, Black 1972) agents who are more easily able to borrow can take risk by holding levered portfolios of low risk assets. Risk tolerant agents who are borrowing constrained must hold unlevered portfolios of high risk assets, bidding up the prices of these assets. The intermediary's ability to hold a diversified pool of assets that backs a large riskless tranche of debt is the advantage it has in borrowing.

Non-financial firm's problem This section shows how non-financial firms issue securities to exploit asset market segmentation. The intermediary is willing to pay more than the household for securities with low systematic risk but less for securities with high systematic risk. Non-financial firms therefore find it optimal to sell a low risk security to the intermediary and a high risk security to the household, obtaining a strictly higher valuation than either investor would pay for the entire firm. Under the restrictions imposed below, the firm optimally chooses to issue debt bought by the intermediary and equity bought by the household. Its optimal leverage is determined by the risk preferences of the household and intermediary, illustrating how market segmentation violates the Modigliani-Miller theorem.

Each non-financial firm $i \in [0,1]$ has exogenous cashflows f_i at time 2, subject to aggregate and idiosyncratic shocks. f_i is respectively distributed according to $F(f_i|good)$ and $F(f_i|bad)$ in the good and bad aggregate states. The cashflows of non-financial firms are conditionally independent given the aggregate state. I impose the following condition on f_i. It implies that more senior claims on the firm's cashflows have lower systematic risk, so a more levered firm has debt with higher systematic risk.[13]

[13]Condition 3 (i) is equivalent to the monotone hazard ordering $\frac{f_{f_i}(D|good)}{\Pr(f_i>D|good)} < \frac{f_{f_i}(D|bad)}{\Pr(f_i>D|bad)}$ where $f_{f_i}(.|H)$ is the conditional density of f_i given state H.

Condition 1.1 (i) $\frac{\partial}{\partial D}\frac{\Pr(f_i>D|good)}{\Pr(f_i>D|bad)} > 0$ *for all* $D > 0$.

$(ii)\ \Pr(f_i > 0|good) = \Pr(f_i > 0|bad) = 1$

$(iii)\ \lim_{D\to\infty}\frac{\Pr(f_i>D|good)}{\Pr(f_i>D|bad)} = \infty$

Non-financial firms are subject to the same agency problems as the intermediary between its owners and management and also between owners and other investors. If the true cashflow is f_i and the firm's management gives $f_i' < f_i$ to outside investors, it can divert $C(f_i - f_i')$. The firm faces a second agency problem between its owners and other outside investors, that after management has diverted funds, the owners can either destroy resources or covertly raise additional financing at the market rate (both raised and paid back in period 2). As in (Innes 1990), this agency problem between owners and other investors forces owners to issue securities whose payoffs are increasing in the firm's cashflows. The firm also cannot issue securities whose payoffs explicitly depend on the uncontractible aggregate good or bad state.

The appendix shows that the firm optimally issues debt and equity securities and provides its management with the incentive to never divert resources. The remainder of this section takes this result as given and analyzes the firm's optimal capital structure. In the previous section, it was shown without loss of generality that the intermediary would choose to issue equity and riskless debt, so the optimal behavior of the intermediary is not constrained by this additional agency problem.

Proposition 1.2 *Each non-financial firm i with cashflows* f_i *chooses to pay its management* $C(f_i)$, *which makes it incentive compatible for management to truthfully report the firm's earnings. The remaining cashflows* $x_i = f_i - C(f_i)$ *are optimally divided into a debt security of face value* D_i *which pays* $x_i^D = \min(x_i, D_i)$ *and an equity security which pays* $x_i^E = \max(x_i - D_i, 0)$. *Once* f_i *is reported to the firm's owners, it is optimal for the owners to neither raise additional hidden financing or to destroy resources.*

Firm i's cashflows $x_i = f_i - C(f_i)$ available to outside investors and its choice to issue debt and equity are now taken as given. Since $f_i - C(f_i)$ is strictly increasing in f_i, the

condition imposed on f_i also applies to x_i. The non-financial firm maximizes its total market value by choosing its face value of debt D_i. The firm takes as given asset prices implied by the behavior of the household and intermediary. Proposition 2 implies that the sum of the firm's debt and equity prices can be written as

$$p_E^i + p_D^i = E\frac{u'(c_2)}{u'(c_1)}x_i + \max\left(0, K_1 E_{bad} x_i^D - K_2\left(E_{good} - E_{bad}\right)x_i^D\right) + \max\left(0, K_1 E_{bad} x_i^E - K_2\left(E_{good} - E_{bad}\right)x_i^E\right) \tag{1.12}$$

where $K_1 = \frac{v'\left(\int_0^1 (E_{bad}\delta_s) q_I(s) ds\right)}{u'(c_1)} > 0$ and $K_2 = \frac{u'\left(c_2^{good}\right)}{2u'(c_1)} C'\left(\int_0^1 \left(E_{good}\delta_s - E_{bad}\delta_s\right) q_I(s)\, ds\right) > 0$. The signs of these two constants reflect the fact that the intermediary is willing to pay more than the household for riskless payoffs but less for payoffs in the good state. If $K_1 = K_2 = 0$, which would hold if household and intermediary were willing to pay the same for all securities, firm i's market value would be independent of it's capital structure. The fact that $p_E^i + p_D^i$ depends on the face value of debt D_i illustrates how asset market segmentation violates Modigliani-Miller. This is related to (Baker Hoeyer Wurgler 2016), who argues empirically that market segmentation influences capital structure decisions.[14]

The firm chooses the face value of debt D_i to maximize its market value $p_D^i + p_E^i$. If there is a D_i at which the intermediary buys one security issued by the firm and the household buys the other, $p_E^i + p_D^i$ must be strictly greater than either investor's willingness to pay for the firm's total cashflows x_i. If such a D_i is optimal, it must satisfy the first order condition

$$K_1 \Pr(x_i > D_i | bad) - K_2\left(\Pr(x_i > D_i | good) - \Pr(x_i > D_i | bad)\right) = 0 \tag{1.13}$$

since $\frac{\partial E_H x_i^D}{\partial D_i} = \frac{\partial E_H \min(x_i, D_i)}{\partial D_i} = \Pr(x_i > D_i | H) = -\frac{\partial E_H x_i^E}{\partial D_i}$ for $H = bad$ and $H = good$. This condition implies that a security which pays 1 when $x_i > D_i$ and 0 otherwise is of equal

[14]The analysis in this section provides a somewhat novel framework for analyzing corporate capital structure. The idea that risk aversion heterogeneity can influence corporate capital structure is presented in (Allen Gale 1988) but only in the case where debt is riskless, and the idea does not seem to appear in later literature. The analysis here is mathematically similar to (Simsek 2013)'s study of collateralized margin lending under belief disagreement.

value to the household and the intermediary. Because an increase in D_i increases the payout of debt only in states of the world where $x_i > D_i$, this marginal transfer of resources from equity to debt has no effect on firm i's total market value $p_E^i + p_D^i$.

The first order condition 1.13 uniquely determines the ratio $\frac{\Pr(x_i > D_i|good)}{\Pr(x_i > D_i|bad)}$. For this ratio to determine firm i's capital structure, there must be precisely one D_i for which 1.13 holds, which follows from the assumption that $\frac{\Pr(x_i > D_i|good)}{\Pr(x_i > D_i|bad)}$ is strictly increasing in D_i and has range $[1, \infty)$.

As well as providing a unique solution to equation 1.13 for any $K_1, K_2 > 0$, this condition also implies that

$$\frac{E_{good}\left(\min\left(x_i, D_i\right)\right)}{E_{bad}\left(\min\left(x_i, D_i\right)\right)} < \frac{\Pr\left(x_i > D_i|good\right)}{\Pr\left(x_i > D_i|bad\right)} < \frac{E_{good}\left(\max\left(x_i - D_i, 0\right)\right)}{E_{bad}\left(\max\left(x_i - D_i, 0\right)\right)}. \tag{1.14}$$

When D_i satisfies 1.13, firm i's debt has low enough systematic risk to be bought by the intermediary, while firm i's equity is bought by the household. This verifies that 1.13 determines firm i's unique optimal capital structure. Plugging in the definitions of K_1 and K_2 yields the following proposition.

Proposition 1.3 *(optimal non-financial capital structure) If condition 3 is satisfied, the optimal face value of debt D_i for firm i is the unique D_i which solves*

$$v'\left(\int_0^1 \left[E_{bad}\delta_s\right] q_I\left(s\right) ds\right) - \tag{1.15}$$
$$\frac{1}{2}u'\left(c_2^{good}\right) C'\left(\int_0^1 \left[\left(E_{good} - E_{bad}\right)\delta_s\right] q_I\left(s\right) ds\right)\left(\frac{\Pr\left(x_i > D_i|good\right)}{\Pr\left(x_i > D_i|bad\right)} - 1\right) = 0.$$

When D_i is chosen optimally, firm i's debt and equity are respectively bought by the intermediary and the household.

The intermediary's ability to issue cheap riskless debt implies that non-financial firms are also able to issue cheap debt as long as its systematic risk is low enough. As shown above, the intermediary's cost of capital is reflected in segmented asset prices. This proposition builds on this result by showing how the non-financial sector responds to market segmentation. The household's demand for safe assets (measured by $v'\left(\int_0^1 \left[E_{bad}\delta_s\right] q_I\left(s\right) ds\right)$) and

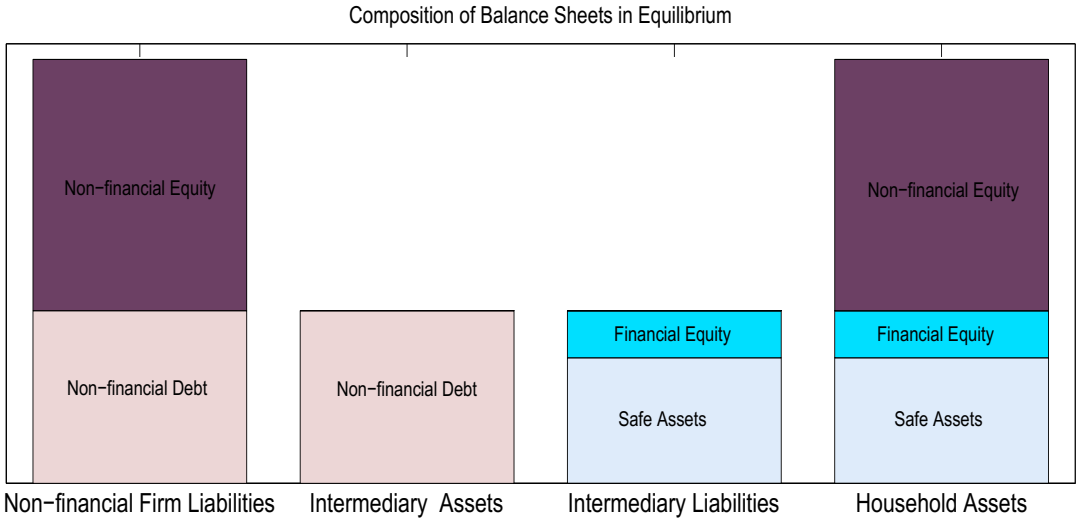

Figure 1.4: *Balance sheets in the model.*

the intermediary's agency cost of equity (measured by $C'\left(\int_0^1 \left[\left(E_{good} - E_{bad}\right)\delta_s\right] q_I(s)\, ds\right)$) jointly determine the non-financial sector's optimal capital structure.

The proposition provides a cross-sectional prediction for capital structure. Firms for whom $\frac{\Pr(x_i > D_i | good)}{\Pr(x_i > D_i | bad)}$ is greater at each D_i choose to issue less debt. This is consistent with (Schwert and Strebulaev 2015)'s finding that firms with more cyclical cashflows are less levered.

The results derived above can be thought of as applying to household borrowing. If the household could buy a durable consumption good providing consumption services x_i and get a collateralized loan of face value D_i backed only by this consumption good (such as a mortgage backed by a house), the optimal amount to borrow would also be described by condition 1.15.

This proposition also determines the composition of household, intermediary, and non-financial firm balance sheets. Households invest in the equity of both the financial and non-financial sectors and also hold safe assets. Intermediaries, who supply the safe assets, invest in the debt of the non-financial sector and must issue a buffer of equity to bear the

risk in their portfolio of debt securities. Non-financial firms sell their debt to intermediaries and equity to households, arbitraging the differing prices of risk for low and high risk securities. The fact that equities are held by households while debt securities are held by the intermediary is endogenous and not assumed. Any agent is able to buy any security, but intermediaries are willing to pay more for debt securities but less for equities than households.[15]

One final implication of this proposition is that it explains the "credit spread puzzle" in debt securities and "low risk anomaly" in equities. The capital structure choices of the non-financial sector ensure debt and equity securities live on opposite sides of the kink in the securities market line. As a result, the debt and equity markets are endogenously segmented, with a greater price of risk in the debt market. As shown in (Huang and Huang 2012), many structural credit risk models underestimate the spreads on corporate bonds when calibrated to data from equity markets, a finding referred to as the credit spread puzzle. Such a result can either be interpreted as a failure of many structural models (and some recent ones do match it in a no arbitrage framework) or taken as evidence that risk is priced more expensively in debt markets than in equity markets, as naturally occurs in my model. The high price of risk in debt markets occurs jointly with a low price of risk in equity markets. This rationalizes the "low risk anomaly" (e.g. Black, Jensen, Scholes 1972, Baker, Bradley, Taliaferro 2014), which finds that for simple measures of risk (such as covariance with returns on an equity market index), the price of systematic risk in equity markets is too small to jointly explain a low risk-free rate and high expected return on equities. This naturally occurs in my model, since the zero beta rate implied by the pricing of equities is strictly above the true risk-free rate, with the spread reflecting the demand for safe assets.

[15]If the non-financial firms were able to issue some riskless debt (ruled out by $\frac{\partial}{\partial D_i}\frac{\Pr(f_i>D_i|good)}{\Pr(f_i>D_i|bad)} > 0$), an equilibrium in which households held both financial debt and a riskless senior tranche of non-financial debt could also occur.

Equilibrium This section characterizes the model's equilibrium, endogenously determining the intermediary's cost of capital, which has been taken as given in the results above.

Definition 1.1 *An equilibrium is a set of consumption allocations* (c_1, c_2), *intermediary and household portfolios* $(q_I(s), q_H(s))_{s\in[0,1]}$, *asset prices* $(p_s)_{s\in[0,1]}$, *deposits* d, *intermediary equity and non-financial firm debt issuance* $(D_i)_{i\in[0,1]}$ *such that*

(i) *The household, intermediary, and non-financial firms behave optimally as described above.*

(ii) *Household and intermediary budget constraints are satisfied.*

(iii) *Consumption at time 2 equals the total output of the non-financial sector,* $c_2 = \int_0^1 f_i di$, *and consumption at time 1 equals output at time 1,* $c_1 = C_1$.

Because the intermediary's portfolio is composed entirely of the debt of the non-financial sector as shown in proposition 4, the quantity d of riskless assets the intermediary can issue and residual payoff e to equityholders in good states are simply

$$d = \int_0^1 E_{bad} \min(x_i, D_i)\, di. \tag{1.16}$$

$$e = \int_0^1 (E_{good} - E_{bad}) \min(x_i, D_i)\, di. \tag{1.17}$$

Plugging these expressions into each firm i's optimal capital structure decision yields

$$v'\left(\int_0^1 E_{bad} \min(x_i, D_i)\, di\right) - \tag{1.18}$$
$$\frac{u'\left(c_2^{good}\right)}{2} C'\left(\int_0^1 (E_{good} - E_{bad}) \min(x_i, D_i)\, di\right)\left(\frac{\Pr(x_i > D_i | good)}{\Pr(x_i > D_i | bad)} - 1\right) = 0.$$

which depends only on exogenous variables and the face value of debt D_i each non-financial firm issues.

Proposition 1.4 *(equilibrium) The model's unique equilibrium is characterized by a face value of debt* D_i *for each non-financial firm* i *that solves equation 1.18*

Proof. Under the regularity conditions on each firm i's cashflows, the ratio $r = \frac{\Pr(x_i > D_i | good)}{\Pr(x_i > D_i | bad)}$ uniquely determines the debt face value D_i of each firm i, and D_i is continuous and increasing in r. The expression in equation 1.18 is a strictly decreasing function of r, $M(r)$, which equals 0 in equilibrium. $M(0) > 0$ and $M(\infty) < 0$, so M crosses zero once and a unique equilibrium exists. ■

This characterization of equilibrium illustrates the interaction between three forces. The household's demand for safe assets reflected in the function $v(.)$ determines how great the incentives are for the intermediary to create riskless assets. The cost of creating riskless assets depends on the severity of the intermediary's agency problem which is reflected in the function $C(.)$, which determines how costly it is for the intermediary to own risky assets. Finally, the cost of creating riskless assets depends on how much risk the intermediary must take in order to back a given quantity of riskless assets. This is determined by the distribution of each firm's marketable cashflows x_i. The more systematic risk non-financial firms are exposed to, the more costly equity financing is required for the intermediary to back deposits.

Equation 1.18 illustrates how the intermediary's portfolio which pays $\int_0^1 \min(x_i, D_i)\, di$ determines the intermediary's cost of capital, both in terms of the premium $v'\left(\int_0^1 E_{bad} \min(x_i, D_i)\, di\right)$ on riskless deposits and the cost $C'\left(\int_0^1 (E_{good} - E_{bad}) \min(x_i, D_i)\, di\right)$ of a marginal increase in the riskiness of the intermediary's portfolio. These costs, which are reflected in equilibrium asset prices then determine the optimal capital structure of the non-financial sector. Because the debt of the non-financial sector is the asset side of the intermediary's balance sheet, ensuring that the non-financial sector issues the optimal amount of debt at the intermediary's equilibrium cost of capital solves for the model's unique equilibrium.

The above diagram summarizes the implications of this equilibrium. The low risk assets owned by the intermediary are now the debt of the non-financial sector, while the high risk assets owned by the household are now the equity of both the financial and non-financial sectors. As a result, the market price of risk is strictly higher in the debt than the equity

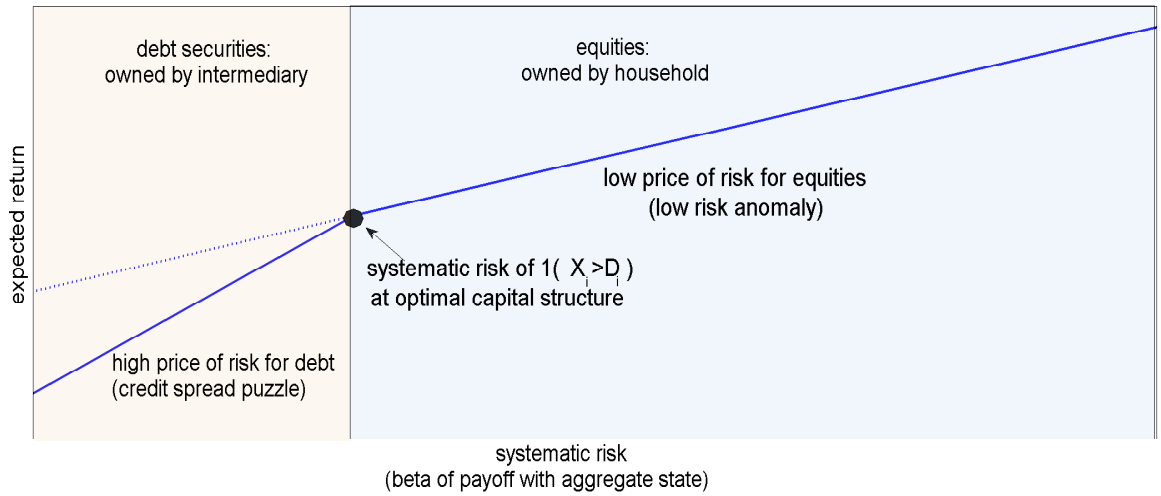

Figure 1.5: *Debt and equity market segmentation*

market as discussed above. The optimal capital structure of the non-financial sector is determined by how segmented the debt and equity markets are. The optimal non-financial capital structure arbitrages between these two markets, with the first order condition that a small increase in leverage has no marginal effect on a non-financial firm's value. This first order condition is summarized by the dot in the above diagram, since the payoff $1(x_i > D_i)$ is the marginal transfer from equity to debt of increasing firm $i's$ leverage. The household and intermediary have the same willingness to pay for $1(x_i > D_i)$, which implies that it must lie at the intersection of their two security market lines. In equilibrium, the gap between the two intercepts is determined by the premium on riskless assets, while the higher slope of the intermediary's security market line reflects its agency costs of owning risky assets.

Discussion Three basic assumptions are crucial for the model's key results. First, there must be a demand for safe, money-like assets that pushes the risk-free rate below the rate implied by the pricing of equities. This gives intermediaries and non-financial firms an incentive to separate their assets into safe and risky tranches in order to borrow as much as possible at the low risk-free rate. Second, the non-financial sector must face some

constraints that make it difficult for them to issue safe assets directly. Because non-financial firms are exposed to (full support) idiosyncratic risk they cannot hedge and issue debt and equity rather than arbitrary Arrow-Debreu securities, they cannot issue riskless assets. Finally, intermediaries must face some cost of bearing risk, so that they choose to only buy low risk debt securities. If intermediaries had no cost of bearing risk, they would buy the entire non-financial sector in order to use its entire output in the bad state of the world to back the safe debt they issue. Equation 1.18 illustrates in a single expression how these three basic assumptions interact. The benefit $v'(.)$ of issuing more riskless securities are balanced against the agency cost $C'(.)$ of increasing the risk on the intermediary's balance sheet, where the amount of risk bearing required is determined by the ratio $\frac{\Pr(x_i > D_i|good)}{\Pr(x_i > D_i|bad)}$ that depends on the riskiness of each non-financial firm's output.[16]

The assumption that households place a special value on riskless assets is common in both recent theoretical and quantitative models, is consistent with empirical evidence referenced in the introduction, but does not have a microfoundation in this paper or the related literature. In my model, only riskless assets are special, whereas one may imagine that bank deposits and money market fund shares are exposed to small risks while still being "money-like." Gorton and Pennacchi (1990) provides a model that shows why riskless assets are the most liquid, but the question of how risky an asset can be while still functioning as a form of money is open. Assuming that deposits must be riskless ignores the possibility of bank runs which could be studied in a similar framework in which depositors withdraw only when deposits become too risky. However, if banks can tap a cheap source of funding by issuing low risk deposits, the basic insight of this paper still holds. Banks and similar intermediaries will choose the assets they hold to issue as many deposits and as little equity as possible if deposits are cheap and equity is expensive.

In order for my model of intermediation to be consistent with the data, the pricing kernels for low and high risk assets must be different. If there are no unexploited arbitrage

[16]The agency problem that makes risk bearing costly for the intermediary also applies to non-financial firms so that all firms are ex ante identical. For non-financial firms whose project risk is exogenous, this agency problem is just an unavoidable cost that has no important implications.

opportunities, an intermediary cannot create value by buying publicly available securities in order to sell other publicly available securities. The paper's asset pricing implications therefore provide a falsifiable way of evaluating the model, which sets it apart from other models of financial intermediation that do not speak to the market for publicly available securities. While my model only provides qualitative predictions, it is consistent with both cross sectional and time series evidence on the credit spread puzzle and low risk anomaly. My model implies that the expected return on the riskiest bonds are close to the securities market line implied by equities, and (Huang and Huang 2012) show that the credit spread puzzle is much less severe for junk than investment grade bonds. In addition, (Gilchrist Zakrajsek 2012) show that in the time series, the severity of the credit spread puzzle comoves strongly with measures of intermediary risk taking. (Frazzini Pedersen 2014) also shows that the low risk (low beta) anomaly is largest when measures of distress in the intermediary sector are high. There is a literature that attempts to rationalize the asset pricing facts I emphasize in a no arbitrage framework, and it is an open question going forward whether a quantitative model with constrained intermediaries best explains the data.

To interpret the model, it is useful to ask what are the financial intermediaries it describes. Intermediaries in my model hold diversified portfolios of debt, issue a safe, senior liability (deposits) backed by this portfolio and a junior liability (equity) that bears the risk in the intermediary's asset portfolio. Banks are the most straightforward fit to the model, though certain elements of the shadow banking system such as broker dealer or investment banks fit as well. Broker dealer and investment banks often fund themselves heavily with short term debt (some of which is collateralized), and this short term debt is often bought by money market funds. Integrating the broker dealer and the money market fund creates an entity like the intermediary in my model, though broker dealers also provide unrelated services such as market making. Life insurers also similar to my model, though their liabilities are longer duration than banks and not money-like, so the demand for their liabilities is conceptually distinct. Key features missing from my model are capital requirements and deposit insurance, which may be important for ensuring that even agents who do not

understand an intermediary's assets can assume their liabilities are safe.[17]

1.2 Applications

Changes in the Supply and Demand for Safe Assets The model developed in the previous section can be used to understand the general equilibrium effects of changes in the supply and demand for safe assets. Because the model endogenously determines asset prices, intermediary portfolios and leverage, and the capital structure of the non-financial sector, all of these will adjust in order to clear the market for safe assets. This provides a framework for understanding how the financial system responds to a safe asset shortage, which a macroeconomic literature (e.g. Caballero Farhi 2017) argues has been a key driving force behind the low real interest rates in recent decades. My model implies that a growing demand for safe assets causes something akin to the subprime boom of the 2000s. In particular, the financial sector expands and invests in riskier assets than it previously did, which leads to an increase in the leverage of the non-financial sector due to a reduction in its cost of borrowing.

To increase the demand for safe assets, I take the comparative static of increasing $v'(d)$ for all d by one unit.[18] The effect of this is characterized by implicitly differentiating the equilibrium condition 1.18. For any x, let $\frac{\partial x}{\partial v}$ be the derivative of x with respect to increasing v' by one unit. The ratio $r = \frac{\Pr(x_i > D_i | good)}{\Pr(x_i > D_i | bad)}$ that characterizes the intermediary's portfolio satisfies

$$\overbrace{v''(d)\frac{\partial d}{\partial r}\frac{\partial r}{\partial v}}^{\text{change in } v'(d)} - \overbrace{\frac{u'\left(c_2^{good}\right)}{2}C''(e)(r-1)\frac{\partial e}{\partial r}\frac{\partial r}{\partial v}}^{\text{change in } C'(e)} - \overbrace{\frac{u'\left(c_2^{good}\right)}{2}C'(e)\frac{\partial r}{\partial v}}^{\text{direct effect of change in } r} + 1 = 0 \tag{1.19}$$

[17] Deposit insurance can be thought of as a promise by the government to pay off depositors in states of the world where the intermediary is unable to pay. In my model, this is equivalent to providing the intermediary with payoffs in the bad state of the world that allow them to increase the supply of safe assets.

[18] Formally, if $v_\lambda(d) = v(d) + \lambda d$ is a family of functions indexed by λ, I am taking the (Gateaux) derivative with respect to λ $\frac{d}{d\lambda}v_\lambda(d) = \lim_{\lambda' \to \lambda} \frac{(v(d)+\lambda' d)-(v(d)+\lambda d)}{\lambda' - \lambda}$.

where

$$\frac{\partial d}{\partial r} = \int_0^1 \Pr\left(x_i > D_i | bad\right) \frac{\partial D_i}{\partial r} di \tag{1.20}$$

$$\frac{\partial e}{\partial r} = \int_0^1 \left(\left[\Pr\left(x_i > D_i | good\right) - \Pr\left(x_i > D_i | bad\right)\right]\right) \frac{\partial D_i}{\partial r} di \tag{1.21}$$

$$\frac{\partial D_i}{\partial r} = \frac{\Pr\left(x_i > D_i | bad\right)^2}{\Pr\left(x_i > D_i | bad\right) f_{i,good}\left(D_i\right) - \Pr\left(x_i > D_i | good\right) f_{i,bad}\left(D_i\right)}. \tag{1.22}$$

The expression for $\frac{\partial D_i}{\partial r}$ comes from implicitly differentiating $\frac{\Pr(x_i > D_i | good)}{\Pr(x_i > D_i | bad)} = r$, and $\frac{\partial D_i}{\partial r} > 0$ is implied by the assumption $\frac{\partial}{\partial D_i} \frac{\Pr(x_i > D_i | good)}{\Pr(x_i > D_i | bad)} > 0$. $\frac{\partial d}{\partial r}$ and $\frac{\partial e}{\partial r}$ are therefore strictly positive. The change in the ratio $r = \frac{\Pr(x_i > D_i | good)}{\Pr(x_i > D_i | bad)}$ that parametrizes each firm's optimal leverage changes as

$$\frac{\partial r}{\partial v} = \frac{1}{\frac{u'\left(c_2^{good}\right)}{2}\left(C''\left(e\right)\left(r-1\right)\frac{\partial e}{\partial r} + C'\left(e\right)\right) - v''\left(d\right)\frac{\partial d}{\partial r}} > 0. \tag{1.23}$$

The quantity of debt issued by firm i, deposits d issued by the intermediary, and good state equity payout e from the intermediary change as

$$\frac{\partial D_i}{\partial v} = \frac{\partial D_i}{\partial r}\frac{\partial r}{\partial v} > 0, \qquad \frac{\partial d}{\partial v} = \frac{\partial d}{\partial r}\frac{\partial r}{\partial v} > 0, \qquad \frac{\partial e}{\partial v} = \frac{\partial e}{\partial r}\frac{\partial r}{\partial v} > 0. \tag{1.24}$$

The change in the safety premium $v'\left(d\right)$ equals $1 - v''\left(d\right)\frac{\partial d}{\partial r}\frac{\partial r}{\partial v}$ which satisfies

$$1 > 1 - v''\left(d\right)\frac{\partial d}{\partial r}\frac{\partial r}{\partial v} = \left(C''\left(e\right)\left(r-1\right)\frac{\partial e}{\partial r} + C'\left(e\right)\right)\frac{\partial r}{\partial v} > 0 \tag{1.25}$$

while the intermediary's willingness to pay for good state payoffs changes as

$$-\frac{1}{2}\frac{u'\left(c_2^{good}\right)}{u'\left(c_1\right)}C''\left(e\right)\frac{\partial e}{\partial v} < 0 \tag{1.26}$$

The increased safety premium and decreased value of good state payoffs to the interme-

diary implies that it is willing to pay more for sufficiently low (systematic) risk securities but less for sufficiently high risk securities. While some securities are so risky that the intermediary's willingness to pay for them decreases, the borrowing costs of all non-financial firms decrease. This can be seen from the fact that $\frac{\partial r}{\partial v} > 0$, implying that the intermediary is now willing to pay the same as the household for an asset of greater systematic risk r. Because $\frac{E_{good}\min(x_i,D_i)}{E_{bad}\min(x_i,D_i)} < r$ for all firms i, the intermediary is also willing to pay strictly more for each firm's debt, reducing each firm's cost of borrowing. This completes the proof of the following result. As noted on the section on non-financial firms, non-financial firm debt can be relabeled to represent mortgage debt, so this result also implies household mortgage borrowing would increase.

Proposition 1.5 *(safe asset demand) An increase in the demand for riskless securities, modeled as an increase in the function $v'(d)$ causes:*

1. An increase in the quantity d of riskless securities and intermediary equity issuance e.

2. A reduction in the risk-free interest rate and increase in credit spreads, with an overall reduction in borrowing costs for all firms.

3. An increase in the leverage of the non-financial sector

The second comparative static, creating a supply μ of riskless securities backed by lump sum taxes on the household, simply increases the supply of safe assets from the liability d issued by the intermediary to the sum $d + \mu$. This crowds out the intermediary's incentive to perform safety transformation by providing a supply of safe assets that do not lie on the intermediary's balance sheet. For any given quantity d of deposits, the safety premium $v'(d + \mu)$ is decreasing in μ. The effect of this decrease is therefore precisely the opposite of the increase in $v'(d)$ considered in the first comparative static. While the model in the previous section does not explicitly have government debt, riskless government debt can be mapped into the framework above by simply replacing $v'(d)$ with $v'(d + \mu)$. The calculations for the effect of an increase in the demand for safe assets therefore also imply the following. Closed form derivatives for how variables adjust are simply $-v''(d)$ times the results derived above for the increase in safe asset demand.

Proposition 1.6 *(government debt supply) An increase in the supply μ of riskless securities issued by the government causes*

1. An increase in the quantity $d + \mu$ of total riskless securities, a decrease in riskless securities d issued by the intermediary, and decrease in intermediary equity issuance e.

2. An increase in the risk-free interest rate and compression of credit spreads, with an overall increase in borrowing costs for all firms.

3. A decrease in non-financial leverage.

Quantitative Easing A third possible policy experiment is to consider the effects of quantitative easing policies, in which the government issues safe debt in order to purchase risky securities. If the government buys equities, which are held by households, the effect on asset prices, leverage, and intermediary portfolios is identical to simply increasing the supply of government debt backed by more taxes. However, the effects are more subtle when the government buys debt securities which are owned by intermediaries. Such a transaction replaces risky assets owned by the intermediary with riskless government debt and therefore can be seen as a combination of adding riskless assets to the intermediary's portfolio and removing good state payoffs. This has the effect of both increasing the supply of safe assets and decreasing the amount of risk the intermediary needs to bear.

To derive the effects of such asset purchases, I first compute the effect of removing good state payoffs from the intermediaries balance sheet. For any variable m I denote $\frac{\partial m}{\partial good}$ the change in m that occurs when good state payoffs are removed from the intermediary's portfolio.

$$\left[v''(d)\frac{\partial d}{\partial r} - \frac{u'\left(c_2^{good}\right)}{2}\left[C''(e)\frac{\partial e}{\partial r}(r-1) + C'(e)\right]\right]\frac{\partial r}{\partial good} = -\frac{u'\left(c_2^{good}\right)}{2}C''(e)(r-1) \tag{1.27}$$

$$\frac{\partial r}{\partial good} = -\frac{\frac{u'\left(c_2^{good}\right)}{2} C''(e)(r-1)}{v''(d)\frac{\partial d}{\partial r} - \frac{u'\left(c_2^{good}\right)}{2}\left[C''(e)\frac{\partial e}{\partial r}(r-1) + C'(e)\right]} > 0 \tag{1.28}$$

$$\frac{\partial d}{\partial good} = \frac{\partial d}{\partial r}\frac{\partial r}{\partial good} > 0\,, \qquad \frac{\partial e}{\partial good} = \frac{\partial e}{\partial r}\frac{\partial r}{\partial good} > 0 \tag{1.29}$$

The change in the safety premium $v'(d)$ is

$$v''(d)\frac{\partial d}{\partial r}\frac{\partial r}{\partial good} < 0 \tag{1.30}$$

.

The change in the cost of equity $\frac{u'\left(c_2^{good}\right)}{2} C'(e)$ is equal to

$$\left[v''(d)\frac{\partial d}{\partial r} - \frac{u'\left(c_2^{good}\right)}{2}\left[C''(e)\frac{\partial e}{\partial r}(r-1) + C'(e)\right]\right]\frac{\partial r}{\partial good} < 0. \tag{1.31}$$

As noted above, a purchase of risky assets owned by the intermediary financed by the issuance of riskless government debt increases the supply of riskless assets and removes good state payoffs from the intermediary's balance sheet. To compute the effects of asset purchases, I must figure out what weights to place on the effects of adding riskless assets and removing good state payoffs from the intermediary's balance sheet. An asset purchase occurs at market prices, so the assets bought and sold must have the same price. If the government issues d units of debt to buy an asset that has an (expected) payoff of δ_{good} in the good state and δ_{bad} in the bad state, this can be seen as increasing the supply of riskless assets held by the intermediary by $\mu = d - \delta_{bad}$ units while reducing the amount of good state payoffs on its balance sheet by $r_{good} = \delta_{good} - d$. A transaction at market prices must satisfy

$$\left[Eu'(c_2) + v'(d)\right]\mu = \frac{1}{2}u'\left(c_2^{good}\right)\left(1 - C'(e)\right)\left(-r_{good}\right) \tag{1.32}$$

An asset purchase removes $\frac{[Eu'(c_2)+v'(d)]}{\frac{1}{2}u'\left(c_2^{good}\right)(1-C'(e))}$ units of good state payoff from the intermediary's portfolio per unit of riskless payoff added, regardless of which asset is purchased. This adds μ units of bad state payoff to the intermediary's portfolio while removing $r_{good}-\mu$ good state payoffs. μ is a sufficient statistic for the effect of the asset purchase. For any variable v, let $\frac{\partial v}{\partial QE}$ be the change in v from purchases that increase the bad state payoff of the intermediary's portfolio by 1 unit, so $\frac{\partial v}{\partial QE}=\frac{\partial v}{\partial \mu}-\frac{[Eu'(c_2)+v'(d)]}{\frac{1}{2}u'\left(c_2^{good}\right)(1-C'(e))}\frac{\partial v}{\partial good}$. These comparative statics have the same signs for the following variables, proving the following result.

Proposition 1.7 *(asset purchases 1) Purchasing risky assets owned by the intermediary financed by the issuance of riskless government debt causes*

1. An increase in the quantity $d+\mu$ of total riskless securities, a decrease in riskless securities d issued by the intermediary, and decrease in intermediary equity e.

2. An increase in the risk-free interest rate and compression of credit spreads.

The effect on corporate leverage, however, is ambiguous.

$$\left[v''(d)\frac{\partial d}{\partial r}-\frac{u'\left(c_2^{good}\right)}{2}\left[C''(e)\frac{\partial e}{\partial r}(r-1)+C'(e)\right]\right]\frac{\partial r}{\partial QE} = \tag{1.33}$$
$$-v''(d)-\frac{[Eu'(c_2)+v'(d)]}{\frac{1}{2}u'\left(c_2^{good}\right)(1-C'(e))}\left[\frac{u'\left(c_2^{good}\right)}{2}C''(e)(r-1)\right]$$

and has the opposite sign as the right hand side of equation 1.33. If $\frac{\partial r}{\partial QE}<0$, then arguments made above imply that all firms have an increase in their borrowing costs. However, if $\frac{\partial r}{\partial QE}>0$, then firms for whom $\frac{E_{good}\min(x_i,D_i)}{E_{bad}\min(x_i,D_i)}$ is sufficiently close to r will have a decrease in borrowing costs, while firms for which this ratio is small enough face an increase in borrowing costs.

Proposition 1.8 *(asset purchases 2) If*

$$-v''(d)-\frac{[Eu'(c_2)+v'(d)]}{\frac{1}{2}u'\left(c_2^{good}\right)(1-C'(e))}\left[\frac{u'\left(c_2^{good}\right)}{2}C''(e)(r-1)\right]>0 \tag{1.34}$$

Purchasing assets owned by the intermediary financed by issuing riskless debt causes a decrease in corporate leverage and an increase in borrowing costs for all firms.

If this expression is negative, asset purchases cause an increase in leverage for all firms, an increase in borrowing costs for firms with $\frac{E_{good}\min(x_i,D_i)}{E_{bad}\min(x_i,D_i)}$ *sufficiently small, and a decrease in borrowing costs for firms with* $\frac{E_{good}\min(x_i,D_i)}{E_{bad}\min(x_i,D_i)}$ *sufficiently large.*

Nominal Rigidities and The Zero Lower Bound This section adds a binding zero lower bound on monetary policy to the model developed above into a simple framework with nominal rigidities, which is the context under which the Federal Reserve's quantitative easing policies were performed. To maintain tractability, I make the extreme assumption that goods prices are perfectly rigid, following the original liquidity trap analysis of (Krugman 1998). Given this price rigidity, I assume that the central bank sets the interest rate i_d subject to the zero lower bound constraint $i_d \geq 0$ which is motivated by the possibility that households will swap riskless bonds for cash when interest rates are negative.

Under flexible prices, the household's optimality condition for investing in riskless securities

$$u'(c_1) = (1+i_d)\left[E\left(u'(c_2) + v'(d)\right)\right] \tag{1.35}$$

determines the risk-free rate taking as given consumption (c_1, c_2) and the supply of riskless assets d. With sticky prices in the goods market at time 1, the variables at time 2 (c_2, d) and the risk-free rate i_d set by the central bank determine the amount of consumption c_1 that occurs at time 1, so long as c_1 is not greater than the supply C_1 of resources available to consume. When $c_1 < C_1$, a shortage of aggregate demand depresses output in a recession.

When interest rates are fixed at the zero lower bound $c_1 < C_1$, this first order condition implies that reducing the demand shortage at time 1 requires either $Eu'(c_2)$ or $v'(d)$ to decrease. To reduce $Eu'(c_2)$, a policy in the original zero lower bound analysis of (Krugman 1998) is to commit at time 1 to stimulating future demand by keeping interest rates below their natural level, which is called forward guidance. This is not considered in my model

since we are in an endowment economy where c_2 is held fixed for simplicity. A second policy considered here is to reduce $v'(d)$ by either increasing the supply of safe assets. A shortage of aggregate demand due to a scarcity of safe assets is termed a "safety trap" by (Caballero Farhi 2017), and the stimulating effects of reducing the scarcity of safe assets in my analysis is similar to what they show. The three comparative statics considered above all change the safety premium $v'(d)$ and therefore influence aggregate demand when the zero lower bound constrains conventional interest rate policy.

The novelty of my analysis is that it considers how changing the scarcity of safe assets leads to changes in the portfolio choices of financial intermediaries and the leverage of the financial and non-financial sectors. This allows me to understand the effects of quantitative easing on financial stability, which has worried some policymakers. Relatedly, (Stein 2012b) is a policy speech arguing that debt issuance by the non-financial sector in order to repurchase stock could possibly weaken the effects of quantitative easing, and my model's determination of non-financial capital structure allows me to speak to this concern.

The fact that the equilibrium of my model is characterized by equation 1.18 makes it quite tractable to analyze the effects of nominal rigidities, since c_1 does not appear in this expression at all. This single equation can be used to solve for all corporate capital structure decisions and the assets and liabilities of the financial intermediary and yields the same answer with and without nominal rigidities. Changes in the supply and demand for safe assets and central bank asset purchases have exactly the same effect on intermediary portfolio choices and corporate capital structure decisions whether or not nominal rigidities cause a shortage of aggregate demand at time 1. This is summarized in the following proposition. One particularly important implication is that asset purchases reduce risk taking by financial intermediaries, since the policy discussion about asset purchases has considered their financial stability implications.

Proposition 1.9 *(irrelevance of nominal rigidities for portfolios and capital structure) The leverage decisions of the intermediary and non-financial sector, the portfolio choice of the intermediary, the intermediary's marginal cost of equity $C'(e)$, and the safety premium $v'(d)$ have the same response*

to asset purchases or changes in the supply and demand for safe assets with our without nominal rigidities. The results proved above for changes in these variables continue to hold at the zero lower bound.

Because the changes in $v'(d)$ computed without nominal rigidities continue to hold, it is immediate to determine the effect of the comparative statics considered above on consumption at time 1. This is true because when $i_d = 0$ it must be the case that for any perturbation μ

$$\frac{\partial c_1}{\partial \mu} u''(c_1) = \frac{\partial u'(c_1)}{\partial \mu} = \frac{\partial v'(d)}{\partial \mu} \tag{1.36}$$

to ensure the risk-free rate remains at 0. Since $u'' < 0$, it follows that any decrease in the safety premium $v'(d)$ must also increase time 1 consumption.

Proposition 1.10 *(the safe asset premium and aggregate demand) While at the zero lower bound, increasing the demand for safe assets reduces consumption at time 1 while increases in the supply of safe assets or risky asset purchases financed by the issuance of government debt increase consumption at time 1.*

The response of asset prices to asset purchases or safe asset supply and demand changes does depend on whether there are nominal rigidities. The risk-free rate is held fixed at the zero lower bound while previously it was free to adjust and ensure the goods market at time 1 is able to clear. In addition, increasing aggregate demand at time 1 reduces the marginal utility of consuming c_1, providing an additional mechanism that boosts asset prices only in a shortage of aggregate demand. The price of an equity security paying δ_e at time 2 now changes as

$$\frac{d}{d\mu} E \frac{u'(c_2)}{u'(c_1)} \delta_e = \left(E u'(c_2)\, \delta_e\right) \frac{-1}{\left(u'(c_1)\right)^2} \frac{\partial u'(c_1)}{\partial \mu} = \left(E u'(c_2)\, \delta_e\right) \frac{-1}{\left(u'(c_1)\right)^2} \frac{\partial v'(d)}{\partial \mu} \tag{1.37}$$

while the price of a debt security paying δ_{debt} now changes as

$$\frac{\partial}{\partial \mu}\left[E_{bad}\left(\delta_{debt}\right)\left(1+i_d\right)+\frac{1}{2}\frac{u'\left(c_2^{good}\right)}{u'\left(c_1\right)}\left(1-C'\left(e\right)\right)\left(E_{good}-E_{bad}\right)\delta_{debt}\right]= \tag{1.38}$$

$$\frac{1}{2}\left[\frac{u'\left(c_2^{good}\right)}{u'\left(c_1\right)^2}\left(-\frac{\partial v'\left(d\right)}{\partial \mu}\right)\left(1-C'\left(e\right)\right)+\frac{u'\left(c_2^{good}\right)}{u'\left(c_1\right)}\left(-C''\left(e\right)\frac{\partial e}{\partial \mu}\right)\right]\left(E_{good}-E_{bad}\right)\delta_{debt}.$$

Relative to the equity market, the debt market has no change in risk-free rate but a greater proportional change in the price of systematic risk due to the additional effect of changing $C'(e)$. This is because asset purchases only effect the pricing of risk in the equity market through the indirect effect on consumption, while the pricing of risk in the bond market depends explicitly on the intermediary's cost of equity capital as well as on consumption. These calculations prove the following proposition.

Proposition 1.11 *(asset price responses at the zero lower bound) At the zero lower bound,*

(i) *An increase in the demand for safe assets reduces the prices of debt and equity securities.*

(ii) *An increase in the supply of safe assets increases the prices of debt and equity securities.*

(iii) *Purchasing risky assets financed by the issuance of riskless government debt increases the prices of debt and equity securities. The risk-free rate implied by equity prices decreases while the risk-free rate implied by bond prices remains at zero. The price of risk in both markets decreases, with a greater proportional decrease in the debt market.*

Of particular interest, asset purchases now lower the cost of borrowing for the non-financial sector, since the risk-free rate stays fixed and the price of systematic risk decreases with asset purchases. This is related to the verbal argument (Stein 2012b) gives in a policy speech that asset purchases may induce firms to borrow in order to repurchase stock as a result of their decreased borrowing cost. My model provides a particularly relevant framework for evaluating this reasoning, since unlike in existing models it is the relative cost of debt and equity financing that determines the leverage of the non-financial sector in my framework. Consistent with event study evidence (Neely 2011, Chodorow-Reich 2014), at the zero lower bound asset purchases boost both debt and equity prices, and it is the

relative cost of debt and equity financing that determines optimal capital structure. As a result, the non-financial sector may decrease its leverage despite the reduced borrowing cost, as characterized in the section without nominal rigidities. One limitation of my analysis is that it only formalizes a broad "portfolio balance" channel in which all debt securities are priced by preferences of the same intermediary, while there is some empirical evidence (Krishnamurthy Vissing-Jorgensen 2011) that segmentation between markets for individual assets plays an important role in the transmission mechanism of quantitative easing.

Conclusion This paper develops a general equilibrium model of how the financial system is organized to meet a demand for safe assets. In the model, financial intermediaries face the same financing frictions as other firms and have the same information and investment opportunities as households. The role played by intermediaries is to pool the debt of non-financial firms, who cannot issue riskless assets because of idiosyncratic risk, and issue riskless securities and a risky equity tranche backed by this debt portfolio. The debt and equity markets are endogenously segmented, and the non-financial sector's optimal capital structure arbitrages these segmented markets. While previous models of financial intermediation emphasize the illiquidity of intermediary balance sheets, this model provides a framework that can explain intermediaries' large holdings of liquid, publicly available securities. In addition, the model shows how a growing demand for safe assets causes a subprime boom and provides a framework for understanding the transmission mechanism of quantitative easing policies and their implications for financial stability.

Several features of the model suggest a future research agenda. First, the model takes as given the demand for safe, money-like assets. A more fundamental framework where the demand for money and the role of intermediaries as creators of money are both endogenous may provide additional insights. Second, existing safe assets are typically denominated in a currency. A framework with safe assets in multiple currencies may be useful for understanding the international spillovers of quantitative easing and the role of the dollar in the international financial system. The perspective taken in this model, where the demand for liabilities issued by intermediaries determines their asset portfolio, may be a useful and

tractable framework for many questions about the role of intermediaries in macroeconomics and finance.

Chapter 2

Money, Prices, and Financial Fragility

Bank deposits are sometimes referred to as a form of privately created money. In order to pay for goods, it is common to make a withdrawal from a bank account in order to pay a price denominated in a form of currency. To facilitate these transactions, bank deposits themselves are also denominated in currency, so that deposits and currency can be easily exchanged for each other without fluctuations in the exchange rate. For a bank to issue deposits that can readily be exchanged for currency, it must hold enough currency and other assets to be able to meet any withdrawals it is likely to face. If depositors are wary of a bank's ability to meet withdrawals, deposits may no longer function as a money-like asset, and a run on the bank may occur.

While historically money and banking were considered a unified topic, the literatures that study the two topics have grown separately and do not speak to their interaction. While the need for money whose value is easily verified to perform transactions has been studied (Wright Williamson 1994, Banerjee Maskin 1996), and a separate literature emphasizes the demand for informationally insensitive bank deposits (Gorton Pennachi 1993, Dang Gorton Holmstrom Ordonez 2017), the ideas in these literatures have had essentially no interaction. As a result, no existing framework is able to pose the question of how the health of the banking system is related to the supply of money, and how this resulting supply of money determines the price level. Given that this story is at the heart of the monetarist explanation of the great depression (Friedman and Schwartz 1963), such a framework may be a useful tool for understanding how the price stability and financial stability mandates of central banks relate to each other. In addition, reserve requirements are often used as a tool to implement monetary policy (historically in many countries, and currently in China), and the current cashless New Keynesian framework (Woodford 2003) cannot be used to understand this.

This paper makes several contributions. First, it provides a simplification of existing models of asymmetric information in monetary transactions, in which the role of matching, bargaining, and heterogenous preferences are abstracted away. As a result, it can isolate the effects of asymmetric information about the quality of goods in transactions, replicating

findings in the existing literature more transparently. In particular, the insight of (Banerjee and Maskin 1996) that goods of most easily verifiable quality are overproduced are replicated in a more tractable framework. This simplification is particularly useful for this paper's main goal of building a model of banking on top of existing models of money and may provide a use intermediate input for other papers that want to integrate models of money into other literatures. A particularly tractable special case occurs using the technique of (Lester Postlewaite Wright 2012) where all but one good "gold" can be costlessly counterfeited. Because the model features competitive markets with identical consumers rather than search and matching, the model provides a transparent way of showing how the price level of goods is determined by the money supply and how this leads to a pecuniary externality where money is overproduced.

Second, the papers shows how the price level and money supply are determined when agents are unsure of when they need to consume and hold a precautionary liquidity buffer. Similar to (Diamond Dybvig 1983), I assume that there are "early" and "late" periods in the model, and consumers obtain uncontractible private information about when they want to consume. In this setting, consumers who want to consume in the late period of the model must hold idle balances of gold in the early period that could otherwise have been spent on consumption. In addition, consumers who turn out to want to consume in the early period of the model end up consuming their gold at the end of the early period since they have no incentive to save. Both of these channels reduce the quantity of gold available for performing transactions. As a result, a lower price level is required to clear the market with only a fraction of existing gold being used to finance all goods market transactions. This deflation in turn increases the private incentive to inefficiently overproduce gold. This draws a connection between frictions commonly studied in the banking literature following (Diamond Dybvig 1983) and the connection between the money supply and inflation studied in the literature following (Lagos Wright 2005).

Third, the model shows how a bank that issues deposits that can be exchanged for gold on demand is able to ensure that gold is held precisely by those agents who need it to

consume. The bank is able to do this by holding a portfolio composed both of liquid gold and other illiquid goods, ensuring that its gold reserves are large enough to finance all transactions in the goods market. In this sense, the bank is performing fractional reserve banking, so that the "inside money" supply created by the bank is strictly larger than the base "outside money" supply in which goods transactions are performed. However, because the bank does not have 100% reserves, if all agents choose to withdraw from the bank at the same time, it will be forced into a costly liquidation. In this sense, the bank is able to increase the effective money supply if and only if it chooses to expose itself to costly runs. Similar to (Calvo 1988), the bank's expectations of the probability of a run determines how much it chooses to expand the money supply.

While the model is primarily about an older banking system where transactions were settled in gold, it also has some implications for a more modern setting. Just as it is costly to physically mine gold, and the private benefits of producing gold is influenced by its role in the monetary system, backing riskless assets such as treasury bills which function as money requires government fiscal capacity. Because the price level is a function of the supply of money-like assets, private money creating agents do not internalize the effects of their money production on the aggregate economy. Just as banks used to hold fractional reserves of gold to back the deposits they issue, in the modern world only a fraction of bank assets are reserves, treasuries, or other highly liquid assets. As a result, the liquidity creation by modern banks is conceptually akin to this model.

2.1 A model of information frictions and media of exchange

This section presents a model of a goods market where trade is composed of anonymous, bilateral transactions in a competitive market. These transactions are composed of a "buyer" and a "seller", where the quality of goods sold by the seller are public information while the goods used as payment by the buyer are private information. As a result, goods traded by buyers are subject to a lemons problem, in which only the lowest quality goods are used by buyers to trade. There are many varieties of goods with differing distributions of quality,

and sellers will accept each variety only at an exchange rate that is consistent with the lemons problem they face. As a result, there is a "scarcity" of goods that can be used by buyers, in the sense that one unit of quality offered by a buyer purchases strictly more than one unit of quality sold by a seller. Across all transactions, the amount of quality offered by the seller per unit of quality provided by a buyer must be the same, and this "terms of trade" parameter characterizes the severity of the price distortion induced by asymmetric information.

The model has two periods $t = 0, 1$. There is a continuum of identical consumers indexed by $i \in [0, 1]$. At time 0, goods are produced and exchanged which are then consumed at time 1. A set of possible varieties of goods indexed by $v \in [0, 1]$ may be produced, and each good has a "quality level" $q \in [0, \infty)$. For each good variety v, the distribution of quality levels of goods of varity g is $m_v(q)$, which has density $f_v(q)$. A consumer who consumes $c(v, q)$ of quality v goods of variety v at time 1 obtains utility

$$\int_0^1 \int_0^\infty qc(v, q)\, dqdv.$$

where the integral sums across the quality levels q of each variety v that is consumed and the different varieties $v \in [0, 1]$. The amount of utility derived from consuming a good is proportional to its quality q. However, consumers are unable to consume goods which they produce themselves, providing a reason for them to exchange their output for the output of others.

The goods produced by each consumer may either be "sold at a shop" or taken as a "means of payment" to purchase items at the shops of other consumers. The quality of goods sold in shops are observable, while the quality of goods used as means of payment are privately known by the consumer making the purchase. This assumption captures the idea that when consumers are shopping for something they want, they understand their own preferences, but that shops who sell goods know nothing about what consumers will offer them to pay for their purchases. This stark assumption provides a clean way to ensure

that media of exchange are subject to an asymmetric information problem, and the most "money-like" goods are those for which this problem are the least severe. Such a problem could perhaps be microfounded with endogenous information aquisition, where consumers are willing to learn the quality of goods they will themselves consume, but producers will not be willing to learn about their many different customers.

Each consumer faces market prices $p_{v_{payment}}(v_{sold}, q_{sold})$determining how many goods of variety $v_{payment}$ can be obtained by one selling unit of good (v_{sold}, q_{sold}). The consumer also chooses which of its goods it will use to buy goods in the shops of other consumers. The prices $p_{v_{payment}}(v_{sold}, q_{sold})$ cannot depend directly on the quality $q_{payment}$ of a good offered as payment, since the quality of goods used as means of payment is private information. If sellers refuse goods of variety v payment, this can be thought of as a price ∞.

Given the market prices $p_{v_{payment}}(v_{sold}, q_{sold})$, the consumer's shopping problem is to choose which goods (v_{sold}, q_{sold}) to buy with any variety $v_{payment}$ it uses as a means of payment. It obtains $\frac{q_{sold}}{p_{v_{payment}}(v_{sold}, q_{sold})}$ of utility by buying $p_{v_{payment}}(v_{sold}, q_{sold})$ with one unit of a good of variety $v_{payment}$, so it only buys goods for which $\frac{q_{sold}}{p_{v_{payment}}(v_{sold}, q_{sold})}$ attains its maximum. Call this expression $q_{exchange}(v_{payment})$, which is the maximum number of units of quality that can be obtained by shopping with one unit of a good of variety $v_{payment}$. The consumer's optimal shopping problem is solved if and only if it buys goods with means of payment $g_{payment}$ such that

$$\frac{q_{sold}}{p_{v_{payment}}(v_{sold}, q_{sold})} = q_{exchange}(v_{payment}). \tag{2.1}$$

In addition to its shopping problem, the consumer has to decide which goods it produces will be used to buy other goods and which will be sold. If a consumer chooses to use a good (v, q) to buy other goods, it obtains utility $q_{exchange}(v)$ for doing so when it shops optimally. It will therefore choose to sell goods (v, q) for which it can obtain weakly more utility than $q_{exchange}(v)$ in the sale. If the good is sold for a means of payment of variety v, $\frac{q}{q_{exchange}(v_{payment})}$ units of good $v_{payment}$ will be obtained from the sale. The utility obtained

from this sale is therefore equal to

$$\frac{q_{sold}}{q_{exchange}\left(v_{payment}\right)}\frac{\int q_{payment}d\mu_{v_{payment}}\left(q_{payment}\right)}{\int d\mu_{v_{payment}}\left(q_{payment}\right)},$$

where $\mu_{v_{payment}}(.)$ is the quality distribution of goods of variety $v_{payment}$ used as a means of payment. This expression is the product of the quantity of goods of variety $v_{payment}$ paid in the transaction, $\frac{q_{sold}}{q_{exchange}\left(v_{payment}\right)}$, times the average quality $\frac{\int q_{payment}d\mu_{v_{payment}}\left(q_{payment}\right)}{\int d\mu_{v_{payment}}\left(q_{payment}\right)}$ of goods of variety $v_{payment}$ used as a medium of exchange. Note that this expression depends only on the good (v_{sold}, q_{sold}) through its quality level q. The maximum of this expression over all varieties $g_{payment}$ defines a linear function

$$q_{sales}(q) = q \max_{v_{payment}} \frac{\int q_{payment}d\mu_{v_{payment}}\left(q_{payment}\right)}{q_{exchange}\left(v_{payment}\right)\int d\mu_{v_{payment}}\left(q_{payment}\right)} = Tq, \tag{2.2}$$

which is the highest amount of utility that can be obtained by choosing to sell a good of quality q. The constant T is the "terms of trade" between units of quality on the buy and sell sides of the market. Sellers are only willing to accept goods for which they obtain T units of quality in means of payment per unit of quality sold. The constant $P = \frac{1}{T}$ can be interpeted as a "price level", so that one unit of quality on the buyer's side purchases P units of quality on the seller's side.

Since the returns to selling a good is strictly increasing in its quality, while the returns to buying a good with it does not depend on its quality, it follows that every consumer sells its highest quality goods of each variety and uses its lowest quality goods to buy goods in the shops of other consumers. For each variety g, the consumer's problem of determining what to buy in sell is determined by a cutoff quality $q^*(v)$ for which

$$q_{sales}\left(q^*(v)\right) = q_{exchange}(v). \tag{2.3}$$

All goods (v, q) with $q > q^*(v)$ are optimally sold, while those with $q < q^*(v)$ are optimally used as a means of payment. This formalizes the idea that "bad money drives out good", in that among a set of indistinguishable goods, the least valuable are the ones that circulate as a medium of exchange.

Given these properties of the consumer's optimal sales and shopping decisions, it is now straightforward to characterize the models equilibrium. First, the set of goods sold can be characterized by a cutoff $q^*(v)$ such that for each v, a good (v,q) is sold only if $q \leq q^*(v)$. Second, the price $p_{v_{payment}}(v_{sold}, q_{sold})$ of every good sold must satisfy

$$p_{v_{payment}}(v_{sold}, q_{sold}) = q_{sold} T \frac{\int_0^{q^*(v)} q dm_{v_{payment}}(q)}{\int_0^{q^*(v)} dm_{v_{payment}}(q)}, \tag{2.4}$$

so that the sale yields T units of expected quality per unit of quality sold. Recall that $m_v(.)$ is the unconditional quality distribution of goods of variety v, regardless of whether these goods are bought or sold. Third, the cutoff $q^*(v)$ must be chosen optimally, so that $q_{sales}(q^*(g)) = q_{exchange}(g)$, which can be written as

$$Tq^*(v) = \frac{1}{T} \frac{\int_0^{q^*(v)} q dm_v(q)}{\int_0^{q^*(v)} dm_v(q)}. \tag{2.5}$$

Finally, for the goods market to clear, the value of goods sold must equal the value of goods used as a medium of exchange

$$\int_0^1 \left[\int_0^{q^*(v)} q dm_v(q) \right] dv = \frac{1}{T} \int_0^1 \left(\int_{q^*(v)}^{\infty} q dm_v(q) \right) dv. \tag{2.6}$$

If the distribution of quality of goods of variety g μ_g is nonatomic for all g, it immediately follows that there is a unique equilibrium. For each g, the $q^*(v)$ which solves 2.5 is monotone and continuous. Plugging this value of $q^*(v)$ into equation 2.6 yields a monotone, continuous function which crosses zero once. Moreover, this point of crossing must be at some point $T < 1$ if each μ_g is nonatomic. The following proposition summarizes the results shown so far.

Proposition 2.1 *A unique equilibrium of the model exists. In this equilibrium*

(i) For each variety of goods g there is a quality cutoff $q^*(v)$ such that high quality goods with $q > q^*(v)$ are sold while low quality goods with $q < q^*(g)$ are used as a means of payment.

(ii) The price at which every good is sold is determined by 2.4, so that the average quality of goods $\frac{\int_0^{q^*(v)} q dm_v(q)}{\int_0^{q^*(v)} dm_v(q)}$

of variety v used as a means of payment determines the quality q of goods that can be purchased with variety v.

(iii) Every good variety v can be used to purchase goods of quality $\frac{1}{T}\frac{\int_0^{q^*(v)} q dm_v(q)}{\int_0^{q^*(v)} dm_v(q)}$*, where the average quality of v goods used as a means of payment is multiplied by a terms of trade parameter* $T < 1$ *that is common across all transactions.*

(iv) The equilibrum is charecterized by the unique solution for T *and* $q^*(g)$ *for each* g *to 2.5 and 2.6*

A key property of the equilibrium is that is characterized entirely by the terms of trade T or equivalently price level $P = \frac{1}{T}$. A value of $T < 1$ (or $P > 1$) implies that there is a "liquidity premium", in the sense that assets which are easily verifiable media of exchange buy goods which are strictly more desirable to consume than they themselves are. This premium is large when the supply of "money-like" goods is scarce. If a premium of $T = 1$ occurred in equilibrium (which can happen when there is a sufficient supply of goods which are not subject to a lemons problem when used as a means of payment), the equilibrium is equivalent to one in which there is no information friction in bilateral transactions. $T < 1$ can never happen, since once $T = 1$ is attained, money-like goods will optimally be sold as well as used as a means of payment to ensure the market still clears.

With endogenous production, the liquidity premium $T < 1$ distorts production decisions away from the social optimum. Suppose it requires $l(v, q, C)$ units of disutility to create C units of good (v, q). Since all goods produced are eventually consumed, and utility is determined by total consumption, the social value of a good produced is precisely its value to consume. It follows that the optimal production decision must satisfy $\frac{\partial}{\partial C} l(v, q, C) = q$, since q units of labor disutility are spent to produce q units of consumption utility. However, the privately optimal production decisions satisfy

$\frac{\partial}{\partial C} l(v, q, C) = Tq$ for goods with $q \geq q^*(v)$ and $\frac{\partial}{\partial C} l(v, q, C) = \frac{1}{T}\frac{\int_0^{q^*(v)} q dm_v(q)}{\int_0^{q^*(v)} dm(q)}$ for goods with $q \leq q^*(v)$. The market underproduces goods high quality goods which must be sold

to an informed buyer and overproduces low quality goods that can be used as a means of payment. This can be viewed as a pecuniary externality where agents take the price level P as given when choosing how much "money" to produce, while the supply of "money" relative to the quantity of goods sold determines the overall price level. The private benefits of money production are (in part) dissipated socially because they cause inflation.

2.2 A special case with a unique medium of exchange

The above equilibrium, in general, has the low quality goods of many different varieties being used as media of exchange. To build a framework which is sufficiently tractable to add a financial sector, it is desirable for there to be a single good that circulates as a means of payment. In order for this to be the case, I will assume that there is a single good "gold" indexed by g_{gold} which can only be produced at a quality level $q = 1$ (and therefore is mechanically not subject to any asymmetric information problem about its quality). For all other goods g, I will assume that zero quality goods of variety g can be costlessly produced. As a result if any non-gold good is accepted as a means of payment, infinite quantities of worthless zero quality goods will be produced to exploit the good's purchasing power. It therefore follows that gold is the only asset which can be accepted as a means of payment in a transaction. This special case, adapted from (Lester Postlewaite Wright 2012), can be viewed as a microfoundation for a cash in advance constraint, where only a good whose quality is verifiable functions as cash.

In this special case, equilibrium is particularly tractable. Given a supply G of quality units of gold, and a total quality Q of non-gold goods produced, the price level P must satisfy the following expression reminiscient of the quantity theory of money

$$PQ = G. \tag{2.7}$$

in order for the goods market to clear.

This expression implies that the model behaves quite similarly to cash-in-advance models,

except for the fact that the "money supply" G is privately produced. Holding fixed Q, a doubling in the quantity of gold G leads to a double in the price level. However, as the price level adjusts, the private incentives to produce gold relative to other goods that cannot be used in transactions adjusts. If there is a disutility cost $c_g(G)$ of producing gold and $c_Q(Q)$ of producing other non-gold goods, then in equilibrium the supply of gold and other goods satisfies

$$\begin{aligned} c_g'(G) &= P \\ c_Q'(Q) &= \frac{1}{P}. \end{aligned} \tag{2.8}$$

This yields a unique equilibrium with $P > 1$ if $c_g(G)$ and $c_Q(Q)$ are increasing, strictly convex, $c_g' > c_Q'$ pointwise, and both c_g' c_Q' have full support on the positive real numbers.

However, since the utility attained from consumption is simply $Q + G$, optimal production decisions must satisfy

$$c_g'(G) = c_Q'(Q) = 1. \tag{2.9}$$

The market equilibrium therefore can only implement the socially optimal production decisions only at $P = 1$. With $P > 1$, gold is overproduced while other goods are underproduced. To implement the efficient allocation, a social planner can levy a proportional tax of rate $1 - P$ on gold and provide a subsidy of rate $\frac{1}{P} - 1$ on the production of other goods. This tax overcomes the private incentives to overproduce money, whose price is distorted by its liquidity value in exchange. These results are summarized in the following proposition:

Proposition 2.2 *Under the above regularity conditions on the production cost functions c_g and c_Q,*

(i) There exists a unique equilibrium under which 2.7 and 2.8 are satisfied.

(ii) In this equilibrium, a price level $P > 1$ is obtained, so one quality unit of gold buys strictly

more than one unit of other goods. In this sense, gold is "scarce" and earns a "liquidity premium."

(iii) *Whenever there is a liquidity premium $P > 1$, the privately optimal production decisions lead to too much gold and too few non-gold goods being produced relative to the social optimum.*

(iv) *A social planner who can levy proportional taxes on gold production and proportional subsidies for non-gold production can implement the social optimum.*

2.3 Uncertain liquidity needs, and inefficient consumption of money-like goods

The analysis so far has demonstrated how information frictions determines what properties goods have to endogenously circulate as a form of money. However, there has been no role for financial intermediaries to create money-like assets. In order to introduce banks, I first add frictions similar to (Diamond Dybvig 1983), in which consumers are uncertain of their future consumption needs. As a result, they want to hold a precautionary buffer of money in case they want to consume. In the absence of intermediaries, this leads to an inefficient hoarding of money like assets, which are then overproduced in equilibrium. Intermediaries, who offer gold on demand, improve the allocation by allowing consumers to withdraw gold precisely when they need to consume, reducing the incentives to produce more gold than is socially efficient.

The model now has three time periods $t = 0, 1, 2$. At time 1, a fraction μ of consumers realize that they only demand consumption at time 1, while a fraction $(1 - \mu)$ realize they only demand consumption at time 2. The market for goods at $t = 1$ and $t = 2$ is precisely the economy with private information in exchange that is analyzed above. In addition, trees which produce goods at time 2 may also be sold at time 1, and are subject to the same information frictions that require using gold as a means of payment. At time 2, a supply G_2 of gold and non-gold goods Q_2 are available. At time 1, there is additionally a supply Q_1 of goods available to consume at time 1 and a total supply G_1 of gold. At time 0, a supply (Q_1, Q_2, G) of gold and non-gold goods are produced, required disutility

$l_1(Q_1) + l_2(Q_2) + l_G(G)$ to produce.

The analysis of the equilibrium at $t = 2$ is straightforward. Given the supply (Q_2, G_2), the price level at time 2 P_2 satisfies

$$G_2 = P_2 Q_2 \tag{2.10}$$

As a result, each late consumer who prefers $t = 2$ consumption gets P unit of utility from owning goods and $\frac{1}{P}$ units of utility from owning gold.

Given the equilibrium at $t = 2$, the equilibrium at time $t = 1$ is similar to that above. Because early and late consumers did not know their identities when choosing a portfolio at $t = 0$, they enter period 1 with the same portfolios (Q_1, Q_2, G). Given this portfolio, early consumers will spend all of their gold on buying goods Q_1, while late consumers will invest their gold in buying goods Q_2. In addition, early consumers will sell their supply of period 2 output μQ_2. Market clearing therefore requires

$$\begin{aligned} \mu G &= P_1 Q_1 \\ (1-\mu) G &= P_a \mu Q_2 \end{aligned} \tag{2.11}$$

where P_1 is the price of goods at time 1 and P_a is the price of assets that yield 1 unit of goods at time 2. After this asset market exchange, early consumers will have no desire to save for the future and will therefore consume their gold holdings. They will have owned a proportional share μQ_1 of the period 1 goods sold plus μQ_2 of the period 2 goods, for which . They therefore consume $\left(\mu^2 + (1-\mu)\right) G$ of gold at the end of the period. It follows that $G_2 = \left(\mu - \mu^2\right) G$.

To complete the analysis at $t = 0$, it is necessary to determine the utility a consume gets from owning a portfolio, which can then be used to compute optimal production decisions. A portfolio (Q_1, Q_2, G) provides utility $PQ_1 + \frac{G}{P} + P_a Q_2$ to an early consumer, since it can trade its early goods for PQ_1 units of gold, its late goods for $P_a Q_2$ units of gold, and its

gold for $\frac{G}{P}$ units of early goods. This same portfolio provides utility $\frac{P_2 G}{P_a} + \frac{P_a}{P_2} Q_2 + \frac{P}{P_2} Q_1$ to the late consumer. The expected utility to the consumer of this portfolio at $t = 0$ is therefore $\mu \left(P Q_1 + \frac{G}{P} + P_a Q_2 \right) + (1 - \mu) \left(\frac{P_2 G}{P_a} + \frac{1}{P_2} Q_2 + \frac{P}{P_2} Q_1 \right)$. The optimal portfolio choices therefore satisfy

$$\begin{aligned} \mu P_1 + (1 - \mu) \frac{P_2}{P_a} &= l_1' (Q_1) \\ \mu P_a + (1 - \mu) \frac{1}{P_2} Q_2 &= l_2' (Q_2) \\ \frac{\mu}{P_1} + \frac{P_2 (1 - \mu)}{P_a} &= l_G' (G) . \end{aligned} \tag{2.12}$$

A unique equilibrium exists if each production technology has decreasing returns to scale and satisfies Inada conditions (follows from the Gale-Nikaido global inverse function theorem).

One important aspect of this equilibrium is that at after transactions are performed in period 1, early consumers own a quantity of gold. These early consumers then consume the gold since they get no utility from consuming in period 2, leading to a decrease in the total gold supply from G to $G_2 < G$. This reduction in the gold supply in turn reduces the price level in period 2. This reduced price level implies that gold buys an increased quantity of consumption goods, thereby increasing the incentives to produce gold. As analyzed above, the equilibrium in inefficient because gold is over produced, while other goods are underproduced. If there were a way to more efficiently allocate resources, so that only consumption goods Q_1 would be consumed by early consumers, the incentives to overproduce gold would be mitigated.

In addition, while the gold supply G produced at $t = 0$ is entirely available at $t = 1$, only a fraction μG of it ends up in the hands of early consumers who want to consume. If the fraction of gold used to buy goods could be increased, the resulting inflation would mitigate the incentive to produce more gold than is socially desirable. The following proposition summarizes the results in this section.

Proposition 2.3 *When consumers are unsure at $t = 0$ whether they will want to consume early at*

$t = 1$ or late at $t = 2$ (where μ consume early and $1 - \mu$ consume late)

(i) Late consumers end up holding gold at $t = 1$, so the price of goods P_1 is determined by 2.11 and is strictly lower than it would be if consumers knew when they would want to consume.

(ii) Early consumers have no desire to save from $t = 1$ to $t = 2$ and end up consuming gold instead. This reduces the gold supply at time 2 to a fraction $(\mu - \mu^2)$ of the initial gold supply G. The P_2 at time 2 given by 2.10 is therefore also a fraction $\mu - \mu^2$ of what it would be with no gold consumption at $t = 1$.

(iii) As a result of the deflation of goods prices at $t = 1$ and $t = 2$, at $t = 0$ producers produce more gold and less non-gold goods than they would have in an alternative setting where consumers knew in which period they would consume. This reduces social welfare, since gold is overproduced and other goods underproduced in equilibrium.

2.4 Financial intermediaries as creators of money like assets

This section shows how introducing a bank, that purchases assets and issues demandable claims to provide gold, is welfare enhancing. In addition to the real assets that exist in the economy, a "bank" is able to buy portfolios of assets and issue liabilities backed by the portfolio. In the best possible equilibrium, adding a bank to the framework in section 3 is able to replicate the allocation that would occur where consumers knew at $t = 0$ whether they would want to consume at $t = 1$ or $t = 2$. The bank effectively increases the money supply by issuing claims demandable for gold which are only backed by a fraction of gold reserves. This exposes the bank to runs, since only gold is liquid, and the bank is able to increase the money supply if and only if a bad equilibrium with a run exists.

As in section 3, the model now runs from $t = 0, 1, 2$ and is identical to the section above except for the role of banks. At time 0, banks now can be created which purchase assets and issue liabilities in a competitive market. These liabilities of intermediaries are allowed to be demandable, in the sense that at time 1, investors may choose between a payoff at time 1 or a promise of a different payoff at time 2. Because investors learn at time 1 when they want to consume, this allows them to consume precisely in states of the world when

they desire it. These payoffs must be denominated in gold, since for the reasons presented in section 2 promises of any other good will be met by creating a worthless counterfeit. However, demandable claims can only be met by the assets an intermediary owns. If the intermediary is unable to meet a payment, a fraction $0 \leq l \leq 1$ of its assets are destroyed in "bankruptcy". Once investors choose at time 1 whether to exercise their demandable claims, the economy at times 1 and 2 is identical to that in section 3.

The model is most tractably solved backwards. At $t = 2$, all demandable claims have been exercised, so the analysis is identical to that in section 3. The price level is mechanically

$$Q_2 P_2 = G_2$$

since consumers will own all gold to perform transactions.

At $t = 1$, the intermediary owns $\left(Q^I_{1,2}, Q^I_{1,1}, G^I_1\right)$. It has issued D demandable claims which provide 1 unit of gold at time 1 or $1 + R$ units of gold at time 2.

It is only able to meet withdrawals $W \leq G^I_1$ without going bankrupt, in which case a fraction $l = 1$ of any tree it owns is lost. Because deposits cannot be consumed, every early consumer withdraws. For a late consumer, its marginal utility from successfully withdrawing gold is 1. It therefore does not withdraw whenever it gets a greater return than 1 from keeping money in the bank. The bank never sells its long dated goods (if P_a is not 1). It uses its remaining gold to buy long dated goods. It therefore has assets worth $Q^I_{1,2} + PQ^I_{1,1} + P_a\left(G^I_1 - W\right)$. If this is weakly greater than $(1 - \mu) D$, then there is an equilibrium where late consumers do not withdraw.

The equilibrium is most straightforwardly analyzed backwards, first deriving what happens after consumers have chosen whether or not to withdraw from the bank. In this case, early consumers have a portfolio $\left(Q^{early}_1, Q^{early}_2, G^{early}\right)$ while late consumers (plus banks, who act to maximize the value of their payoff at time 2) have a portfolio $\left(Q^{late}_1, Q^{late}_2, G^{late}\right)$. The price of goods at time 1 must satisfy

$$G^{early} = P_1\left(Q^{early}_1 + Q^{late}_1\right).$$

while if the return on holding long dated goods dominates gold we have a price P_a for long dated assets

$$G^{late} = P_a Q_2^{early}$$

. In this setting, the marginal return to gold for an early consumer is P, while for a late consumer it is $\frac{P_a}{P_2}$.

Given this, I now determine at time 1 which consumers will choose to withdraw from the bank. Suppose consumers enter time 1 with a portfolio (Q_1, Q_2, G, D)

where D is a quantity of deposits redeemable for gold. All early consumers will always withdraw, since long dated consumption has no value. If the bank has a portfolio $\left(Q_1^{bank}, Q_2^{bank}, G^{bank}\right)$, it follows that $\left(Q_1^{bank}, Q_2^{bank}, G^{bank} - \mu D\right)$ remains if only early consumers withdraw. This yields a payoff of $PQ_1^{bank} + Q_2^{bank} + P_a\left(G^{bank} - \mu D\right)$. If $\frac{PQ_1^{bank} + Q_2^{bank} + P_a\left(G^{bank} - \mu D\right)}{(1-\mu)D} > P_a Q_2^{early}$, then there is an equilibrium where only early consumers withdraw. If however, all bank assets are destroyed in the case of a run, then running is an equilibrium as long as $1 > \frac{G_{bank}}{D} \geq 0$, since this implies a nonnegative return to running and zero return to not running.

Under the no run equilibrium, a deposit is worth P to an early consumer and $\frac{PQ_1^{bank} + Q_2^{bank} + P_a\left(G^{bank} - \mu D\right)}{(1-\mu)D}$ to a late consumer. Under the run equilibrium, a deposit is worth $P\frac{G_{bank}}{D}$ to an early consumer and $P_a\frac{G_{bank}}{D}$ to a late consumer. As a result, if a bank expects a run, it should choose a portfolio with large gold reserves and no illiquid assets. Conversely, if the bank does not expect a run, it need only hold enough gold to meet the withdrawals of the early consumer.

At $t = 0$, intermediaries choose their assets based on their expectations about what equilibrium will occur in periods 1 and 2. Because their are multiple equilibria, the intermediary's portfolio choice is a function of whether or not it believes there will be a run, and these expectations will in equilibrium be consistent with future outcomes.

Assuming the no run equilibrium is played, the constrained best allocation can be

implemented. If a bank buys all of the gold and all goods (which by market competition requires precisely all of the resources it raises by issuing deposits), then all gold can be withdrawn by the early consumer at time 1. Since the early consumer owns only gold now and no long dated goods, no long dated goods need to be sold to clear the market. The bank will then earn all the gold from selling early goods to the consumer (because this must be true to clear the market), which can then be withdrawn by the late consumer at time 2.

To see this formally, suppose the bank's portfolio is (Q_1, Q_2, G), and consumers invest only in deposits, which are sold in quantity $D = \frac{G}{\mu}$. If only early consumers withdraw at time 1, then precisely G is withdrawn, so the price of goods satisfies $G = PQ_1$. After period 1, the bank has (Q_2, G) on its balance sheet.

Since no gold is used to buy long dated assets, we must have $P_a = 1$. It follows that no run is an equilibrium at t=1 if $\frac{PQ_1^{bank} + Q_2^{bank} + P_a\left(G^{bank} - \mu D\right)}{(1-\mu)D} = \frac{G+Q_2}{(1-\mu)\frac{G}{\mu}} > 1$.

To show that (Q_1, Q_2, G) is the optimal bank portfolio, note that any other portfolio that could be bought either implements the same allocation or requires gold to be hoarded or consumed inefficiently, and therefore will provide less utility at the same market prices.

Under this equilibrium, gold buys strictly fewer real goods than in a setting where early consumers have less gold at time 1. The production of real resources now satisfies

$$\begin{aligned} P_1 &= l_1'(Q_1) \\ P_2 &= l_2'(Q_2) \\ \frac{1}{P_1} &= l_G'(G) \end{aligned}$$

which yields strictly more welfare than the case with no intermediaries. In effect, the money creation performed by intermediaries reduces the distortions of the role gold plays as money in production decisions.

Assuming the run equilibrium is played, an outcome with no intermediation is the unique equilibrium. Because everyone will withdraw at time 1, (which is always an equilibrium as shown above), the bank must invest all deposits in gold. This is formally

identical to people holding gold directly instead of deposits. The following proposition summarizies the results in this section.

Proposition 2.4 *Introducing a bank to the framework in section 3 leads to 2 equilibria*

(i) *One good equilibrium where all gold is held by early consumers at* $t = 1$ *and by late consumers at* $t = 2$ *and no gold is consumed at* $t = 1$*. This implements the same allocation as if consumers knew their type at* $t = 0$*. Goods prices are strictly higher than in section 3, so there is less overproduction of gold.*

(ii) *A run equilibrium, where both early and late consumers withdraw gold at* $t = 1$ *and force the bank to inefficiently liquidate.*

(iii) *The best allocation is only implemented when the good equilibrium is expected at* $t = 0$*, and if the bad equilibrium is expected banks will create no money at all.*

An interesting path going forward in this model is to study fundamental driven runs as in (Goldstein Pauzner 2005) to figure out the optimal reserve holdings of a bank that creates money.

Conclusion

This paper presents a model where money-like informationally insensitive goods are required to make anonymous bilateral transactions and shows how fractional reserve banking increases the effective supply of money. By integrating the financial intermediation literature that studies the connections between bank's assets and liabilities with the monetary economics literature studying the connection between the money supply and prices, the model draws novel connections between the health of the financial sector and the price level in the economy. Only a bank that exposes itself to the risk of a run is able to create money. In future work, tieing the risks of bank runs to the fundamentals of the financial system will provide a framework for understanding the connections between the price stability and financial stability mandates of central banks.

Chapter 3

Latent Indices in Assortative Matching Models [1]

[1] Co-authored with Nikhil Agarwal

Assortative matching along a variety of dimensions has been well documented in many matching markets. There has been growing interest in estimating the underlying preferences that generate these patterns.[2] This is an important step for quantitatively evaluating economic questions such as equilibrium effects of policy interventions or changes in market structure. However, a researcher often has access only to data on matches instead of direct information on preferences and only on a limited set of characteristics. Unobserved characteristics result in deviations from the central assortative tendency observed in the data, and they can be important in understanding the distribution of preferences.

We study the identification and estimation of preferences in a large matching market in which the attractiveness of agents to the other side of the market can be summarized using a single dimensional index that aggregates an unobserved characteristic and multidimensional observed characteristics. We assume that the matching is positive assortative along this latent index. The positive assortative match is the unique pairwise stable match if utility is non-transferable, but also if utility is transferable and the total surplus is supermodular. While the single index model is canonical in the theoretical literature (see Becker, 1973), it is clearly restrictive as it rules out heterogeneity in preferences. At the cost of this restriction, compared to the large body of empirical work following Choo and Siow (2006), we present a non-parametric approach to identification that not only allows for unobserved agent characteristics that are valued by the other side,[3] but that is also agnostic about whether utility is transferable. This single-index assumption has been useful in empirical analyses of the marriage market (see Chiappori *et al.*, 2009, 2012, for example). The model may also provide an approximation in labor or education markets in which workers or students are primarily differentiated by skill and firms or colleges are primarily differentiated by quality. Further, the insights and results from our analysis have proven useful to empirical approaches in related models (see Agarwal, 2015; Vissing, 2016; Jiang, 2016; Agarwal, 2017).

[2]See Fox (2009) and Chiappori and Salanié (2015) for surveys.

[3]Choo and Siow (2006) assume that the (pre-transfer) utility of agent i for partner j is given by $u_{ij} = \phi\left(x_i, z_j\right) + \varepsilon_i\left(z_j\right)$, where x_i and z_j are observed. Therefore, characteristics of agent j that are not observed in the data do not directly affect the utility of agent i.

Estimates of the distribution of the latent index as a function of observables are useful for the analysis of many economic questions. For instance, quantitatively evaluating the trade-off for firms between workers' experience or education and unobserved productivity, or the trade-off for workers between wages and the value of amenities such as on-the-job training may require estimates that account for unobserved characteristics. Similarly, evaluating the consequences of a market reform (such as policies that place limits on college tuition) can require estimating the distribution of latent indices on both sides of the market. Identifying and estimating preferences of agents on both sides of the market may be a challenging exercise because equilibrium matches are jointly determined by both sets of preferences: when we see a student enrolling at a particular college, it need not be the case that the college is her most preferred option because she may have not been accepted at a more preferred institution.

We study these problems assuming that the available data are from a large market. This approach is motivated by the fact that data from several matching markets with the same underlying structure are rare compared to data from a few markets with many agents. For example, public high school markets, colleges, the medical residency market, and marriage markets have at least several thousand participating agents. For similar reasons, recent papers in the theoretical matching literature have utilized large market approximations for analyzing strategic behavior and the structure of equilibria.[4] In our analysis, large market approximations highlight and account for important interdependence between matches within a market in the asymptotic analysis of our estimator.

Even with the stark restriction that preferences are homogeneous, our first result on identification is negative. We show that the distribution of the latent index is not identified from data from a single large market with one-to-one matches. Indeed, we construct an example parametric family of models of one-to-one matching that are observationally equivalent. This example illustrates that our non-identification result is not pathological. Intuitively, the observed joint distribution of agents and their match partners, which we

[4]See Immorlica and Mahdian (2005); Kojima and Pathak (2009); Azevedo and Budish (2017), for example.

refer to as sorting patterns, does not allow us to condition on unobservables. The non-identification arises because unobservable characteristics of either side of the market could be driving these sorting patterns. These results imply limitations on what can be learned using data on one-to-one matches and guide the use of empirical techniques. For instance, they weigh against estimating the distribution of the latent index in marriage markets using data from a single market. Nonetheless, data from one-to-one matches may still be useful for certain questions. We show that the relative value of various observed characteristics are identified with one-to-one matches. However, this limits the scope of questions that may be answered with such data.

In contrast to the non-identification result with one-to-one matches, we show that the distribution of latent indices on both sides of the market is non-parametrically identified from data on many-to-one matches. The key insight is that the same value of the unobservable characteristic of an agent determines multiple matches for that agent. The formal result requires that each agent on one side of the market is matched to at least two agents on the other side, a requirement that is likely satisfied in many education and labor markets. To the best of our knowledge, this difference between the empirical content of one-to-one matching and many-to-one matching has not been previously exploited to obtain non-parametric identification results of a model with unobserved characteristics. Our proof is based on interpreting the matching model with two-to-one matches in terms of a measurement error model (Hu and Schennach, 2008). This reinterpretation makes the additional empirical content of many-to-one matches ex-post intuitive: the observable components of a worker's quality provide a noisy measure of the overall quality of her colleagues. As in measurement error models, we use the repeated measurements made available when many workers match with the same firm to identify the model.

We also use simulations from a parametrized family of models to illustrate the additional identifying information available in many-to-one matches. Our simulations suggest that moments that only use information available in sorting patterns are not able to distinguish between a large set of parameter values. In the context of one-to-one matching, this is

the only information observed in a dataset from a single large market. In contrast, our simulations also suggest that additional moments constructed from many-to-one matching can be used to distinguish parameter values that yield indistinguishable sorting patterns. An objective function constructed from both sets of moments has a global minimum near the true parameter. These simulations suggest that using such information is important in empirical applications. For example, they suggest that moments such as the within-firm variance in worker observables contain information about primitives beyond what can be learned from the covariance between worker and firm observables. We therefore recommend empirical strategies that use information from many-to-one matching, when available.[5]

We then study the asymptotic properties of a minimum distance estimator for a parametric model based on a criterion function that uses moments from many-to-one matching as well as sorting patterns. As in the identification analysis, we develop an asymptotic theory based on data from a single market with the number of agents growing large. This approach requires us to deal with technical challenges that arise from the dependence of each match on the characteristics of all agents in the market. We prove both consistency and $\sqrt{N}-$asymptotic normality of the estimator. For simplicity, we restrict attention to the case with two-to-one matching. To our knowledge, ours is the first result on asymptotic theory for an estimator in a single large matching market.

Our asymptotic theory requires us to confront the fact that the observed matches, as well as the model predictions, are a non-linear function of the observables and unobservables in the entire market. We separately analyze the sampling distribution of the moments in the data and the map from the structural parameters to these moments. To prove a limit theorem for the sampling distribution of the moments in the data, we use the fact that the distribution of the observed characteristics of matched pairs depends only on the latent index. Hence, the conditional distribution of the observables given the latent indices are

[5]If data from many-to-one matches is not available, it may be possible to use variation in market composition to identify the distribution of latent indices. We are not aware of any formal results that show that such variation is sufficient for identification. This approach may require assuming that the parameters governing the primitives are constant across the markets.

independent on the two sides of the market. This insight allows us to derive the asymptotic distribution of the moments of the data.

Then, we study the model's prediction for the moments as a function of the structural parameters and the observables in the data. Analyzing this map is challenging because the matches depend on the characteristics of all agents in the market. This generates dependency that cannot be analyzed using standard empirical process techniques for i.i.d. data (e.g. van der Vaart and Wellner, 2000). In particular, deriving the sensitivity of the matches between extremely desirable or extremely undesirable agents to the parameter requires controlling the tail behavior of the latent index. We make progress by first showing that this map, ignoring the tails of the latent utilities, is smooth – specifically, Hadamard differentiable – in the sampled observed characteristics. This allows us to use continuous mapping theorems and the functional delta method to show convergence properties, except at the tails. When the tails are negligible, the limit as the size of the tails we ignore goes to zero yields large sample properties of the moment function.

The dependence inherent in the model also complicates the analysis of these tails. We show that the tails are negligible by adapting a chaining argument from the empirical process literature (Pollard, 2002), replacing a tail bound for i.i.d. data used in the existing proof with a concentration of measure inequality (Boucheron *et al.*, 2003) suitable to the dependent data in our problem. This method allows us to prove the equicontinuity results necessary for the limit theorem. For simplicity of exposition in the main text, the technical regularity conditions on the primitives that justify this approximation are detailed in the appendix. Finally, we use Monte Carlo simulations to study the property of a simulated minimum distance estimator.

The paper starts with a brief discussion of the related literature, after which we present the model (Section 3.1). Section 3.2 discusses identification with one-to-one and many-to-one matching, Section 3.3 presents our asymptotic analysis of the estimator, and Section **??** presents Monte Carlo results. All proofs are in the Appendix.

Related Literature: Most of the recent literature on identification and estimation of matching games studies the transferable utility (TU) model, in which the equilibrium governs the matches as well as the surplus split between the agents with quasi-linear preferences for money (Choo and Siow, 2006; Sorensen, 2007; Fox, forthcoming; **?**; Galichon and Salanie, 2012; Chiappori *et al.*, 2015, among others). The equilibrium transfers are such that no two unmatched agents can find a profitable transfer in which they would like to match with each other. The typical goal in these studies is to recover a single aggregate surplus that determines the equilibrium matches. A branch of this literature, following the work of Choo and Siow (2006), proposes identification and estimation of a transferable utility model based on the assumption that each agent's utility depends only on observed characteristics and an unobserved taste shock drawn from a specified distribution. Using this assumption, the papers propose estimation and identification of group-specific surplus functions (Choo and Siow, 2006; Galichon and Salanie, 2012; Chiappori *et al.*, 2015). A different approach to identification in transferable utility models, due to Fox (forthcoming), is based on assuming that the structural unobservables are such that the probability of observing a particular match is higher if the total systematic, observable component of utility is larger than an alternative match. Compared to these approaches, our study is restricted to a single index model but incorporates both TU and NTU matching in a non-parametric framework. We also allow for unobserved characteristics of the partner to affect agent preferences and are interested in identifying the distribution of unobservable characteristics, which are not considered in the maximum score approach by Fox (forthcoming).

In many applications, inflexible monetary transfers or counterfactual analyses that require estimates of preferences for agents on both sides of the market motivate the use of a non-transferable utility model (c.f. Roth and Sotomayor, 1992). Previous analysis of NTU models have resulted in only partial identification. Hsieh (2011) follows Choo and Siow (2006) in assuming that agents belong to finitely many observed groups and that agents have idiosyncratic tastes for these groups. The main identification result in Hsieh (2011) shows that the model can rationalize any distribution of matchings in this setting, implying

that the identified set is non-empty. Menzel (2015) studies identification and estimation in a non-transferable utility model in a large market where agent preferences are heterogeneous due to idiosyncratic match-specific tastes with a distribution in the domain of attraction of the Generalized Extreme Value (GEV) family and in which observable characteristics have bounded support. Menzel (2015) finds that only the sum of the surplus of both sides obtained from matching is identified from data on one-to-one matching. The result that identification is incomplete with one-to-one matching is similar in spirit to our negative result on identification. While these papers focus on the one-to-one matching case, our results exploit data on many-to-one matches to non-parametrically identify preferences of both sides of the market, although our results come at the cost of assuming homogeneous preferences.

With the exception of Chiappori *et al.* (2012) and Galichon *et al.* (2014), previous models are typically restricted to either non-transferable or transferable utility. The objective in Galichon *et al.* (2014) is to generalize the Choo and Siow (2006) framework to models of imperfectly transferable utility. Our framework is closer to that of Chiappori *et al.* (2012), which studies a marriage market with positive assortative matching. They also assume a single index model and allow for both transferable and non-transferable utility matching. They show that the marginal rates of substitution between two observable characteristics is identified using data on one-to-one matching. Our identification results with data on one-to-one matching are consistent with their results, but may also explain why Chiappori *et al.* (2012) may not have estimated the distribution of the latent index with their data. Specifically, we show that a many-to-one matching market is needed for such identification. Agarwal (2015) and Vissing (2016) use our insight on the information in many-to-one matching to respectively estimate preferences in the market for medical residents and the market for oil drilling contracts using simulated minimum distance estimators. This approach is different from work by Logan *et al.* (2008) and Boyd *et al.* (2013), who propose techniques that use only the sorting of observed characteristics of agents as given by the matches (sorting patterns) to recover primitives. Our result on non-identification of a single-index model with data only

on sorting patterns implies that a more general model with heterogenous preferences will also not be identified. Therefore, our results suggest that point estimates obtained using only information in sorting patterns may be sensitive to parametric assumptions.

A few empirical papers estimate sets of preference parameters that are consistent with pairwise stability (Menzel, 2011; Uetake and Watanabe, 2013). The concern that preferences need not be point identified with one-to-one matches does not necessarily apply to these approaches. For example, Menzel (2011) uses two-sided matching to illustrate a Bayesian approach for estimating a set of parameters consistent with an incomplete structural model. Our results on non-identification and subsequent simulations that use information on sorting patterns suggest that a rather large set of parameters are observationally equivalent. While these results imply that the identified set may be large, these approaches may still be informative for certain questions of interest.

Our finding that data from many-to-one matching is important in identification is related to work by Fox (forthcoming, 2010) on many-to-many matching. In these papers, many-to-many matching games allow identification of certain features of the observable component of the surplus function when agents share some but not all partners. This allows differencing the surplus generated from common match partners to learn valuations. In our setting, many-to-one matching plays a different role in that it allows us to learn the extent to which unobservable characteristics of each side of the market drive the observed patterns.

The results on identification with many-to-one matching are based on techniques for identifying non-linear measurement error models developed in Hu and Schennach (2008). These techniques have been applied to identify auction models with unobserved heterogeneity (Hu *et al.*, 2013), and dynamic models with unobserved states (Hu and Shum, 2012). To our knowledge, these techniques have not been previously used to identify matching models.

Finally, we use a novel approach for dependent data to prove our limit theorems because standard empirical process theories for i.i.d. data are not applicable in our context. This feature of our model may be shared with other contexts, such as network formation models

(c.f. Graham, 2017; Boucher and Mourifie, 2012; Chandrasekhar and Jackson, 2015; Leung, 2015). A common technique in the asymptotic analysis of network models is based on assuming that dependence across links decays with a notion of distance between two nodes. Our application of concentration of measure inequalities removes the need for an analogous assumption in our model.

3.1 Model

We consider a two-sided matching market with one side labelled as workers and the other labeled firms. Although these labels are suggestive of a labor market, the model may be applied to other two-sided matching markets, including matching of students to schools, and the marriage market. Our model does not presume a monetary transfer between the two sides of the market, and will include both non-transferable and transferable utility cases. We first describes the latent indices that will be the object of interest in our identification and estimation analysis before discussing their interpretation in transferable and non-transferable utility models. Finally, we discuss questions of interest that may be answered in this framework.

3.1.1 Latent Indices

Most datasets have information on a limited number of characteristics on each side of the market. Let the observable characteristics of worker i be x_i and the observable characteristics of firm j be z_j. Given our focus on positive assortative matching, we posit two latent quality indices, v_i and u_j, one for each side of the market. These indices simply order workers and firms by quality and do not impose cardinal restrictions. For instance, firms may have heterogeneous production functions that take human capital (v_i) as an input. The latent indices can depend on both observable characteristics as well as unobserved characteristics. Specifically, we assume that worker i's human capital index is given by the additively

separable form

$$v_i = h(x_i) + \varepsilon_i, \tag{3.1}$$

where we set the location normalizations $h(\bar{x}) = 0$ for some $\bar{x}$, assume that ε is median zero, and set the scale normalization $|\nabla h(\bar{x})| = 1$. Because an additively separable representation of preferences is unique up to a positive affine transformation, the scale and location normalizations are without loss of generality. These normalizations ensure that the latent indices in our model are well defined.

The scalar unobservable ε_i aggregates the effect of all relevant determinants worker quality that are not observed in the dataset. Additive separability in ε_i implies that the marginal value of observable traits does not depend on the unobservable.

As for the model for the human capital index, we assume that quality of firm j is given by

$$u_j = g(z_j) + \eta_j, \tag{3.2}$$

with the normalizations $g(\bar{z}) = 0$, η is median zero and $|\nabla g(\bar{z})| = 1$. The quality of the firm may reflect productivity differences or on-the-job amenities for workers. For instance, one may also include wages in this model through one of the characteristics z_j if they are not negotiated during the matching process. This approach may be used to model medical residency markets or colleges/schools in countries with non-negotiable tuition rates.

We make the following assumptions on the model:

Axiom 3.1 *(i) ε and η are independent of X and Z respectively*

(ii) ε and η have bounded, differentiable densities, f_ε and f_η, with full support on $\mathbb{R}$, and non-vanishing characteristic functions

(iii) $h(\cdot)$ and $g(\cdot)$ are differentiable and have full support over $\mathbb{R}$

(iv) The random variables $h(X)$ and $g(Z)$ admit bounded continuous densities f_h and f_g

Assumption 3.1 (i) assumes independence of the unobservables. On its own, inde-

pendence is not particularly strong, but the restriction of additive separability makes this restrictive. Additive separability with independence is commonly used in discrete choice literature as it significantly eases the analysis. Assumption 3.1 (ii) requires that ε and η have large support and imposes technical regularity conditions on their distributions that will be useful in our identification analysis. The support conditions in Assumption 3.1 (iii) ensures that the observables can trace out the distribution of ε and η in the tails as well, and Assumption 3.1 (iv) requires at least one covariate to be sufficiently smooth while others may be discrete.

3.1.2 Positive Assortative Matching

The composition of the market is described by a pair of probability measures, $\mu_{X,\varepsilon}$ and $\mu_{Z,\eta}$. Here, $\mu_{X,\varepsilon}$ is the joint distribution of workers' observable traits $x \in \chi \subseteq \mathbb{R}^{k_x}$ and unobservable traits $\varepsilon \in \mathbb{R}$. Likewise, $\mu_{Z,\eta}$ is the joint distribution of firms' observable traits $z \in \zeta \subseteq \mathbb{R}^{k_z}$ and unobservable traits $\eta \in \mathbb{R}$. This formulation allows us to consider large but finite economies, as well as a continuum limit in a unified notational framework. For instance, an economy with N agents on each side can be represented with the measures $\mu_{(X,\varepsilon)_N} = \frac{1}{N}\sum_{i=1}^{n} \delta_{(X_i,\varepsilon_i)}$ and $\mu_{(Z,\eta)_N} = \frac{1}{N}\sum_{j=1}^{n} \delta_{\left(Z_j,\eta_j\right)}$, where δ_Y is the dirac delta measure at Y.

A one-to-one **match** is a probability measure μ on $(\chi \times \mathbb{R}) \times (\zeta \times \mathbb{R})$ with marginals $\mu_{X,\varepsilon}$ and $\mu_{Z,\eta}$ respectively. The measure μ could be used to represent a continuum limit as well as a finite-economy match. The traditional definition of a finite-market match used in Roth and Sotomayor (1992) is based on a matching function $\mu^*(i) \mapsto \mathcal{J} \cup \{i\}$, where $\mathcal{J}$ is the set of firms. For an economy of size N, with probability 1, such a function defines a unique counting measure of the form $\mu_N = \frac{1}{N}\sum_{i,j=1}^{N} \delta_{\left(X_i,\varepsilon_i,Z_j,\eta_j\right)}$, where $\delta_{\left(X_i,\varepsilon_i,Z_j,\eta_j\right)} > 0$ only if i is matched to j in a finite sample. When η and ε admit a density, in a finite economy, (z,η) (respectively (x,ε)) identifies a unique firm (respectively worker) with probability 1.[6]

[6]In addition to a traditional matching function, in a finite sample our definition also allows for fractional matchings. However, such realizations are not observed in typical datasets on matches.

A many-to-one match with M partners on one side is defined analogously as a measure μ on $(\chi \times \mathbb{R})^M \times (\zeta \times \mathbb{R})$.

The match μ is **positive assortative** if there do not exist two (measurable) sets $S_I \subseteq \chi \times \mathbb{R}$ and $S_J \subseteq \zeta \times \mathbb{R}$ in the supports of $\mu_{X,\varepsilon}$ and $\mu_{Z,\eta}$ respectively, such that $\int_{S_I} (h(X) + \varepsilon)\, d\mu_{X,\varepsilon} > \int_{S_I} (h(X) + \varepsilon)\, d\mu(\cdot, S_J)$ and $\int_{S_J} (g(Z) + \eta)\, d\mu_{Z,\eta} > \int_{S_J} (g(Z) + \eta)\, d\mu(S_I, \cdot)$. This definition considers two potential sets of agents S_I and S_J. If $\int_{S_I} (h(X) + \varepsilon)\, d\mu_{X,\varepsilon} > \int_{S_I} (h(X) + \varepsilon)\, d\mu(\cdot, S_J)$, then the expected value of the latent indices of agents in S_I are larger than those matched with S_J. The analogous inequality for agents in S_J yields the second condition. Hence, there are no such sets if these inequalities are not simultaneously satisfied for any pair S_I and S_J, and the matching is assortative.[7]

This formulation presents a unified definition for assortativity in continuum markets as well as markets with a finite number of agents. In the finite market case, consider a match in which an agent with characteristics (x, ε) (respectively (x', ε')) is matched with an agent with characteristics (z, η) (respectively (z', η')). Now, consider singleton sets $S_I = \{(x, \varepsilon)\}$ and $S_J = \{(z', \eta')\}$. The inequalities above imply that either $h(x) + \varepsilon \leq h(x') + \varepsilon'$ or that $g(z') + \eta' \leq g(z) + \eta$. Therefore, there are no such pairs of sets in the finite markets if the conditions of our definition are satisfied. In what follows, we simply assume that the market is characterized by positive assortative matching. As we discuss in the next few sections, this assumption encompasses both transferable and non-transferable utility models.

Further, our model requires that the matching only depends on the latent index. This assumption is vacuous in finite samples because ties are zero-probability events. Shi and Shum (2014) formalize this as "random matching" in a continuum version of the Beckerian marriage model. They note that without this assumption, the distribution of observed characteristics of matched partners is indeterminate. Our consistency results imply that

[7] We do not consider unmatched agents for two reasons. First, different equilibrium notions matching (TU or NTU) impose different restrictions on preferences for unmatched agents. Using the implications of positive assortative matching alone allows us to be agnostic about the nature of transfers. Second, many datasets do not have information on unmatched agents. For example, typical employer-employee matched datasets do not contain the number of job openings, and Agarwal (2015) does not have information on medical residents that were not placed at residency programs.

the moments of the finite sample data naturally converge to a population analog with this property. Therefore, the data generating process we consider has the following property in a positive assortative match:

Remark 3.1 *A positive assortative match μ has support on $(x, \varepsilon, z, \eta)$ only if $F_U(u(z,\eta)) = F_V(v(x,\varepsilon))$, where F_U and F_V are the cumulative distributions of u and v respectively. Further, the latent index is a sufficient statistic for the distribution of match partners.*

Hence, the firm with the q-th quantile position of value to the worker is matched with the worker with the q-th quantile of desirability to the firm. The dependence only on the latent index, in the one-to-one case, implies that $\mu_{X,\varepsilon|Z,\eta} = \mu_{X,\varepsilon|Z',\eta'}$ if $g(Z) + \eta = g(Z') + \eta'$ and $\mu_{Z,\eta|X,\varepsilon} = \mu_{Z,\eta|X',\varepsilon'}$ if $h(X) + \varepsilon = h(X') + \varepsilon'$. Our paper studies identification and estimation of the latent utility indices using data from a matching market with this property. As described below, it turns out that positive assortative matching on v and u can result from both non-transferable and transferable utility models.

Non-transferable Utility (NTU) Matching

Models of matching markets in which transfers between the parties are prohibited or restricted are commonly used in the theoretical literature (c.f. Roth and Sotomayor, 1992). Motivating examples include marriage markets, public schooling, and colleges. In such a model, the latent indices v_i and u_j are interpreted as representing the *ordinal* preference relation for their match partners. Because these indices are ordinal, the framework allows for each firm j to have a separate production function $\Phi_j(v)$ as long as Φ_j is strictly increasing in v. In the many-to-one matching case, a focus of this paper, we will assume that Φ_j is increasing in each of its components. Specifically, $\Phi_j(v_1, v_2)$ is increasing in both v_1 and v_2 when a firm is matched with two workers.

The typical equilibrium assumption is that of pairwise stability, which makes two restrictions. First, there is no worker-firm pair such that both agents prefer matching with each other to their current match (where the firm can fire a currently matched worker, if

necessary). Second, no worker or firm is matched with an unacceptable partner. Existence of a pairwise stable match follows in a finite market because preferences are *responsive* (Roth and Sotomayor, 1992) and uniqueness follows from *alignment of preferences* as discussed in Clark (2006) and Niederle and Yariv (2009). It is easy to see that the unique pairwise stable match is positive assortative on the latent indices v_i and u_j. Given our focus on positive assortative matching, we assume that all workers and firms are acceptable to the other side.

Although the models are referred to as non-transferable utility models, the model can incorporate transfers that are not simultaneously determined with the matching. In this case, one of the observables includes the salary offered by program j. Estimating the latent index allows one to measure the willingness to pay for various on-the-job amenities by assuming a functional form, say

$$u_j = z_j\beta + w_j + \eta_j. \tag{3.3}$$

For instance, Agarwal (2015) uses a similar model to quantify the value for various attributes of medical residency programs such as size, prestige, and patient mix.

An important restriction in the latent index framework is that agents have homogeneous ordinal preferences over their match partners. While the theoretical literature assumes very general preferences when studying the existence of stable matchings, formal identification and estimation analysis is yet to incorporate significant heterogeneity in preferences.

Transferable Utility (TU) Matching

Our latent index framework fits well into the classical Beckerian model of the marriage market. This matching model posits men and women differentiated by a one-dimensional characteristics that split a surplus from marriage given by $\Phi\left(u_j, v_i\right)$. A matching is pairwise stable if there are transfers t_{ij} (possibly negative), such that no man-woman pair find it mutually beneficial to agree to a transfer and match with each other. As is well known, the unique pairwise stable match is positive assortative on u and v if $\Phi\left(u_j, v_i\right)$ is supermodular. This elegant model has received a considerable amount of attention, and patterns of positive

assortative matching observed along age, income, and education in various marriage markets have been well documented.

A thrust of our paper is the consideration of many-to-one matching. In this case, we assume separability of the surplus function across matches in order to maintain positive assortative matching on the latent indices. Specifically, we assume that the surplus generated by firm with index u_j that is matched with workers v_i and v_k is given by

$$\Phi\left(v_i, v_k, u_j\right) = \tilde{\Phi}\left(v_i, u_j\right) + \tilde{\Phi}\left(v_k, u_j\right), \tag{3.4}$$

where $\tilde{\Phi}_{\cdot}$ is supermodular. The assumption rules out complementarities across matches but retains positive assortativity in a pairwise stable match. It also assumes that the multiple matches for an agent are symmetric. For example, in the worker-firm context, the model is best suited for a market in which firms are hiring multiple workers with the same job description.

3.1.3 Unobserved Characteristics

The lack of perfect positive assortative matching on observable characteristics may be attributable to unaccounted preference heterogeneity or unobserved characteristics. These unobserved characteristics are important for rationalizing the data. Previous approaches have typically focussed on the identification of observable components of utility, often under parametric assumptions on the distribution of unobserved characteristics.[8] For instance, Chiappori *et al.* (2012) study a single index model like ours and obtain identification of the

[8]For instance, Galichon and Salanie (2012) generalize the model by Choo and Siow (2006) and show that $\bar{\Phi}_{x_i z_j}$ is identified for a separable surplus function of the form $\Phi_{ij} = \bar{\Phi}_{x_i z_j} + \varepsilon_i\left(z_j\right) + \eta_j\left(x_i\right)$ with known distributions of $\varepsilon_i\left(z_j\right)$ and $\eta_j\left(x_i\right)$. These models therefore allow for unobserved preferences for observed characteristics, but do not allow for unobserved characteristics. Menzel (2015) studies an NTU model with a light restriction on tail behavior of the unobservables to identify the sum of the match surpluses accruing to each side due to observables.

marginal rates of substitution

$$\frac{\partial h(x)/\partial x_1}{\partial h(x)/\partial x_2} \text{ and } \frac{\partial g(z)/\partial z_1}{\partial g(z)/\partial z_2}.$$

These quantities can be used to determine the trade-offs between observables, such as the trade-off between a worker's education and experience. Some economic questions, however, may require an analysis of unobservables. For example, one may be interested in knowing how much of a worker's human capital is explained by experience and/or education. This exercise may require decomposing the variance of human capital into observable and unobservable components. Similarly, questions about compensating differentials in labor markets require valuing on-the-job amenities or training, some components of which may not be observed.

While several objects of interest can be measured through these marginal rates of substitution between observed characteristics, many economic questions require a deeper understanding of how agents' preferences respond to interventions in matching markets. For example, one may be interested in the effect of a subsidy on college tuition on matches that occur in equilibrium. To predict the counterfactual matches, one needs to measure the effect of this subsidy on the relative desirability of various colleges to students. Changes in the relative desirability of colleges depend on the monetary value students place on unobserved college characteristics. Therefore, an important objective in this paper is to understand when the distributions of ε and η are identified, which in turn implies identification of the probabilities

$$\mathbb{P}\left(h(x_1) + \varepsilon_1 > h(x_2) + \varepsilon_2 | x_1, x_2\right) \text{ and } \mathbb{P}\left(g(z_1) + \eta_1 > g(z_2) + \eta_2 | z_1, z_2\right). \tag{3.5}$$

These choice probabilities are also fundamental in the analysis of counterfactual changes in market structure, market composition, and other policy-relevant counterfactuals.

It is important to note that the latent indices we analyze, u and v, are ordinal measures of the desirability of agents in the market. Identification of the total surplus function in the transferable utility case, $\Phi(u, v)$, or a cardinal measure of utilities in the non-transferable

utility case will require additional assumptions. For example, one may simply interpret the latent index as a utility measure in the NTU case or assume a particular structure for the surplus function in the TU case if this is desirable for the empirical application being considered. We avoid these assumptions for simplicity and to retain generality with respect to these choices. In applications where one of the observed traits presents a natural measure of value, our indices can be interpreted in units of this metric for value.

3.2 Identification

This section starts by showing that data from a single matching market are sufficient for identifying certain features of preferences. Specifically, one can identify the indices $h(x)$ and $g(z)$. We then show that data from one-to-one matches is unable to identify the distribution of the latent indices if there are unobserved characteristics on both sides of the market. Next, we show that data from many-to-one matching restores full identification of the distribution of preferences. Finally, we illustrate these results using simulations.

3.2.1 Sorting Patterns, Indifference Curves and a Sign-restriction

We now study what can be learned from the joint distribution μ_{XZ} of observed firm and worker traits. This is the marginal of μ on the observables, and it summarizes all information available in data from one-to-one matching. It allows the assessment of the sorting of worker observable traits to firm observables. We therefore refer to features of this distribution as "sorting patterns." As our first result shows, these features of the data allow us to identify the indices $h(x)$ and $g(z)$ up to monotonic transformations.

Lemma 3.1 *Under Assumption 3.1, the level sets of the functions $h(\cdot)$ and $g(\cdot)$ are identified from data on a one-to-one match, i.e. μ_{XZ} is observed.*

Proof. See Appendix B.1.1. ■

The result states that we can determine whether or not two worker types x and x' are equally desirable from the sorting patterns observed in a one-to-one matching market (hence,

also if many-to-one matches are observed). Intuitively, if two worker types have equal values of $h(\cdot)$, then the distributions of their desirability to firms are identical. Consequently, the distribution of firms they match with are also identical. In a positive assortative match, under the additive structure of equations (3.1) and (3.2) and the independence of unobserved traits, the distribution of firm observable types these workers are matched with turns out to be identical. Conversely, if two worker types are matched with different distributions of firm observables, they cannot be identical in observable quality. This result is similar to those in Chiappori *et al.* (2012) that show that differentiability of $h(\cdot)$ and $g(\cdot)$ implies identification of marginal rates of substitution, which are pinned down by indifference curves.

While the level sets of $h(\cdot)$ and $g(\cdot)$ are known, we cannot yet determine $h(\cdot)$ and $g(\cdot)$ even up to positive monotone transformations. In particular, we cannot tell whether a given worker trait is desirable or not. Intuitively, assortative matching between, say, firm size and worker age, may result from either both traits being desirable or both traits being undesirable. The next result shows that a **sign restriction** is sufficient for identifying $h(\cdot)$ and $g(\cdot)$ up to positive monotone transformations.

Axiom 3.2 *(i) The functions $h(x)$ and $g(z)$ are strictly increasing in their first arguments*

(ii) Further, for each $x_{-1} = (x_2, \ldots, x_{k_x})$ and $z_{-1} = (z_2, \ldots, z_{k_z})$, $h(x_1, x_{-1})$ and $g(z_1, z_{-1})$ have full support on $\mathbb{R}$

Part (i) imposes a sign restriction that requires that the latent index is strictly increasing in at least one observable characteristic. It is often natural to impose this restriction in matching markets. For example, it is reasonable to argue that the desirability of workers is increasing in education, holding all else fixed. Given such an assumption, our next result shows that $h(\cdot)$ and $g(\cdot)$ can be determined up to positive monotone transformations. Part (ii) makes a large support assumption that allows ordering all the level sets of $h(x)$.

Proposition 3.1 *If Assumptions 3.1 - 3.2 are satisfied then, $h(\cdot)$ and $g(\cdot)$ are identified up to positive monotone transformations.*

Proof. Identification of h and g up to a positive monotone transformation follows immediately from Lemma 3.1 and Assumption 3.2. ■

The sign restriction allows us to order the level sets of h and g.

3.2.2 Limitations of Sorting Patterns

As mentioned earlier, typical datasets do not contain all relevant characteristics of agents on both sides of the market. The dispersion around a central positive assortative trend is a manifestation of these unobservables. Remark 3.1 reflects the importance of unobservables as workers with characteristic (x, ε) are matched with firms with characteristics (z, η) if

$$h(x) = F_V^{-1} \circ F_U(g(z) + \eta) - \varepsilon. \tag{3.6}$$

This expression indicates that there are two sources of unobservables that result in imperfect assortativity, namely η and ε. Without these unobservables, a researcher would observe perfect positive assortativity along the estimated indices $h(x)$ and $g(z)$.

A question remains about whether we can learn about the distribution of both these unobservables with data on one-to-one matches, which only contains information in F_{XZ}. The following stylized example shows that the answer is negative. A wide range of parameters could be consistent with the data, even a highly parametric case.

Example 3.1 *Let $h(x) = x$ and $g(z) = z$. Assume that X, Z are distributed as $N(0,1)$ and ε, η are distributed as $N(0, \sigma_\varepsilon^2)$ and $N\left(0, \sigma_\eta^2\right)$ respectively. The distributions of U and V are therefore $N\left(0, 1 + \sigma_\eta^2\right)$ and $N(0, 1 + \sigma_\varepsilon^2)$ respectively. It is straightforward to show that $X|V = v \sim N\left(\frac{1}{1+\sigma_\varepsilon^2} v, \frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2}\right)$, that $U|Z \sim N\left(Z, \sigma_\eta^2\right)$, and that $F_V^{-1} \circ F_U = \left[\frac{1+\sigma_\varepsilon^2}{1+\sigma_\eta^2}\right]^{1/2}$. Therefore, the distribution of $X|Z = z$ is given by the distribution of*

$$\frac{1}{1+\sigma_\varepsilon^2} F_V^{-1} \circ F_U(z + \eta) + \varepsilon_1,$$

where $\varepsilon_1 \sim N\left(0, \frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2}\right)$ and $\eta \sim N\left(0, \sigma_\eta^2\right)$, independently of X and Z. Hence, $X|Z = z$ is

distributed as

$$N\left(\frac{z}{\kappa^{1/2}}, 1-\frac{1}{\kappa}\right), \tag{3.7}$$

where $\kappa = \left(1+\sigma_\varepsilon^2\right)\left(1+\sigma_\eta^2\right)$.

The distribution in equation (3.7) is identical for all pairs $\left(\sigma_\varepsilon, \sigma_\eta\right)$ *with* $\left(1+\sigma_\eta^2\right)\left(1+\sigma_\varepsilon^2\right) = \kappa$. *Thus, the family of matching models with* $\left(1+\sigma_\eta^2\right)\left(1+\sigma_\varepsilon^2\right) = \kappa$ *are observationally equivalent when only data from one-to-one matches is available.*

The example above shows that data on one-to-one matches cannot be used to identify the distribution of the two latent indices in the presence of unobservables on both sides of the market. This highlights a central limitation of data from a market with one-to-one matching such as the marriage market.[9] Section 3.2.4 illustrates this limitation using a simulated objective function.

The failure of identification can be understood by considering the case in which $\varepsilon \equiv 0$. Equation (3.6) reduces to

$$h\left(x\right) = F_V^{-1} \circ F_U\left(g\left(z\right) + \eta\right).$$

This expression shows that when $\varepsilon \equiv 0$, the model is mathematically identical to the well-studied transformation model (Ekeland *et al.*, 2004; Chiappori and Komunjer, 2008). Appendix C.1.1 uses results from Chiappori and Komunjer (2008) to formally derive conditions under which any distribution F_{XZ} can be rationalized.

These results imply that a model with unobservables on both sides is under-identified. This non-identification is despite imposing additional regularity conditions. Hence, empirical strategies to estimate the distribution of latent preferences using information in sorting patterns may be suspect. Logan *et al.* (2008) and Boyd *et al.* (2013) employ empirical strategies that only use sorting patterns to estimate preferences for models that include preference heterogeneity. Our non-identification result suggests that point estimates from

[9] This observation suggests one reason why Chiappori *et al.* (2012) do not estimate the distribution of the latent index in their paper on the marriage market.

this approach, including for models more general than the single index model, may be sensitive to parametric assumptions. Such non-identification is problematic for counterfactuals relying on the probability of choices. For instance, the result implies that the data can be rationalized in a model in which any worker with trait x is preferred to any worker with trait x' if $h(x) > h(x')$, even if this is not the case.

3.2.3 Identification from Many-to-One Matches

We now show that data from many-to-one matching markets can be used to identify the model. Consider a dataset in which there are a large number of firms, and each firm has two workers hired at the same position. Therefore, we may arbitrarily label the slots occupied by each worker as slots 1 and 2, independently of the firm and worker characteristics. The data are summarized by the joint distribution $F_{X_1,X_2,Z}$, where X_1 and X_2 are the observed characteristics of the two workers employed at a firm with observable characteristic Z.

To see why multiple matches per partner can be useful for identification, note that the observed worker/firm characteristics present noisy measures of the true quality of the partners matched with each other. Remark 3.1 implies that the following two equalities when workers with characteristics (x_1, ε_1) and (x_2, ε_2) are matched with a firm with characteristics (z, η):

$$\begin{aligned} h(x_1) &= F_V^{-1} \circ F_U(g(z) + \eta) - \varepsilon_1 \\ h(x_2) &= F_V^{-1} \circ F_U(g(z) + \eta) - \varepsilon_2. \end{aligned}$$

Agarwal (2015) uses this insight and discusses it in the context of the medical residency market. The argument is that if the medical school quality of a resident is highly predictive of human capital, then the variation within programs in human capital should be low. If unobservables such as test scores and recommendations are important, then residency programs should be matched with medical residents from medical schools of varying quality. Our result below formally shows the usefulness of data from many-to-one matching. We therefore recommend the use of this information when available.

Theorem 3.1 *Under Assumptions 3.1 - 3.2, the functions $h(\cdot)$, $g(\cdot)$ and the densities f_η and f_ε are identified when data from two-to-one matching is observed, i.e. $F_{X_1,X_2,Z}$ is observed.*

Proof. See Appendix B.1.2. ■

The proof proceeds by interpreting our model in terms of a nonlinear measurement error model and employing techniques in Hu and Schennach (2008) to prove identification. To understand the analogy, note that the distribution observables of matched partners depends only on the latent index. Positive assortative matching implies that all partners have the same quantile of the latent index. Therefore, to write the joint distributions of the observables given a quantile q, we need to consider the conditional densities of the observables X_1, X_2 and Z given q. For expositional simplicity, assume that these densities exist. Therefore, the joint distribution $f_{X_1,X_2,Z,q}$ can be factored as follows:

$$f_{X_1,X_2,Z,q}(x_1,x_2,z,q) = f_{X_1|q}(x_1|q)\ f_{X_2|q}(x_2|q)\ f_{Z|q}(z|q)\ f_q(q),$$

where $f_q(q) = 1$ for $q \in [0,1]$ and 0 otherwise because quantiles are uniformly distributed, $f_{X_1|q}(x_1|q)$ is the conditional density at x_1 given that $h(x_1) + \varepsilon = F_V^{-1}(q)$, and $f_{X_2|q}(x_2|q)$ and $f_{Z|q}(z|q)$ are defined analogously. Integrating this quantity with respect to q yields the observable quantity

$$f_{X_1,X_2,Z}(x_1,x_2,z) = \int_0^1 f_{X_1|q}(x_1|q)\ f_{X_2|q}(x_2|q)\ f_{Z|q}(z|q)\,dq.$$

Intuitively, this simplification arises from the latent index assumption and positive assortative matching on v and u. Mathematically, this equation is identical to the nonlinear measurement error model of Hu and Schennach (2008), with the latent variable q governing the distribution of the observables.[10] This formulation clarifies the intuition that the observable characteristics of matched partners are noisy signals of the underlying latent index, and it allows us to identify the distributions of X and Z conditional on the quantile q. We then identify the model using the scale and location normalizations on h, g, f_ε, f_η, and

[10] A technical difference with Hu and Schennach (2008) is that we replace Assumptions 1 and 5 in their paper with implications of Assumptions 3.1. See appendix for details.

Assumption 3.1.

While these results are derived in the specific context of a latent index model with no preference heterogeneity, they highlight the fact that data from many-to-one matches has additional empirical information that cannot be obtained from one-to-one matches. This insight has enabled and guided the empirical analyses of more flexible models of the medical match (Agarwal, 2015) and the market for oil drilling rights (Vissing, 2016). An extension of our analogy of a matching model to a measurement error model has also been used to prove identification results for and study a labor market model with data on worker productivity (Jiang, 2016).

3.2.4 Importance of Many-to-one Match Data: Simulation Evidence

The identification results presented in the previous section relied on observing data from many-to-one matching, and they show that the model is not identified using data from one-to-one matches. In this section, we present simulation evidence from a parametric version of the model to elaborate on the nature of non-identification and to illustrate the importance of using information from many-to-one matching in estimation. To mimic realistic empirical applications, our simulations have firms with varying capacity instead of the fixed number of workers per firm.

We simulate a dataset using known parameters and then compare objective functions of various minimum distance estimators. Specifically, we compare an objective function that exclusively uses moments based on sorting patterns to another that also uses information from many-to-one matching. We model the latent indices as

$$\begin{aligned} v_i &= x_i\alpha + \varepsilon_i \\ u_j &= z_j\beta + \eta_j, \end{aligned}$$

where x_i, z_j, ε_i, η_j are distributed as standard normal random variables. These parametric assumptions are identical to those used in Example 3.1. We generate a sample using $J = 500$ firms. Each firm j has capacity q_j drawn uniformly at random from $\{1, \ldots, 10\}$. The

number of workers in the simulation is $N = \sum c_j$. A pairwise stable match $\mu : \{1, \ldots, N\} \to \{1, \ldots, J\}$ is computed for $\alpha = 1$ and $\beta = 1$. Using the same dataset of observables and firm capacities, the variables ε_i and η_j are simulated $S = 1000$ times, and a pairwise stable match μ_s^θ can be computed for each $s \in \{1, \ldots, S\}$ as a function of $\theta = (\alpha, \beta)$. We then compute two sets of moments

$$\hat{\psi}_{ov} = \frac{1}{N} \sum_i x_i z_{\mu(i)} \tag{3.8}$$

$$\hat{\psi}_{ov}^S(\theta) = \frac{1}{S} \sum_s \frac{1}{N} \sum_i x_i z_{\mu_s^\theta(i)} \tag{3.9}$$

and

$$\hat{\psi}_w = \frac{1}{N} \sum_i \left(x_i - \frac{1}{|\mu^{-1}(\mu(i))|} \sum_{i' \in \mu^{-1}(\mu(i))} x_{i'} \right)^2 \tag{3.10}$$

$$\hat{\psi}_w^S(\theta) = \frac{1}{S} \sum_s \frac{1}{N} \sum_i \left(x_i - \frac{1}{\left|(\mu_s^\theta)^{-1}(\mu_s^\theta(i))\right|} \sum_{i' \in (\mu_s^\theta)^{-1}(\mu_s^\theta(i))} x_{i'} \right)^2. \tag{3.11}$$

The first set, $\hat{\psi}_{ov}$ and $\hat{\psi}_{ov}^S(\theta)$, captures the degree of assortativity between the characteristics x and z in the pairwise stable matches in the generated data and as a function of θ. For a given $\alpha > 0$ (likewise $\beta > 0$), this covariance should be increasing in β (likewise α). The second set, $\hat{\psi}_w$ and $\hat{\psi}_w^S(\theta)$ captures the within-firm variation in the characteristic x. If the value of α is large, we can expect that workers with very different values of x are unlikely to be of the same quantile. Hence, the within-firm variation in x will be small. Using both sets of moments, we construct an objective function $\hat{Q}^S(\theta) = \left\| \hat{\psi} - \hat{\psi}^S(\theta) \right\|_W$, where $\hat{\psi} = \left(\hat{\psi}_{ov}, \hat{\psi}_w\right)'$, $\hat{\psi}^S(\theta) = \left(\hat{\psi}_{ov}^S(\theta), \hat{\psi}_w^S(\theta)\right)'$ and W indexes the norm.

Figure **??**(a) presents a contour plot of an objective function that only penalizes deviations of $\hat{\psi}_{ov}$ from $\hat{\psi}_{ov}^S(\theta)$. This objective function only uses information on the sorting between x and z to differentiate values of θ. We see that pairs of parameters, α and β, with large values of α and small values of β yield identical values of the objective function. These contour sets result from identical values of $\hat{\psi}_{ov}^S(\theta)$, illustrating that this moment cannot distinguish between values along this set. In particular, the figure shows that the objective function has

a trough containing the true parameter vector with many values of θ yielding similar values of the objective function.

In Figure **??**(b), we consider an objective function that only penalizes deviations of $\hat{\psi}_w$ from $\hat{\psi}_w^S(\theta)$. The vertical contours indicate that the moment is able to clearly distinguish values of α because the moment $\hat{\psi}_w^S(\theta)$ is strictly decreasing in α. However, the shape of the objective function indicates that this moment cannot distinguish different values of β.

Finally, the plot of an objective function that penalizes deviations from both $\hat{\psi}_w$ and $\hat{\psi}_{ov}$ (Figure **??**(c)) shows that we can combine information from both sets of moments to identify the true parameter. Unlike the other two figures, this objective function displays a unique minimum close to the true parameter. Together, Figures **??**(a)-(c) illustrate the importance of using both these types of moments in estimating our model.

3.3 Estimation

This section develops an estimator for the latent index model considered above. We then study the limit properties of this estimator and derive conditions under which the estimator is consistent and asymptotically normal. As in the identification analysis, we consider a dataset from a single large matching market. This choice is motivated by the fact that researchers typically have data on a single (or few) matching markets with many participants.[11] This includes applications in labor markets, marriage markets, and education markets. The analysis of asymptotic properties in a single large market is technically challenging because the characteristics of any individual's match partner depends on the composition of the entire market. To our knowledge, consistency or asymptotic theory has not been previously established for parametric models, even with a single latent index.[12]

There are several technical insights that allow us to solve this problem. First, we use the

[11] In cases where many matching markets are observed, it may not always be appropriate to assume that the underlying preference parameters are the same across all markets.

[12] Even proving consistency is non-trivial. For example, **?** show that the canonical correlation estimator suggested by Becker (1973) is inconsistent. A previously circulated version of this paper (Agarwal and Diamond, 2014) shows consistency of the estimator studied here under weaker conditions on the primitives.

property that we observe a positive assortative match along a single latent index. This allows us to re-write the dependence of the matches in terms of the latent indices. While restrictive on the nature of primitives, our model allows for a large parametric class of models and both transferable and non-transferable utility. Second, the problem can be decomposed into separately analyzing two distinct pieces. The first problem is to show limit theorems for the observed moments of the data as the market size increases. Separately, we must show a uniform limit theorem for the map from structural parameters to these moments. Third, we find that analyzing this map by first ignoring the behavior in the tails of the latent indices and then showing that the tails are negligible is the most tractable approach. Finally, to ensure that tails are negligible, we adapt a chaining argument from the empirical process literature, using a concentration of measure inequality to replace tail bounds for i.i.d. data that do not apply in our setting.

In this section, we assume that the latent indices of workers for firms and vice versa are known up to a finite dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}^{K_\theta}$. The latent indices are generated by

$$\begin{aligned} u(z, \eta; \theta) &= g(z; \theta) + \eta \\ v(x, \varepsilon; \theta) &= h(x; \theta) + \varepsilon, \end{aligned}$$

where $g : \zeta \times \Theta \to \mathbb{R}$ and $h : \chi \times \Theta \to \mathbb{R}$ are known functions that are Lipschitz continuous in θ for each x and z with constants $g_{LC}(z)$ and $h_{LC}(x)$ respectively. We assume that the densities f_ε and f_η are known, and ε and η are independent of x and z respectively.

We adopt a parametric approach for several reasons. First, our identification argument does not directly suggest a non-parametric estimator. Second, our focus is on solving issues that arise from the dependent data nature of the problem. Relaxing the parametric assumption would further complicate the analysis. Finally, computational burden in empirical applications have often prevented extremely flexible functional forms from being implemented. Similar parametric assumptions are common in the discrete choice literature where one typically assumes a normal or an extreme value type I distribution for the unobservable

ε.

We assume that the data contains a sample of J firms, each with $\bar{c}$ slots, and consider the properties of an estimator as $J \to \infty$. The number of workers is $N = \bar{c}J$. The characteristics of each worker are sampled i.i.d. from the measure $\mu_{X,\varepsilon}$ and the characteristics of the firm are sampled i.i.d. from $\mu_{Z,\eta}$. For simplicity of analysis and notation, we set $\bar{c} = 2$.

3.3.1 A Minimum Distance Estimator

We propose an estimator based on a minimum distance criterion function. Specifically, let $\Psi(x_1, x_2, z) \in \mathbb{R}^{K_\Psi}$ be a bounded vector-valued moment function, i.e. $\|\Psi\|_\infty < \infty$, where x_1 and x_2 are the observed characteristics of two workers and z is the observed characteristics of the firm. We assume that Ψ is symmetric in x_1 and x_2 because the data do not make a distinction between two workers hired at the same firm (for the same position). The data consist of matches between $N = 2J$ workers and J firms. Therefore, we observe $N/2$ triples $\left\{\left(x_{2j-1}, x_{2j}, z_j\right)\right\}_{j=1}^{N/2}$, which can be used to construct empirical moments of the form

$$\psi_N = \frac{1}{N/2} \sum_{j=1}^{N/2} \Psi\left(x_{2j-1}, x_{2j}, z_j\right). \tag{3.12}$$

The moments discussed in equations (3.8) and (3.10) are given by particular choices for Ψ.

We now describe the value of the moment as a function of θ. Instead of writing the sampling process as drawing pairs of (x_i, ε_i) and $\left(z_j, \eta_j\right)$, it will be convenient to rewrite the sampling distribution via Bayes' rule as sampling N and J draws from the population distributions of v_i and u_i respectively, and then sampling $x_i|v_i$ and $z_j|u_j$ from their respective conditional distributions. This sampling process has an identical distribution for (x_i, ε_i) and $\left(z_j, \eta_j\right)$ as sampling directly from their respective distributions. This rewriting uses the feature that the final matches depend on the latent indices rather than directly on observable and unobservable traits. Further, conditional on the latent indices, the observable traits of two workers matched to the same firm or different firms are independent. Therefore, given the utilities v_1, v_2, and u, at parameter vector θ and any two measures m_X and m_Z for the

observable traits, the value of the moment is:

$$\tilde{\psi}\left[m_X, m_Z\right](v_1, v_2, u; \theta) = \int \Psi\left(X_1, X_2, Z\right) f_{X|v_1;\theta}\left(X_1\right) f_{X|v_2;\theta}\left(X_2\right) f_{Z|u;\theta}\left(Z\right) dX_1 dX_2 dZ,$$

where $f_{X|v_1;\theta}\left(X\right)$ and $f_{Z|u;\theta}\left(Z\right)$ are the conditional densities (w.r.t. m_X and m_Z respectively) of the observable traits at θ given latent indices v and u, and m_X and m_Z. These distributions govern the observed traits of the workers and firms at any given quality.

In the limiting large market match, firms with the q-th quantile of firm quality are matched with workers on the q-th quantile of the worker quality distribution. Hence, the expected value of the moment of the q-th quantile match is given by $\tilde{\psi}$ evaluated at

$$(v_1, v_2, u) = \left(F^{-1}_{V;\theta,m_X}\left(q\right), F^{-1}_{V;\theta,m_X}\left(q\right), F^{-1}_{U;\theta,m_Z}\left(q\right)\right),$$

where $F_{V;\theta,m_X}\left(v\right)$ and $F_{U;\theta,m_Z}\left(u\right)$ are respectively the cumulative distributions of the worker and firm qualities (given θ, m_X and m_Z). This quantity must be integrated to obtain the moment as a function of the parameter θ:

$$\psi\left[m_X, m_Z\right](\theta) = \int_0^1 \tilde{\psi}\left[m_X, m_Z\right]\left(F^{-1}_{V;\theta,m_X}\left(q\right), F^{-1}_{V;\theta,m_X}\left(q\right), F^{-1}_{U;\theta,m_Z}\left(q\right); \theta\right) dq, \tag{3.13}$$

where $F_{V;\theta,m_X}\left(v\right) = \int_{-\infty}^{v} F_\varepsilon\left(v - h\left(X;\theta\right)\right) dm_X$, $F_{U;\theta,m_Z}\left(u\right) = \int_{-\infty}^{u} F_\eta\left(u - g\left(Z;\theta\right)\right) dm_Z$. This expression can be evaluated at any pair of measures m_X and m_Z governing the distribution of observed traits. Of particular interest are the quantities $\psi\left[\mu_X, \mu_Z\right](\theta)$ and $\psi\left[\mu_{X_N}, \mu_{Z_N}\right](\theta)$, which correspond to the values at the population and empirical measures of observables traits respectively. In this notation, the population analog of ψ_N in equation (3.12) is therefore $\psi\left[\mu_X, \mu_Z\right](\theta)$ evaluated at θ_0. For simplicity of notation, when referencing the moment function at populations measures μ_X and μ_Z, we will write $\psi\left(\theta\right) = \psi\left[\mu_X, \mu_Z\right](\theta)$. Similarly, when referencing their empirical analog μ_{X_N} and μ_{Z_N}, we will write $\psi_N\left(\theta\right) = \psi\left[\mu_{X_N}, \mu_{Z_N}\right](\theta)$.[13]

[13] $\psi_N\left(\theta\right)$ can be approximated by first drawing ε and η to simulate $F_{N,V;\theta} = F_{V;\theta,m_{X_N}}$ and $F_{N,U;\theta} = F_{U;\theta,m_{Z_N}}$, and then using the expression in equation (3.13). One can also create a simulation analog of $\psi_N\left(\theta\right)$ that uses a second simulation step to approximate the integral. More specifically, we may independently sample from the conditional distributions of X and Z given the measures μ_{X_N} and μ_{Z_N} and simulated values of v_i and u_j.

We now define our minimum distance estimator:

$$\widehat{\theta}_N = \arg\min_{\theta \in \Theta} \|\psi_N - \psi_N(\theta)\|_W, \tag{3.14}$$

where ψ_N are the moments computed from the sample as given in equation (3.12), $\psi_N(\theta)$ are computed from the observed sample of firms and workers as a function of θ, $\|\psi_N - \psi_N(\theta)\|_W = \left[(\psi_N - \psi_N(\theta))' W (\psi_N - \psi_N(\theta))\right]^{1/2}$ and W is a positive definite symmetric weight matrix. This minimum distance estimator finds the value of θ that best predicts the features of the data summarized by the moment function. For example, one can specify Ψ to summarize the overall sorting patterns and the many-to-one match moments used previously to illustrate the importance of using this information.

The next section presents conditions under which the estimator above is consistent and asymptotically normal.

3.3.2 Limit Properties

In this section, we outline a fairly standard set of convergence conditions on $\psi_N - \psi_N(\theta)$ and show that they imply limit properties for the estimator in equation (3.14). We will verify these conditions under large market asymptotics. These results are presented in the subsequent sections. We follow this organization to highlight the main ideas in the proof and clarify the contribution. We separate the conditions needed for consistency, which are weaker than those necessary for asymptotic normality of our estimator.

We require the following properties for the moment function at the population distribution of observable and unobservable traits.

Axiom 3.3 *(i) For any $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|\psi(\theta) - \psi(\theta_0)\|_W < \delta \Rightarrow \|\theta - \theta_0\| < \varepsilon$.*

(ii) $\psi(\theta)$ is continuously differentiable at θ_0 with an invertible Jacobian, $\psi'(\theta_0)$.

Part (i) assumes that the distance in the population $\|\psi(\theta) - \psi(\theta_0)\|_W$ is zero only if $\theta = \theta_0$. It implies that $\psi(\theta)$ identifies the parameter θ_0. Further, it requires that parameter

values outside a neighborhood of the true value cannot yield a distance arbitrarily close to 0.[14] This assumption, along with the convergence condition below, will guarantee consistency of our estimator. Part (ii) is used to prove that the estimator is asymptotically normal. The commonly made assumption that the Jacobian at the limit is invertible allows us to use Taylor approximations.

We will derive limiting properties of the estimator by showing conditions under which the following properties are satisfied:

Condition 1 *(i)* $(\psi_N - \psi(\theta_0)) - (\psi_N(\theta) - \psi(\theta))$ *converges in probability to 0, uniformly in* θ.

(ii) a. $\sqrt{N}(\psi_N - \psi_N(\theta_0))$ *converges in distribution to* $N(0, \Sigma)$

b. for every sequence $\{b_N\}$ *of positive numbers that converges to* 0,

$$\sqrt{N} \sup_{\|\theta-\theta_0\| \leq b_N} \|(\psi_N(\theta) - \psi(\theta)) - (\psi_N(\theta_0) - \psi(\theta_0))\|_\infty = o_p(1).$$

The first conditions would follow from a uniform law of large numbers. The second condition would follow from a central limit theorem and stochastic equicontinuity. These results are not obvious a priori because the matches depend on the composition of the entire market. The following sections prove these results under large market asymptotics. Along with Assumption 3.3, these conditions imply consistency and asymptotic normality of our estimator:

Theorem 3.2 *Suppose that the parameter space* Θ *is compact* θ_0 *lies in the interior of* Θ.

(i) If Assumption 3.3(i) and Condition 1(i) are satisfied, then $\widehat{\theta}_N$ *converges in probability to* θ_0.

(ii) If Assumption 3.3 and Condition 1 are satisfied, then

$$\sqrt{N}\left(\widehat{\theta}_N - \theta_0\right) \to N(0, \Omega),$$

$$\Omega = \left(\psi'(\theta_0)' C'C\psi'(\theta_0)\right)^{-1} \psi'(\theta_0) C\Sigma C'\psi'(\theta_0)' \left(\psi'(\theta_0)' C'C\psi'(\theta_0)\right)^{-1}$$

and $W = C'C$.

[14] A sufficient condition for this requirement is that $\theta \in \Theta$ is compact, $\psi(\theta)$ is continuous and $\psi(\theta) = \psi(\theta_0) \Rightarrow \theta = \theta_0$.

Proof. Part (i) follows from the arguments in Newey and McFadden (1994), Theorem 2.1. We use Theorem 3.3 in Pakes and Pollard (1989) to show part (ii). Let $G_N(\theta)$ (in the notation of Pakes and Pollard (1989)) by given by $(\psi_N - \psi_N(\theta))' C'$. Assumption 3.3(ii) and the definition of the estimator imply requirements (i), (ii) and (v) of Theorem 3.3 in Pakes and Pollard (1989). Requirement (iii) of Theorem 3.3 in Pakes and Pollard (1989) follows from Condition 1(ii)b. Requirements (iv) in Theorem 3.3 of Pakes and Pollard (1989) follow from Condition 1(ii)a. ■

This theorem shows that Assumption 3.3 and Condition 1 imply consistency and asymptotic normality in our setting. Therefore, the main difficulty in obtaining limit properties of our estimator is verifying Condition 1. This is not straightforward for two reasons. First, the triples (x_{2j-1}, x_{2j}, z_j) in the expression for our sample moments ψ_N in equation (3.12) are not sampled independently. This dependence occurs because their distribution is determined by the observed and unobserved characteristics of the entire sample. Second, equation (3.13) shows that $\psi_N(\theta) = \psi[\mu_{X_N}, \mu_{Z_N}](\theta)$ is also a function of the entire sample of observed characteristics.

To prove the required properties, we split the argument into two conceptually separate pieces. The first piece studies the distribution of sample moments ψ_N, and the second studies properties of the sample moment function $\psi_N(\theta)$. There are two reasons why this distinction helps analyze their limit distributions. First, the observed moments, ψ_N, are a function of both the sampled observed and unobserved characteristics because the realized assortative match depends on the latent indices of all agents in the market. On the other hand, $\psi_N(\theta)$, is a function only of observed traits because equation (3.13) shows that it is an integral with respect to the (known) distribution of unobservables. Second, ψ_N depends only on θ_0, while $\psi_N(\theta)$ is a stochastic process that must be studied uniformly in θ. The first reason complicates the analysis of the distribution of ψ_N, while the second reason complicates the analysis of $\psi_N(\theta)$.[15]

[15] An additional complication for analyzing the limit distribution of $\sqrt{N}(\psi_N - \psi_N(\theta))$ is that our convergence results must be joint with the empirical processes on X and Z.

Before proceeding, we formally show that is it sufficient to treat ψ_N and $\psi\left[m_X, m_Z\right](\theta)$ as scalars.

Proposition 3.2 *(i) Suppose that for each $k \in \{1, \ldots, K_\Psi\}$, the k-th component of $(\psi_N - \psi(\theta_0)) - (\psi_N(\theta) - \psi(\theta))$ converges in probability to 0, uniformly in θ, then $(\psi_N - \psi(\theta_0)) - (\psi_N(\theta) - \psi(\theta))$ converges in probability to 0, uniformly in θ.*

(ii) Suppose that for any $a \in \mathbb{R}^{K_\Psi}$, $\sqrt{N}(\psi_N - \psi_N(\theta_0)) \cdot a$ converges in distribution to $N(0, a'\Sigma a)$, and for every sequence $\{b_N\}$ of positive numbers that converges to 0,

$$\sqrt{N} \sup_{\|\theta - \theta_0\| \leq b_N} \left| \left[(\psi_N(\theta) - \psi(\theta)) - (\psi_N(\theta_0) - \psi(\theta_0)) \right] \cdot a \right| = o_p(1),$$

then, Condition 1(ii) is satisfied.

Proof. Part (i) follows from the definition of convergence in probability. To verify part (ii), note that Condition 1(ii) a. follows from the Cramer-Wold theorem. Condition 1(ii) b. follows from the fact that

$$\begin{aligned} & \sqrt{N} \sup_{\|\theta - \theta_0\| \leq b_N} \left\| (\psi_N(\theta) - \psi(\theta)) - (\psi_N(\theta_0) - \psi(\theta_0)) \right\|_\infty \\ = & \max_{a \in \left\{e_1, \ldots, e_{K_\Psi}\right\}} \sqrt{N} \sup_{\|\theta - \theta_0\| \leq b_N} \left| \left[(\psi_N(\theta) - \psi(\theta)) - (\psi_N(\theta_0) - \psi(\theta_0)) \right] \cdot a \right|, \end{aligned}$$

where $\{e_1, \ldots, e_{K_\Psi}\}$ are the standard basis vectors of $\mathbb{R}^{K_\Psi}$. ■

The following subsections derive regularity properties under which condition Condition 1 is satisfied, assuming that Ψ is a scalar-valued function. We first analyze the limiting properties of ψ_N, and then we analyze the properties of the function $\psi_N(\theta)$.

3.3.3 Convergence of the Data Generating Process

The first challenge is to study the large sample properties of the sample moments, ψ_N in equation (3.12). The primary technical difficulty arises from the dependence of the observed matches (X_1, X_2, Z) on the observable (and unobservable) characteristics of all agents in the market. We make progress by re-writing the sampling process as one in which the utilities u and v are drawn first. This allows us to condition on the matches on latent indices in the

data. The observed characteristics of the matched agents are then sampled conditional on these draws of the latent indices. This sampling process, although identical to drawing the characteristics directly from $\mu_{X,\varepsilon}$ and $\mu_{Z,\eta}$, allows for a more tractable approach to proving limit properties of the moments. The proof technique is based on using the triangular array structure implied by this process: the individual components of the triple (X_1, X_2, Z) are independent conditional on the indices drawn.

Specifically, our approach for obtaining large sample properties of ψ_N is based on the following observations. The observed characteristics X_1, X_2, and Z are a sample from $\mu_{X|v_1}$, $\mu_{X|v_2}$, and $\mu_{Z|u}$, where v_1, v_2,and u are the latent indices for these agents. The expected value of $\Psi(X_1, X_2, Z)$ given the latent indices is therefore $\tilde{\psi}[\mu_X, \mu_Z](v_1, v_2, u; \theta_0)$. Equation (3.13) shows that $\psi[\mu_X, \mu_Z](\theta_0)$ is the integral of $\tilde{\psi}[\mu_X, \mu_Z](v_1, v_2, u; \theta_0)$ over the population values of matched latent indices. This allows us to show that, ψ_N, which is the sample average of $\Psi(X_1, X_2, Z)$ over the matches in the data, approaches the population quantity $\psi[\mu_X, \mu_Z](\theta_0)$.

Below, we present assumptions under which we will prove our result.

Axiom 3.4 *(i) a.* $\tilde{\psi}[\mu_X, \mu_Z](v_1, v_2, u; \theta_0)$ *is Lipschitz continuous in* v_1, v_2 *and* u

b. The random variables ε *and* η *have continuous density with full support*

(ii) a. The derivative of $\tilde{\psi}[\mu_X, \mu_Z]\left(F^{-1}_{V;\theta}(q_1), F^{-1}_{V;\theta}(q_2), F^{-1}_{U;\theta}(q_3); \theta\right)$ *with respect to* $q = (q_1, q_2, q_3)$ *is bounded uniformly in* q, θ

b. The random variables ε *and* η *have continuous density with full support on* $\mathbb{R}$

c. The conditional distributions of X (respectively Z) given any v (respectively u) are not degenerate

Part (i) presents conditions under which we will show that ψ_N converges to $\psi[\mu_X, \mu_Z](\theta_0)$ in probability. Part (i) a. requires Lipschitz continuity of $\tilde{\psi}[\mu_X, \mu_Z]$. This regularity condition implies that the conditional expectation of Ψ is smooth with respect to the latent indices. A more primitive condition is presented in Appendix C.3.1, which shows that the condition

follows from bounds on the densities of X, ε and Z, η and their first derivative.[16] This regularity condition on the expectation of Ψ given the latent indices allows us to approximate the value of $\tilde{\psi}$ at the sampled latent indices for each of the matches. Part (ii) b. is a weak regularity condition on the distribution of the unobservables.

Part (ii) presents stronger assumptions, which we will use to derive the asymptotic distribution of $\sqrt{N}\left(\psi_N - \psi\left[\mu_X, \mu_Z\right](\theta_0)\right)$. Part (ii) a. is analogous to (i) a., but places stronger restrictions on the sensitivity of $\tilde{\psi}$ with respect to the quality of the match. The stronger assumption ensures that $\tilde{\psi}$ is not extremely sensitive to tail behavior. Parts (ii) b. and c. are weak regularity conditions.

Our first result shows that the empirical analog in equation (3.12) converges at the true parameter θ_0 to ψ.

Proposition 3.3 *(i) If Assumption 3.4(i) is satisfied, then $\psi_N - \psi(\theta_0)$ converges in probability to 0.*

(ii) If Assumption 3.4(ii) is satisfied, then for any μ_X- and μ_Z- Donsker classes Γ_X and Γ_Z of bounded functions on X and Z respectively,

$$\begin{bmatrix} \sqrt{N}\left(\psi_N - \psi(\theta_0)\right) \\ \sqrt{N}\left(\mu_{X_N} - \mu_X\right) \\ \sqrt{N/2}\left(\mu_{Z_N} - \mu_Z\right) \end{bmatrix},$$

where $\sqrt{N}\left(\mu_{X_N} - \mu_X\right)$ and $\sqrt{N/2}\left(\mu_{Z_N} - \mu_Z\right)$ are respectively empirical processes indexed by Γ_X and Γ_Z, converges to a mean zero Gaussian process (G_Ψ, G_X, G_Z) with covariance kernel V (given in Appendix B.2.1).

Proof. See Appendix B.2.1. ■

The result derives the large sample properties of $\psi_N - \psi\left[\mu_X, \mu_Z\right](\theta_0)$ based on the Assumption 3.4. The proof is based on studying the large sample properties of $E\left[\psi_N | v_1, \ldots, v_N, u_1, \ldots, u_{N/2}\right]$, the expectation of ψ_N given the sample of latent utilities

[16] See Assumption C.2 and Lemma C.12.

$v_1, \ldots, v_N$ and $u_1, \ldots, u_{N/2}$. Because the observed characteristics are drawn independently given these latent indices, we are able to characterize the large sample properties of $\psi_N - E[\psi_N | v_1, \ldots, v_N, u_1, \ldots, u_{N/2}]$. Next, we show that $E[\psi_N | v_1, \ldots, v_N, u_1, \ldots, u_{N/2}]$ approximates $\psi(\theta_0)$ by appealing to the regularity and smoothness conditions in Assumption 3.4. We do this by relying on smoothness of $\tilde{\psi}$ and noting that the empirical quantiles of the latent indices approximate the limit quantiles. Therefore, the key to the result is that the dependence across the observed matches is only through the latent indices, and that the matching is assortative on these indices.

3.3.4 Differentiability of the Moment Function

The large sample results on ψ_N require evaluating the moment function only at θ_0. To study the limit properties of the estimator defined in equation (3.14), we need to understand the properties of the sample moment function. In this section, we derive conditions under which this map is smooth. This will allow us to use a continuous mapping theorem and the functional delta method for our results.

The approach is based on separately analyzing the behavior of $\psi_N(\theta)$ away from the tails of the latent index distribution, then showing that the tails are negligible. This approach is convenient because deriving the asymptotic distribution of the tails is technically challenging. Specifically, we will show that the functional

$$\psi^{\delta}[\mu_X, \mu_Z](\theta) = \int_{\delta}^{1-\delta} \tilde{\psi}[\mu_X, \mu_Z]\left(F^{-1}_{V;\theta}(q), F^{-1}_{V;\theta}(q), F^{-1}_{U;\theta}(q)\right)(\theta)\, dq$$

is smooth in μ_X, μ_Z for all $\delta \in (0, 1/2)$. The integral above, when evaluated at $\delta = 0$, is equal to $\psi[\mu_X, \mu_Z](\theta)$ in equation. We require the following weak assumption on the distribution of unobservable traits:

Axiom 3.5 *(i) f_ε and f_η are bounded and have continuous, bounded first derivatives. Further, f_ε and f_η are bounded away from zero on any compact interval of $\mathbb{R}$.*

(ii) $h(X;\theta)$ and $g(Z;\theta)$ are uniformly μ_X- and μ_Z- integrable over all $\theta \in \Theta$

Part (i) imposes a weak regularity condition that allows us to show that the conditional distributions of X and Z given the latent indices v and u vary smoothly with θ, except at extreme quantiles of the latent index distribution. This assumption is satisfied for the most commonly used parametric forms in applied analysis. Part (ii) places a weak restriction on the tail behavior of $h(X;\theta)$ and $g(Z;\theta)$ by assuming that, uniformly across θ, with high probability, these random variables belong to a compact set.

To formally state our result on smoothness of ψ^δ, we need to define a metric in which to measure distances in the domain and range of ψ^δ. We use the Banach space of vector-valued functions of $\theta \in \Theta$ endowed with the sup-norm, denoted by L_∞^Θ, as the range. We use L_∞^Γ for the domain, which is the space of measures (m_X, m_Z) endowed with the sup-norm over the class of sets Γ. We let $\Gamma = \Gamma_X \cup \Gamma_Z$, where Γ_X is a class of sets that includes

1. $\Psi(x_1, x_2, z) f_\varepsilon\left(F_{V;\theta}^{-1}(q) - h(x_1;\theta)\right) f_\varepsilon\left(F_{V;\theta}^{-1}(q) - h(x_2;\theta)\right) f_\eta\left(F_{U;\theta}^{-1}(q) - g(z;\theta)\right)$ and $\Psi(x_1, x_2, z) f_\varepsilon'\left(F_{V;\theta}^{-1}(q) - h(x_1;\theta)\right) f_\varepsilon\left(F_{V;\theta}^{-1}(q) - h(x_2;\theta)\right) f_\eta\left(F_{U;\theta}^{-1}(q) - g(z;\theta)\right)$ indexed by (x_1, z, q, θ)[17]
2. $F_\varepsilon(v - h(x;\theta))$, $f_\varepsilon(v - h(x;\theta))$ and $f_\varepsilon'(v - h(x;\theta))$ indexed by (v, θ)
3. $1\{c_1 \leq x \leq c_2\}$ indexed by c_1 and c_2[18]

and Γ_Z is a class of sets that includes

1. $\Psi(x_1, x_2, z) f_\varepsilon\left(F_{V;\theta}^{-1}(q) - h(x_1;\theta)\right) f_\varepsilon\left(F_{V;\theta}^{-1}(q) - h(x_2;\theta)\right) f_\eta\left(F_{U;\theta}^{-1}(q) - g(z;\theta)\right)$ and $\Psi(x_1, x_2, z) f_\varepsilon\left(F_{V;\theta}^{-1}(q) - h(x_1;\theta)\right) f_\varepsilon\left(F_{V;\theta}^{-1}(q) - h(x_2;\theta)\right) f_\eta'\left(F_{U;\theta}^{-1}(q) - g(z;\theta)\right)$ indexed by (x_1, x_2, q, θ)
2. $F_\eta(u - g(z;\theta))$, $f_\eta(u - g(z;\theta))$ and $f_\eta'(u - g(z;\theta))$ indexed by (u, θ)
3. $1\{c_1 \leq z \leq c_2\}$ indexed by c_1 and c_2

[17] Since the functions considered are symmetric in x_1 and x_2, we have implicitly also included the analogous class of functions, indexed by (x_2, z, q, θ).

[18] If a and b are vectors, we say that $a \leq b$ if each element of a is weakly less than each element of b.

Therefore, we will consider smoothness of the map $\psi^\delta : L_\infty^\Gamma \to L_\infty^\Theta$. The class Γ defines a norm in which we measure distances between two pairs (m_X, m_Z) and (m'_X, m'_Z). The first two groups of sets in Γ_X and Γ_Z arise from Taylor expansions of terms in the expression for ψ^δ. The last two sets are intersections of half-spaces. To use the continuous mapping theorem and the functional delta method, we will need to ensure that the empirical measures μ_{X_N} and μ_{Z_N} converge to the population measures with distance measured in this norm. The required properties on the primitives to ensure that Γ_X and Γ_Z are respectively μ_X- and μ_Z- Donsker classes are stated formally in the Online Appendix (Proposition C.2).

We are now ready to state the main results in this section.

Proposition 3.4 *If Assumption 3.5 is satisfied, then for each* $\delta \in \left(0, \frac{1}{2}\right)$, $\psi^\delta : L_\infty^\Gamma \to L_\infty^\Theta$ *is Hadamard differentiable tangentially to the space of bounded uniformly continuous functions at* (μ_X, μ_Z). *The Hadamard derivative at* (μ_X, μ_Z) *in the direction* (G_X, G_Z) *is* $\nabla_{(G_X, G_Z)} \psi^\delta [\mu_X, \mu_Z]$ *(given in Appendix C.2.2).*

Proof. See Appendix B.2.2 for a sketch of the proof and Appendix C.2.2 for details. ■

This result formalizes the idea that the small perturbations of the measures μ_X, μ_Z result in small deviations in the value of the moments (outside the tails) as a function of θ. This is useful because we expect the empirical distributions of X and Z to be close to μ_X and μ_Z in a large sample. Assuming that tails are negligible, the result implies that the moment function in a large sample approximates the population moment function. The next section uses this result and Proposition 3.3 to verify Condition 1.

3.3.5 Verifying Condition 1

We now put together the results in the previous sections to show that Condition 1 is satisfied. First we show part (i), which implies consistency of the estimator by Theorem 3.2(i). We will use a continuous mapping theorem and the following assumption for this result:

Axiom 3.6 (i) Γ_X *and* Γ_Z *are respectively* μ_X*- and* μ_Z*- Glivenko Cantelli.*

This assumption implies that the expectations of functions in Γ_X and Γ_Z evaluated at the empirical measures μ_{X_N} and μ_{Z_N} respectively converge (in probability) to the population values. Further, the Glivenko-Cantelli theorem implies that the convergence is uniform over all functions in these classes. The assumption is satisfied under weak conditions on the elements of Γ_X and Γ_Z.[19] We now formally state that Condition 1(i) is satisfied for our model and sketch the proof.

Proposition 3.5 (i) *If Assumptions 3.4(i), 3.5 and 3.6(i) are satisfied, then $\psi_N - \psi_N(\theta)$ converges in probability to $\psi - \psi(\theta)$, uniformly in θ.*

Proof. See Online Appendix C.2.3, part (i). ■

The result shows that the difference between the empirical distance function $\psi_N - \psi_N(\theta)$ and the population analog $\psi - \psi(\theta)$ converges to zero (in probability) as the sample increases in size. The proof is proceeds by using the triangle inequality to observe that this difference is at most $|\psi_N - \psi| + |\psi_N(\theta) - \psi(\theta)|$. Proposition 3.3 implies that the first term, which measures the distance between the empirical and population values of the moments, converges in probability to zero. The second term, which measures the distance of the sample moment function to the population function at θ, is $\psi^0[\mu_{X_N}, \mu_{Z_N}](\theta) - \psi^0[\mu_X, \mu_Z](\theta)$ by definition. To show that this term also converges in probability to zero (uniformly in θ), we approximate ψ^0 with ψ^δ. Specifically, $\psi_N(\theta)$ and $\psi(\theta)$ can be approximated by $\psi^\delta[\mu_{X_N}, \mu_{Z_N}](\theta)$ and $\psi^\delta[\mu_X, \mu_Z](\theta)$, where the error is of the order of δ because Ψ is bounded. Proposition 3.4 and Assumption 3.6 imply, by the continuous mapping theorem, that $\psi^\delta[\mu_{X_N}, \mu_{Z_N}](\theta)$ converges in probability to $\psi^\delta[\mu_X, \mu_Z](\theta)$ uniformly in θ. Together, these observations imply the result.

The approach to a limit theorem that verifies Condition 1(ii) is similar in spirit, but technically more demanding. Proposition 3.3 provides a result for the term $\sqrt{N}(\psi_N - \psi)$. Our next challenge is to prove a limit theorem for $\sqrt{N}(\psi_N(\hat{\theta}) - \psi(\hat{\theta}))$, where $\hat{\theta}$ is our estimator. We do this by approximating $\sqrt{N}(\psi_N(\hat{\theta}) - \psi(\hat{\theta}))$ with

[19] Proposition C.2 formally states conditions on primitives under which Γ_X and Γ_Z are Donsker classes.

$\sqrt{N}\left(\psi_N^\delta(\theta_0) - \psi^\delta(\theta_0)\right)$. The functional delta method and Proposition 3.4 imply that asymptotic distribution of $\sqrt{N}\left(\psi^\delta(\theta_0) - \psi_N^\delta(\theta_0)\right)$ is given by $\nabla_G\psi^\delta(\theta_0) = \left(\nabla\psi^\delta \circ G\right)(\theta_0)$, where G is a mean zero Gaussian process on L_∞^Γ. The remaining term is the approximation error $\left(\nabla_G\psi^\delta - \nabla_G\psi^0\right)(\theta_0)$. Therefore, we need to ensure that the errors in approximating $\nabla_G\psi^0(\theta_0)$ with $\nabla_G\psi^\delta(\theta_0)$ and approximating $\sqrt{N}(\psi_N(\theta) - \psi(\theta))$ in a neighborhood of θ_0 with $\sqrt{N}\left(\psi_N^\delta(\theta_0) - \psi^\delta(\theta_0)\right)$ are negligible. Ensuring that these errors do not affect the limit distribution of $\sqrt{N}((\psi_N - \psi) - (\psi_N(\theta) - \psi(\theta)))$ requires tighter controls of the tails than our consistency result. Specifically, the limit theorem requires us to replace Assumption 3.6(i) with the following stronger requirement:

Axiom 3.7 (ii) *a.* Γ_X *and* Γ_Z *are respectively* μ_X*- and* μ_Z*- Donsker.*

b. for every sequence $\{b_N\}$ *of positive numbers that converges to* 0*,*

$$\sqrt{N}E\sup_{\|\theta-\theta_0\|\leq b_N}\left|(\psi_N(\theta) - \psi(\theta)) - \left(\psi_N^\delta(\theta) - \psi^\delta(\theta)\right)\right|$$

converges to zero as $\delta \to 0$ *and* $N \to \infty$

c. for fixed $\delta \in \left(0, \frac{1}{2}\right)$ *and every sequence* $\{b_N\}$ *of positive numbers that converges to* 0*,*

$$\sup_{\|\theta-\theta_0\|\leq b_N}\left|\nabla_G\psi^\delta(\theta) - \nabla_G\psi^\delta(\theta_0)\right|$$

converges in probability to zero as $N \to \infty$

d. $\left(\nabla_G\psi^\delta - \nabla_G\psi^0\right)(\theta_0)$ *converges in probability to zero as* $\delta \to 0$*.*

Part a. strengthens Assumption 3.6(i) to allow a functional central limit theorem over the classes Γ_X and Γ_Z. Parts b. and d. are technical assumptions that ensure that tails are negligible. Part b. controls the rate at which the dependence of the moment function on the tails vanishes with the sample size. Part d. assumes that tails have a negligible contribution to the dependence of the moment function on perturbations of the data. Part c. assumes that the process $\nabla_G\psi^\delta(\theta)$ is well-behaved in a neighborhood of θ_0. For completeness, the Online Appendix derives primitive conditions under which each of these requirements are satisfied. Specifically, Theorem C.2.2 shows that smoothness conditions and bounds on the tail behavior of the primitives imply these requirements. Assumption

c. is relatively straightforward to verify and is based on showing that $\nabla_G \psi^\delta(\theta)$ have sample paths continuous in θ by bounding the L^2 covering numbers of the related Gaussian process. Assumption d. follows from showing that an upper bound on the variance of $\left(\nabla_G \psi^\delta - \nabla_G \psi^0\right)(\theta_0)$ converges to 0 as $\delta \to 0$. Verifying assumption b. is the most difficult technical aspect of proving our limit theorem and requires relatively novel proof techniques.

The difficulty in verifying assumption b follows from the fact that $\sqrt{N}\left(\psi_N(\theta) - \psi(\theta)\right)$ is a nonlinear function of the empirical measures (μ_{X_N}, μ_{Z_N}). While the functional delta method is a conceptually straightforward approach to proving a limit theorem for $\sqrt{N}\left(\psi_N^\delta(\theta) - \psi^\delta(\theta)\right)$ with $\delta \in (0,1)$, showing that the tails are negligible requires a proof by first principles. Although direct computations play a large part in this proof, the conceptual core is a modification of the method of chaining with adaptive truncation exposited by Pollard (2002), where it is used to prove Ossiander's bracketing limit theorem for empirical processes. Our proof technique follows a similar approach as Pollard (2002) by similarly approximating Θ using finite subsets of increasing size and similar truncation techniques. After a suitable truncation, the moment generating function of the increments of an empirical process can be bounded using techniques that apply to sums of independent random variables. Because the increments of $(\psi_N(\theta) - \psi(\theta))$ have no simple expression, we use the concentration of measure inequality of Boucheron *et al.* (2003) in order to get the needed bound on the moment generating function. This application of an abstract concentration of measure inequalities within the broader context of a chaining argument may be a more generally useful technique for proving functional limit theorems. This approach is necessary due to the dependent data nature of our problem, which makes standard empirical process techniques for i.i.d. data inapplicable. This feature of our model may be shared with other contexts such as network formation models.

The control of tail behavior implied by these results allow us to verify Condition 1(ii). Formally, we have:

Proposition 3.6 (ii) *If Assumptions 3.4(ii), 3.5 and 3.6(ii) are satisfied, then Condition 1(ii) is satisfied.*

Proof. See Appendix C.2.3, part (ii). ■

As discussed earlier, the basic ideas are similar to the consistency result proved earlier, with a more technically demanding method for handling the approximation in the tails. Proposition 3.5 shows Condition 1 for our model. Therefore, we can use Theorem 3.2 to assure consistency and asymptotic normality of the minimum distance estimator.

3.4 Conclusion

This paper provides results on the identification and estimation of preferences from data from a matching market with positive assortative matching on a latent index when data only on matches are observed. Our results apply to both transferable and non-transferable utility models of matching. We show that using information available in many-to-one matching is necessary and sufficient for non-parametric identification if data on a single large market is observed. These identification results use insights from the analysis of non-linear measurement error models. Intuitively, the observable characteristics of the multiple agents with the same match partner can be seen as noisy measures of the quality of the agents in the match.

We then prove consistency and $\sqrt{N}-$asymptotic normality of an estimator for a parametric class of models. Our limit theorems are based on several insights in this model. First, we use the fact that the matches are determined by the latent indices and that the observables are conditionally independent given these indices. Second, we show that the moment function is smooth in the distribution of observables, except at the extreme quantiles of the latent index. Third, we show that approximating this function by ignoring the tails has a negligible effect on the asymptotic distribution of the estimator using a general concentration of measure inequality for dependent data. Finally, we present Monte Carlo evidence on a simulation-based estimator.

There are several avenues for future research on both identification and estimation for similar models. While we show that it is necessary to use information from many-to-one matching for identification with data on a single large market, it may also be possible to

use variation in the characteristics of participants across markets for identification. This can be particularly important for the empirical study of marriage markets. Our results are also restricted to a single latent index model on each side of the market. Extending this domain of preferences is particularly important. A treatment of heterogeneous preferences on both sides of the market may be of particular interest, but it is likely technically challenging. It may be particularly difficult to analyze both transferable and non-transferable utility models in a single framework. Finally, we have also left the exploration of computationally more tractable estimators for future research.

References

Adrian, T., Etula, E. and Muir, T. (2014). Financial intermediaries and the cross-section of asset returns. *Journal of Finance*, **69** (6), 2557– 2596.

Agarwal, N. (2015). An empirical model of the medical match. *American Economic Review*, **105** (7), 1939–1978.

— (2017). Policy analysis in matching markets. *American Economic Review, Papers and Proceedings*, **107** (5), 246–250.

— and Diamond, W. (2014). Identification and estimation in two-sided matching markets. *Cowles Foundation Discussion Paper No. 1905*.

Allen, F. and Gale, D. (1988). Optimal security design. *Review of Financial Studies*, **1** (3), 229–263.

Anderson, G. F. (1996). What Does Not Explain the Variation in the Direct Costs of Graduate Medical Education. *Academic Medicine*, **71** (2), 164–9.

Azevedo, E. M. and Budish, E. (2017). Strategyproofness in the Large. *Chicago Booth, mimeo*.

Bailey, J. M. (2010). *Health Care Reform, What's in it? Rural Communities and Medical Care.* Tech. rep., Center for Rural Affairs.

Bajari, P., Fox, J. T., Kim, K. I. and Ryan, S. P. (2009). The Random Coefficients Logit Model Is Identified.

Baker, M., Bradley, B. and Taliaferro, R. (2014). The low-risk anomaly: A decomposition into micro and macro effects. *Financial Analysts Journal*, **70** (2), 43–58.

—, Hoeyer, M. F. and Wurgler, J. (2016). The risk anomaly tradeoff of leverage. *working paper*.

Banerjee, A. and Maskin, E. (1996). A walrasian theory of money and barter. *Quarterly Journal of Economics*, **111** (4), 955–1005.

Bansal, R. and Coleman, W. J. (1996). A monetary explanation of the equity premium, term premium, and risk-free rate puzzles. *Journal of Political Economy*, **104** (6), 1135–1171.

Barnish, K., Miller, S. and Rushmore, M. (1997). The new leveraged loan syndication market. *Journal of Applied Corporate Finance*, **10** (1), 79–88.

Becker, G. S. (1973). A Theory of Marriage: Part I. *Journal of Political Economy*, **81** (4), 813 – 46.

— (1975). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, 2nd ed*. New York: National Bureau of Economic Research, Inc.

Berentsena, A., Camera, G. and Waller, C. (2007). Money, credit, and banking. *Journal of Economic Theory*, **135** (1), 171–95.

Berg, T. and Gider, J. (forthcoming). What explains the difference in leverage between banks and non-banks? *Journal of Financial and Quantitative Analysis*.

Bernanke, B., Bertaut, C. C., Demarco, L. P. and Kamin, S. B. (2011). International capital flows and the return to safe assets in the united states, 2003-2007. *Financial Stability Review*, **15**, 13–26.

Berry, S. T., Levinsohn, J. and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, **63** (4), 841 – 890.

— and Pakes, A. (2007). The Pure Characteristics Demand Model. *International Economic Review*, **48** (4), 1193–1225.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of Business*, **45** (3), 444–455.

Blundell, R. and Powell, J. (2003). Endogeneity in Nonparametric and Semiparametric Regression Models. In M. Dewatripont, L. Hansen and S. Turnovky (eds.), *Advances in Economics and Econometrics*, *8*, pp. 312–357.

Boucher, V. and Mourifie, I. (2012). My Friend Far Far Away: Asymptotic Properties of Pairwise Stable Networks. *SSRN Electronic Journal*.

Boucheron, S., Lugosi, G. and Massart, P. (2003). Concentration inequalities using the entropy method. *The Annals of Probability*, **31** (3), 1583–1614.

Boyd, D., Lankford, H., Loeb, S. and Wyckoff, J. (2013). Analyzing the Determinants of the Matching Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets. *Journal of Labor Economics*, **31** (1), 83–117.

Breusch, T. S. and Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, **47** (5), 1287 – 94.

Bulow, J. and Levin, J. (2006). Matching and Price Competition. *The American Economic Review*, **96** (3), 652 – 668.

Caballero, R. J. and Farhi, E. (forthcoming). The safety trap. *Review of Economic Studies*.

Calvo, G. (1988). Servicing the public debt: The role of expectations. *American Economic Review*, **78** (4), 647–61.

Chandrasekhar, A. G. and Jackson, M. O. (2015). A Network Formation Model Based on Subgraphs.

Chiappori, P.-A., Iyigun, M. and Weiss, Y. (2009). Investment in Schooling and the Marriage Market. *American Economic Review*, **99** (5), 1689–1713.

— and Komunjer, I. (2008). Correct Specification and Identification of Nonparametric Transformation Models. *Columbia University, mimeo*.

—, Oreffice, S. and Quintana-Domeque, C. (2012). Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market. *Journal of Political Economy*, **120** (4), 659–695.

— and Salanié, B. (2015). The Econometrics of Matching Models. *Journal of Economic Literature*, **54** (3), 832–861.

—, Salanié, B. and Weiss, Y. (2015). Partner Choice and the Marital College Premium. *CEPR Discussion Paper No. DP10403*.

Chodorow-Reich, G. (2014). Effects of unconventional monetary policy on financial institutions. *Brookings Papers on Economic Activity*, pp. 155–204.

Choo, E. and Siow, A. (2006). Who Marries Whom and Why. *Journal of Political Economy*, **114** (1), 175–201.

Christakis, N. A., Fowler, J. H., Imbens, G. W. and Kalyanaraman, K. (2010). An Empirical Model for Strategic Network Formation.

Clark, S. (2006). The Uniqueness of Stable Matchings. *Contributions in Theoretical Economics*, **6** (1), 1–28.

Csorgo, M. and Revesz, P. (1978). Strong Approximations of the Quantile Process. *The Annals of Statistics*, **6** (4), 882–894.

Dang, T. V., Gorton, G. and Holmstrom, B. (2013). Ignorance, debt and financial crises. *working paper*.

—, —, — and Ordonez, G. (2017). Banks as secret keepers. *American Economic Review*, **107** (4), 1005–1029.

DeAngelo, H. and Stulz, R. M. (2015). Liquid-claim production, risk management, and bank capital structure: Why high leverage is optimal for banks. *Journal of Financial Economics*, **116** (2), 219–236.

DeMarzo, P. (2005). The pooling and tranching of securities: A model of informed intermediation. *Review of Financial Studies*, **18** (1), 1– 35.

— and Duffie, D. (1999). A liquidity-based model of security design. *Econometrica*, **67** (1), 65–99.

Diamond, D. (1984). Financial intermediation and delegated monitoring. *Review of Econom*, **51** (3), 393–414.

— and Dybvig, P. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, **91** (3), 401–19.

— and Rajan, R. (2001). Liquidity risk, liquidity creation, and financial fragility: A theory of banking. *Journal of Political Economy*, **109** (2), 287–327.

Dudley, R. M. (2014). *Uniform Central Limit Theorems*. Cambridge University Press.

Dunford, N. and Schwartz, J. T. (1971). *Linear Operators*. New York: Wiley.

Ekeland, I., Heckman, J. J. and Nesheim, L. (2004). Identification and Estimation of Hedonic Models. *Journal of Political Economy*, **112** (S1).

Fox, J. T. (2009). Structural Empirical Work Using Matching Models. In S. N. Durlauf and L. E. Blume (eds.), *New Palgrave Dictionary of Economics*, online edn.

— (2010). Identification in matching games. *Quantitative Economics*, **1** (2), 203–254.

— (forthcoming). Estimating Matching Games with Transfers. *Quantitative Economics*.

Frazzini, A. and Pedersen, L. (2014). Betting against beta. *Journal of Financial Economics*, **111** (1), 1–25.

Friedman, M. and Schwartz, A. (1963). *A Monetary History of the United States*. Princeton University Press.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, **69** (1), 9–15.

Galichon, A., Kominers, S. D. and Weber, S. (2014). Costly Concessions: An Empirical Framework for Matching with Imperfectly Transferable Utility.

— and Salanie, B. (2012). Cupid's Invisible Hand: Social Surplus and Identification in Matching Models. *Columbia University, mimeo*.

Gilchrist, S. and Zakrajsek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, **102** (4), 1692–1270.

Goldstein, I. and Pauzner, A. (2005). Demand-deposit contracts and the probability of bank runs. *Journal of Finance*, **60** (3), 1293–1327.

Gorton, G. and Pennacchi, G. (1990). Financial intermediaries and liquidity creation. *Journal of Finance*, **45** (1), 49–71.

Graham, B. S. (2017). An Econometric Model of Link Formation with Degree Heterogeneity. *Econometrica*, **85** (4), 1033–1063.

Hanson, S. G., Shleifer, A., Stein, J. and Vishny, R. W. (2015). Banks as patient fixed-income investors. *Journal of Financial*, **117** (3), 449–469.

Hsieh, Y.-W. (2011). Understanding Mate Preferences from Two-Sided Matching Markets: Identification, Estimation and Policy Analysis. *University of Southern California, mimeo*.

Hu, Y., McAdams, D. and Shum, M. (2013). Identification of first-price auctions with non-separable unobserved heterogeneity. *Journal of Econometrics*, **174** (2), 186–193.

— and Schennach, S. M. (2008). Instrumental Variable Treatment of Nonclassical Measurement Error Models. *Econometrica*, **76** (1), 195–216.

— and Shum, M. (2012). Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics*, **171** (1), 32–44.

Huang, J.-Z. and Huang, M. (2012). How much of the corporate-treasury yield spread is due to credit risk? *Review of Asset Pricing Studies*, **2** (2), 153–202.

Imbens, G. W. (2007). Nonadditive Models with Endogenous Regressors. In R. Blundell, W. Newey and T. Persson (eds.), *Advances in Economics and Econometrics, Vol III, 2*, pp. 17 – 46.

— and Newey, W. K. (2009). Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. *Econometrica*, **77** (5), 1481 – 1512.

Immorlica, N. and Mahdian, M. (2005). Marriage, honesty, and stability. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Society for Industrial and Applied Mathematics, pp. 53–62.

Innes, R. (1990). Limited liability and incentive contracting with ex-ante action choices. *Journal of Economic Theory*, **52** (1), 45–67.

Jiang, Y. (2016). Matching Games with Unobserved Heterogeneity: A Structural Analysis of Online Labor Markets.

Johnson, S. G. (2011). The NLopt nonlinear-optimization package.

Jung et.al. v AAMC et.al. (2002). Class Action Complaint, No. 02-CV-00873, D.D.C. May 5, 2002.

Kahn, M. J., Markert, R. J., Lopez, F. A., Specter, S., Randall, H. and Krane, N. K. (2006). Is medical student choice of a primary care residency influenced by debt? *MedGenMed : Medscape general medicine*, **8** (4), 18.

Kamada, Y. and Kojima, F. (2010). Improving Efficiency in Matching Markets with Regional Caps: The Case of the Japan Residency Matching Program.

Kashyap, A., Rajan, R. and Stein, J. C. (2002). Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking. *Journal of Finance*, **57** (1), 33–73.

Kelso, A. S. J. and Crawford, V. P. (1982). Job Matching, Coalition Formation, and Gross Substitutes. *Econometrica*, **50** (6), 1483 – 1504.

Kojima, F. (2007). Matching and Price Competition: Comment. *American Economic Review*, **97** (3), 1027 – 1031.

—, Pathak, P. and Roth, A. E. (2013). Matching with Couples: Stability and Incentives in Large Markets. *Quarterly Journal of Economics*, **128** (4), 1585–1632.

— and Pathak, P. A. (2009). Incentives and Stability in Large Two-Sided Matching Markets. *American Economic Review*, **99** (3), 608 – 27.

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer.

Krasnokutskaya, E. (2011). Identification and Estimation of Auction Models with Unobserved Heterogeneity. *Review of Economic Studies*, **78** (1), 293–327.

Krishnamurthy, A. and He, Z. (2013). Intermediary asset pricing. *American Economic Review*, **103** (2), 732–70.

— and Vissing-Jorgensen, A. (2011). The effects of quantitative easing on interest rates: Channels and implications for policy. *Brookings Papers on Economic Activity*, pp. 215–287.

— and — (2012). The aggregate demand for treasury debt. *Journal of Political Economy*, **120** (2), 233–267.

Krugman, P. (1998). It's baaack: Japan's slump and the return of the liquidity trap. *Brookings Papers on Economic Activity*, **29** (1998-2), 137–206.

Lacker, J. M. and Weinburg, J. A. (1989). Optimal contracts under costly state falsification. *Journal of Political Economy*, **97** (6), 1345–1363.

Lagos, R. and Wright, R. (2005). A unified framework for monetary theory and policy analysis. *Journal of Political Economy*, **113** (3), 463–484.

Lester, B., Postlewaite, A. and Wright, R. (2012). Information, liquidity, asset prices, and monetary policy. *Review of Economic Studies*, **79** (3), 1209–38.

Leung, M. P. (2015). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics*, **188** (1), 182–195.

Logan, J. A., Hoff, P. D. and Newton, M. A. (2008). Two-Sided Estimation of Mate Preferences for Similarities in Age, Education, and Religion. *Journal of the American Statistical Association*, pp. 559 – 569.

Loutskina, E. (2011). The role of securitization in bank liquidity and funding management. *Journal of Financial Economics*, **100** (3), 663–684.

— and Strahan, P. E. (2009). Securitization and the declining impact of bank finance on loan supply: Evidence from mortgage originations. *Journal of Finance*, **64** (2), 861–889.

Manski, C. F. (2007). *Identification for Prediction and Decision*. Cambridge, USA: Harvard University Press.

— and Pepper, J. V. (1998). Monotone Instrumental Variables with an Application to the Returns to Schooling.

Mattner, L. (1993). Some Incomplete But Boundedly Complete Location Families. *The Annals of Statistics*, **21** (4), 2158–2162.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration. *Econometrica*, **57** (5), 995 – 1026.

Menzel, K. (2011). Robust Decisions For Incomplete Structural Models Of Social Interactions. *New York University, mimeo*.

— (2012). Inference for large games with exchangeable players.

— (2015). Large Matching Markets As Two-Sided Demand Systems. *Econometrica*, **83** (3), 897–941.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research, Inc.

Mortensen, D. T. and Pissarides, C. A. (1994). Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies*, **61** (3), 397 – 415.

Nagel, S. (2016). The liquidity premium of near-money assets. *Quarterly Journal of Economics*, **131** (4), 1927–1971.

Neely, C. (2011). The large-scale asset purchases had large international effects. *working paper*.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, **4**, 2111–2245.

Newhouse, J. P. and Wilensky, G. R. (2001). Paying For Graduate Medical Education: The Debate Goes On. *Health Affairs*, **20** (2), 136–147.

Niederle, M. (2007). Competitive Wages in a Match with Ordered Contracts. *American Economic Review*, **97** (5), 1957–1969.

— and Roth, A. E. (2003). Relationship Between Wages and Presence of a Match in Medical Fellowships. *JAMA, Journal of the American Medical Association*, **290** (9).

— and — (2009). The Effects of a Centralized Clearinghouse on Job Placement, Wages, and Hiring Practices. In David Autor (ed.), *NBER Chapters*, pp. 235 – 271.

— and Yariv, L. (2009). Decentralized Matching with Aligned Preferences. *Stanford University, mimeo*.

Pakes, A. and Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, **57** (5), 1027 – 57.

Pencavel, J. (1998). Assortative Mating by Schooling and the Work Behavior of Wives and Husbands. *American Economic Review*, **88** (2), 326–29.

Petrin, A. K. and Train, K. E. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, **47** (1), 3–13.

Polachek, S. W. (2008). Earnings Over the Life Cycle: The Mincer Earnings Function and its Applications. *Foundation and Trends in Microeconomics*, **4** (3), 165–272.

Pollard, D. (2002). Maximal inequalities via bracketing with adaptive truncation. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, **38** (6), 1039–1052.

Postel-Vinay, F. and Robin, J.-M. (2002). Equilibrium Wage Dispersion with Worker and Employer Heterogeneity. *Econometrica*, **70** (6), 2295 – 2350.

Rabinowitz, H. K., Diamond, J. J., Markham, F. W. and Wortman, J. R. (2008). Medical school programs to increase the rural physician supply: a systematic review and projected impact of widespread replication. *Academic Medicine*, **83** (3), 235–43.

Rao, P. B. L. S. (1992). *Identifiability in Stochastic Models: Characterization of Probability Distributions*. New Delhi, India: Academic Press, Inc.

Reschovsky, J. D. and Staiti, A. B. (2005). Physician Incomes in Rural and Urban America. *Issue brief (Center for Studying Health System Change)*, (92), 1–4.

Rosen, S. (1972). Learning and experience in the labor market. *Journal of Human Resources*, **7** (3), 326–342.

— (1983). Specialization and Human Capital. *Journal of Labor Economics*, **1** (1), 43 – 49.

— (1987). The theory of equalizing differences. In O. Ashenfelter and R. Layard (eds.), *Handbook of Labor Economics*, *Handbook of Labor Economics*, vol. 1, *12*, Elsevier, pp. 641–692.

Rosenblatt, R. A. and Hart, L. G. (2000). Physicians and Rural America. *The Western journal of medicine*, **173** (5), 348–51.

Rosenthal, M. B., Zaslavsky, A. and Newhouse, J. P. (2005). The geographic distribution of physicians revisited. *Health services research*, **40** (6 Pt 1), 1931–52.

Roth, A. E. (1984a). Stability and Polarization of Interests in Job Matching. *Econometrica*, **52** (1), 47 – 57.

— (1984b). The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *Journal of Political Economy*, **92** (6), 991–1016.

— (1986). On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets. *Econometrica*, **54** (2), 425 – 27.

— and Peranson, E. (1999). The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Review*, **89** (4), 748–780.

— and Sotomayor, M. A. O. (1992). *Two-Sided Matching*. Cambridge University Press.

— and Vande Vate, J. H. (1990). Random Paths to Stability in Two-Sided Matching. *Econometrica*, **58** (6), 1475 – 80.

— and Xing, X. (1994). Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions. *American Economic Review*, **84** (4), 992–1044.

Rowan, T. H. (1990). *Functional Stability Analysis Of Numerical Algorithms*. Ph.D. thesis.

Schennach, S. M. (2004). Estimation of Nonlinear Models with Measurement Error. *Econometrica*, **72** (1), 33–75.

Shapley, L. S. and Shubik, M. (1971). The assignment game I: The core. *International Journal of Game Theory*, **1** (1), 111–130.

Shi, X. and Shum, M. (2014). On the Empirical Content of the Beckerian Marriage Model. *Caltech, mimeo*.

Shimer, R. and Smith, L. (2000). Assortative Matching and Search. *Econometrica*, **68** (2), 343 – 370.

Shiu, J.-L. and Hu, Y. (2013). Identification and estimation of nonlinear dynamic panel data models with unobserved covariates. *Journal of Econometrics*, **175** (2), 116–131.

Signer, M. M. (2012). *2012 Main Residency Match and SOAP*. Tech. rep., National Resident Matching Program.

Simsek, A. (2013). Belief disagreements and collateral constraints. *Econometrica*, **81** (1), 1–53.

Sorensen, M. (2007). How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital. *Journal of Finance*, **62** (6), 2725 – 2762.

Sotomayor, M. (1999). The Lattice Structure of the Set of Stable Outcomes of the Multiple Partners Assignment Game. *International Journal of Game Theory*, **28** (4), 567 – 583.

Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65** (3), 557 – 586.

Stein, J. C. (2012a). Evaluating large scale asset purchases. *Speech delivered at the Brookings Institution, Washington, October 11.*

— (2012b). Monetary policy as financial-stability regulation. *Quarterly Journal of Economics*, **127** (1), 57–95.

Stern, S. (2004). Do Scientists Pay to Be Scientists? *Management Science*, **50** (6), 835–853.

Stock, J. H., Wright, J. H. and Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, **20** (4), 518–29.

Sunderam, A. (2015). Money creation and the shadow banking system. *Reveiw of Financial Studies*, **28** (4), 939–977.

Talley, R. C. (1990). Graduate medical education and rural health care. *Academic medicine*, **65** (12 Suppl), S22–5.

Teulings, C. N. (1995). The Wage Distribution in a Model of the Assignment of Skills to Jobs. *Journal of Political Economy*, **103** (2), 280 – 315.

Thomas Lemieux (2006). The Mincer Equation Thirty Years After Schooling, Experience, and Earnings. In *Jacob Mincer: A Pioneer of Modern Labor Economics*.

Uetake, K. and Watanabe, Y. (2013). Entry by Merger: Estimates from a Two-Sided Matching Model with Externalities. *Yale University, mimeo*.

van der Vaart, A. and Wellner, J. (2000). *Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics)*. Springer, corr. 2. p edn.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

Vissing, A. (2016). *One-to-Many Matching with Complementary Preferences: An Empirical Study of Natural Gas Lease Quality and Market Power*. Tech. rep., Duke University, Durham.

Weeks, W. B. and Wallace, A. E. (2008). Rural-urban differences in primary care physicians' practice patterns, characteristics, and incomes. *The Journal of rural health : official journal of the American Rural Health Association and the National Rural Health Care Association*, **24** (2), 161–70.

Weitzman, M. L. (1974). Prices vs. Quantities. *Review of Economic Studies*, **41** (4), 477 – 91.

Wetz, R. V., Seelig, C. B., Khoueiry, G. and Weiserbs, K. F. (2010). Out-of-Match Residency Offers: The Possible Extent and Implications of Prematching in Graduate Medical Education. *Journal of Graduate Medical Education*, **2** (3), 327–33.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, **48** (4), 817 – 38.

Williamson, S. and Wright, R. (1994). Barter and monetary exchange under private information. *American Economic Review*, **84** (1), 104–23.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: The MIT Press.

Yang, S., Chen, Y. and Allenby, G. M. (2003). Bayesian Analysis of Simultaneous Demand and Supply. *Quantitative Marketing and Economics*, **1** (3), 251–275.

Youngclaus, J. and Fresne, J. (2012). *Trends in Cost and Debt at U.S. Medical Schools Using a New Measure of Medical School Cost of Attendance*. Tech. rep., Association of Americal Medical Colleges, Washington.

Appendix A

Appendix to Chapter 1

A.1 Proof of Proposition 1.2

This appendix proves that firms optimally issue debt and equity securities and provide enough compensation to management that they do not divert resources. The firm sells securities before its cashflows f_i are privately observed by its management. Because there are only two types of investors, without loss of generality the firm only issues one non-equity security. These securities can have payoffs that depend on the residual cashflows x_i that remain after management has been compensated but not directly on the uncontractible aggregate good or bad state. I first take as given the securities the firm issues and study its optimal compensation of management and then solve for its optimal security issuance.

Suppose the firm has issued a security paying $s(x_i)$, depending only on the residual cashflows x_i, leaving the residual claim $x_i - s(x_i)$ for the firm's equityholders. Equityholders provide compensation to management in order to maximize the value of their residual claim. In general, such a compensation contract can be represented by a mechanism with message space M, so the payment to management is a function $R : M \times X \to \mathbb{R}^+$ where $x \in X$ is the cashflows remaining after management diverts resources. When f_i is realized and management chooses $(m_i, x_i) \in M \times X$, management's payoff is

$$R(m_i, x_i) + C(f_i - x_i) \tag{A.1}$$

. which is maximized by management's strategy $[m_i(f_i), x_i(f_i)]$.

If $x_i(f_i) = x_i(f_i')$, then

$$R(m_i(f_i), x_i(f_i)) + C(f_i - x_i(f_i)) \geq R(m_i'(f_i), x_i(f_i)) + C(f_i - x_i(f_i)) \tag{A.2}$$

$$R(m_i'(f_i), x_i(f_i)) + C(f_i' - x_i(f_i)) \geq R(m_i(f_i), x_i(f_i)) + C(f_i' - x_i(f_i)) \tag{A.3}$$

so $R(m_i(f_i), x_i(f_i)) = R(m_i'(f_i), x_i(f_i))$. It follows that the message space M can be ignored and all allocations depend only on x_i, with management receiving compensation $R(x_i)$.

After x_i is revealed to equityholders, they are able to covertly destroy resources or raise funds and pay them back at the market rate. If x_i is revealed and equityholders destroy resources, they can reduce x_i and receive the payoff $x - s(x)$ for $x \leq x_i$. If equityholders raise hidden funding, they can increase x_i to any $x > x_i$ but must pay back $(x - x_i)$ to the outside source of funding, receiving $(x - s(x)) - (x - x_i)$. Equityholders therefore choose x to maximize

$$G_{x_i}(x) = \left(\{x - s(x)\}_{x \leq x_i} \{x - s(x) - (x - x_i)\}_{x > x_i}\right). \tag{A.4}$$

This menu is pointwise increasing in x_i, so equityholders find it optimal to induce management to turn over the largest feasible x_i given f_i. Because $C' < 1$, this occurs when management receives the smallest payment to induce no diversion, which pays $C(f_i)$ when f_i is realized[1]. Equity then maximizes

$$\left(\{x - s(x)\}_{x \leq f_i - C(f_i)} \{x - s(x) - (x - x_i)\}_{x_i > f_i - C(f_i)}\right) \tag{A.5}$$

.

[1]That is, $R(x_i) + x_i = f_i$ and $R(x_i) = C(f_i)$. This system of equations has a unique solution.

The optimal $x(f_i)$ implies the payment to equity is increasing because $G_{f_i - C(f_i)}(x)$ is pointwise monotone increasing in x, which is preserved under taking a supremum.

Note that

$$\left| \sup_x G_{f_i - C(f_i)}(x) - \sup_x G_{f_i' - C(f_i')}(x) \right| \leq \left| f_i - C(f_i) - f_i' - C(f_i') \right| \tag{A.6}$$

so $f_i - C(f_i) - e(x(f_i - C(f_i)))$is increasing as well.

Note also that if $s(x)$ and $x - s(x)$ are increasing, it is optimal for equity to neither destroy resources nor raise hidden funding.

It follows that the realized payoffs of securities satisfy $s(x) + e(x) \leq x$, and both $e(x)$ and $x - e(x)$ are nonnegative monotone increasing.

As a result, equityholders can increase the market value of $s(.)$ without reducing the payoff to equity by replacing $s(x)$ by $x - e(x)$.

The optimal security issuance therefore satisfies $s(x) + e(x) = x$, with s and e increasing. Since s and e are therefore Lipschitz and thus absolutely continuous and $s(0) = e(0) = 0$, there exist functions e' and s' such that

$$s(f_i - C(f_i)) = \int_0^{f_i - C(f_i)} s'(u')\, du = \int_0^{\infty} s'(u) \{f_i - C(f_i) > u\}\, du \tag{A.7}$$

$$e(f_i - C(f_i)) = \int_0^{\infty} e'(u) \{f_i - C(f_i) > u\}\, du \tag{A.8}$$

where s' and e' are nonnegative and sum to 1.

Each security can therefore be written as a portfolio of assets of the form $\{f_i - C(f_i) > u\}$. Since $\frac{\Pr\{f_i - C(f_i) > u | good\}}{\Pr\{f_i - C(f_i) > u | bad\}}$ is strictly increasing in u, there exists a cut-off u^* such that $\{f_i - C(f_i) > u\}$ is more valuable to the intermediary for $u < u^*$ and to the household for $u > u^*$ since the intermediary is willing to pay more than the

household only for assets with low enough systematic risk. The optimal security design therefore sells the claim $\int_0^{u^*} \{f_i - C(f_i) > u\}\, du = \min(f_i - C(f_i), u^*)$ to the intermediary and $\int_{u^*}^{\infty} \{f_i - C(f_i) > u\}\, du = \max(f_i - C(f_i) - u^*, 0)$ to the household. These are the payoffs of a debt security and an equity security, and because they are both monotone increasing, equityholders will not destroy cashflows or raise hidden funding.

Appendix B

First Appendix to Chapter 3

B.1 Proofs: Identification

B.1.1 Proof of Lemma 3.1

We present the argument for the identification of the level sets of $h(\cdot)$ since the proof for $g(\cdot)$ is identical. The cdf of v conditional on $h(x)$ is given by $F_{V|h(x)}(v) = F_\varepsilon(v - h(x))$. Note that $F_{V|h(x)}(v)$ is increasing in v and decreasing in $h(x)$. Let $F_{q|h(x)}(q|h(x)) = F_{V|h(x)}\left(F_V^{-1}(q) - h(x)\right)$ be the cdf of the quantile of v given $h(x)$. Since F_V^{-1} is an increasing function of q, $F_{q|h(x)}(q|h(x))$ is increasing in q and decreasing in $h(x)$. As noted in Remark 3.1, the q-th quantile of each side matches with the q-th quantile of the other. Therefore, the density of $g(Z)$ that $h(x)$ is matched with is given by

$$\begin{aligned} f_{g(Z)|h(x)}(g|h) &= \int_0^1 f_{g(Z)|q}(g|q) f_{q|h(x)}(q|h)\, dq \\ &= \int_0^1 f_{q|g}(q|g) f_g(g) f_{q|h(x)}(q|h)\, dq \\ &= \int f_\eta(u-g) f_g(g) f_{q|h(x)}(F_U(u)|h)\, du, \end{aligned}$$

where $f_g(\cdot)$ is the density of $g(Z)$. The second equality uses Bayes' rule. The last equality follows from a change of variables $q = F_U(u)$ and the fact that $f_{q|g}(F_U(u)|g) = \frac{f_\eta(u-g) f_g(g)}{f_U(u)}$. Since $f_g(g) > 0$ for all g, and f_η has a non-vanishing characteristic function, $f_{g(Z)|h(x)}(\cdot|h)$

is injective in h. Since $F_{q|h(x)}(q|h)$ is decreasing in h, if $h(x') > h(x)$, then $f_{q|h(x)}(q|h(x')) \neq f_{q|h(x)}(q|h(x))$ for some q. Hence, we have that $f_{g(Z)|h(x)}(g|h(x')) \neq f_{g(Z)|h(x)}(g|h(x))$ if $h(x') \neq h(x)$. If $Z|x \sim Z|x'$ then $g(Z)|x \sim g(Z)|x'$. Therefore, it must be that the distribution of Z given x differs from the distribution of Z given x'. Therefore, the level sets of $h(\cdot)$ are identified.

B.1.2 Proof of Theorem 3.1

In what follows we treat x and z as single dimensional variable that are uniformly distributed on $[0,1]$, and $h(\cdot)$ and $g(\cdot)$ are increasing. This simplification is without loss of generality given identification of $g(x)$ and $h(z)$ up to a positive monotone transformation by Proposition 3.1.

The proof follows from recasting the matching model in terms of the non-classical measurement error model similar to Hu and Schennach (2008), (henceforth HS) to identify $f_{x|q}(x|q)$ and $f_{z|q}(z|q)$, which are the conditional densities of x and z respectively given $h(x) + \varepsilon = F_U^{-1}(q)$ and $g(z) + \eta = F_V^{-1}(q)$, where q is the quantile of the latent index.[1] Lemma C.1 implies that the primitives $h(\cdot)$, $g(\cdot)$, f_η and f_ε, are identified from $f(x|q)$ and $f(z|q)$.

We begin by verifying Assumptions HS.2-HS.4. Assumption HS.2 requires $f_{z|x_1,x_2,q}(z|x_1,x_2,q) = f_{z|q}(z|q)$, and $f_{x_1|z,x_2,q}(x_1|z,x_2,q) = f_{x_1|q}(x_1|q)$. This is satisfied since the quantile of the latent index q is a sufficient statistic for the distribution of observable characteristics in any match.

Assumption HS.3 requires that $L_{x|q}$ and $L_{x_1|x_2}$ are injective, where $L_{x|q}(m) = \int_0^1 f_{x|q}(x|q)\, m(q)\, dq$ and $L_{x_1|x_2}(m) = \int_0^1 f_{x_1|x_2}(x_1|x_2)\, m(x_2)\, dx_2$. Lemmas C.3 and C.4 imply that under Assumption 3.1, $L_{x|q}$ and $L_{x_1|x_2}$ are injective.

Assumption HS.4 requires that for all q_1 and q_2 in $[0,1]$, the set $\left\{z : f_{z|q}(z|q_1) \neq f_{z|q}(z|q_2)\right\}$ has positive probability (under the marginal distribution of z)

[1]The latent variable x^* in HS will be labelled q, the outcome y in HS is instead z, x in HS is x_1 and z in HS is x_2.

if $q_1 \neq q_2$. This assumption is satisfied since

$$\begin{aligned} f_{z|q}(z|q) &= \frac{f_{q|z}(q|z) f_z(z)}{f_q(q)} \\ &= f_{q|z}(q|z) \\ &= \frac{1}{f_U\left(F_U^{-1}(q)\right)} f_\eta\left(F_U^{-1}(q) - g(z)\right) \end{aligned}$$

is complete (Lemma C.2). The first equality follows from Bayes' rule, the second equality uses the fact that z and q are uniformly distributed, and the third equality transforms $u = F_U^{-1}(q)$, using the fact that $f_{u|z}(u|z) = f_\eta(u - g(z))$.

For a function $m(\cdot)$, and any z and q, define the operator $\Delta_{z;q} m(q) = f_{z|q}(z|q) m(q)$ as in HS. Since $f(z, x_1|x_2)$ is observed, for any real valued function m and z, we can compute

$$L_{z;x_1|x_2}(m) = \int_0^1 f(z, x_1|x_2) m(x_2) dx_2 = L_{x_1|q} \circ \Delta_{z;q} \circ L_{q|x_2}(m) \tag{B.1}$$

as shown in HS. They then use Assumption HS.1 to show that (i) $L_{x_1|x_2}^{-1}$ exists and is densely defined, and (ii) $T = L_{z,x_1|x_2} L_{x_1|x_2}^{-1}$ has a unique spectral decomposition. Lemmas C.4 and C.5 respectively show that these results follow under our assumptions (the conditions needed for Lemma C.5 are verified in Lemmas C.4 and C.3). Hence, the conditional densities $f_{z|q}(z|q)$ and $f_{x|q}(x|q)$ are identified up to a reindexing via a bijection $Q(\cdot)$ where $\tilde{q} = Q(q)$. That is, for every pair $\tilde{f}_{x|q}$ and $\tilde{f}_{z|q}$ satisfying our regularity conditions that can rationalize $f(z, x_1|x_2)$, the proof of Theorem 1 in HS shows that there exist bijections $Q_x : [0,1] \to [0,1]$ and $Q_z : [0,1] \to [0,1]$ such that $f_{x|Q_x(q)} = \tilde{f}_{x|q}$ and $f_{z|Q_z(q)} = \tilde{f}_{z|q}$.

This remaining under-identification issue is referred to as the ordering/indexing ambiguity issue in HS. They solve this ambiguity by using Assumption HS.5, which assumes that there is a known functional M such that $M\left[f_{x|q}(\cdot|q)\right] = q$ for all q. Since our model does not deliver such a functional, we instead solve the ordering/indexing ambiguity by using the fact that in our model, the q indexes the quantiles of the latent index and $f_{x|q}$ and f_q must therefore satisfy certain known properties. Specifically, we use Lemma C.6 to show directly, that Q_x and Q_z must be the identity function under the assumptions of our model. To apply Lemma C.6, we need to show that $f_{x|q}(x|q) = \frac{f_\varepsilon\left(F_V^{-1}(q) - h(x)\right)}{f_V\left(F_V^{-1}(q)\right)}$ and $f_{z|q}(z|q) = \frac{f_\eta\left(F_U^{-1}(q) - g(z)\right)}{f_U\left(F_U^{-1}(q)\right)}$

where q are quantiles satisfies Condition C.1. Since the proof is symmetric, we show this only for $f_{x|q}(x|q)$. Condition C.1(i) is satisfied since $f_{x|q}(x|q)$ is complete (Lemma C.2).

To verify condition C.1(ii), we compute $\frac{\partial f_{x|q}(x|q)}{\partial q}$. Note that $f_{x|q}(x|q) = f_{q|x}(q|x)f_x(x)/f_q(q) = f_{q|x}(q|x)$ by Bayes' rule and the (normalized) marginal distributions of x and q. Therefore,

$$\begin{aligned}\frac{\partial f_{x|q}(x|q)}{\partial q} = \frac{\partial f_{q|x}(q|x)}{\partial q} &= \frac{\partial}{\partial q}\frac{f_\varepsilon\left(F_V^{-1}(q) - h(x)\right)}{f_V\left(F_V^{-1}(q)\right)} \\ &= \frac{f_V(F_V^{-1}(q))f_\varepsilon'(F_V^{-1}(q) - h(x)) - f_\varepsilon(F_V^{-1}(q) - h(x))f_V'(F_V^{-1}(q))}{f_V(F_V^{-1}(q))^3}.\end{aligned}$$

Therefore, Condition C.1(ii) follows from Assumption 3.1 since each of the terms is finite and $f_V(v) > 0$ since ε and $h(X)$ have full support on $\mathbb{R}$.

We can verify Condition C.1(iii) by showing that for each $q \in (0,1)$, there exists x such that $\frac{d}{dq}f_{q|x}(q|x) \neq 0$. Towards a contradiction, for a given $q \in (0,1)$, assume that $\frac{d}{dq}f_{q|x}(q|x) = 0$ for all x. As shown above, $\frac{d}{dq}f_{q|x}(q|x) = \frac{d}{dq}\frac{f_\varepsilon(F_V^{-1}(q) - h(x))}{f_V(F_V^{-1}(q))}$. Since $f_V(v) > 0$, $\frac{d}{dq}f_{q|x}(q|x) = 0$ for all x if and only if $\frac{d}{dv}\frac{f_\varepsilon(v - h(x))}{f_V(v)}$ evaluated at $v = F_V^{-1}(q)$ if zero for all x, It must therefore be that

$$\frac{d}{dv}\frac{f_\varepsilon(v - h(x))}{f_V(v)} = \frac{f_V(v)f_\varepsilon'(v - h(x)) - f_\varepsilon(v - h(x))f_V'(v)}{f_V(v)^2}$$

is zero for all x for each $v \in (-\infty, \infty)$. Since $f_V(v) > 0$, it must be that $f_V(v)f_\varepsilon'(v - h(x)) = f_\varepsilon(v - h(x))f_V'(v)$ for all x. Since $h(x)$ has full support on $\mathbb{R}$, this implies that $f_\varepsilon'(\varepsilon) = K_1 f_\varepsilon(\varepsilon)$ for all $\varepsilon \in (\infty, \infty)$. Hence, $f_\varepsilon(\varepsilon) = K_2 \exp(K_1\varepsilon)$ for a constants K_1 and K_2. Note that f_ε is a density with full support, which is a contradiction with this functional form.

Condition C.1(iv) is definitional for the particular model considered since q indexes quantiles. Condition C.1(v) follows from Lemma C.3 under Assumption 3.1. Conditions C.1(vi) is also definitional in our case since $f_{x|q}$ are conditional densities and q indexes quantiles. We have thus verified Condition C.1 for $f_{x|q}$. An identical argument follows for $f_{z|q}$. Therefore, by Lemma C.6, Q_x and Q_z are the identity function. Hence, we have identified $f_{x|q}$ and $f_{z|q}$.

B.2 Proofs: Estimation

B.2.1 Proof of Proposition 3.3

We first rewrite

$$\psi_N - \psi = (\psi_N - E(\psi_N|\mu_{V_N}, \mu_{U_N})) + (E(\psi_N|\mu_{V_N}, \mu_{U_N}) - \psi).$$

Proof of Part (i): Lemma C.8(i) shows that if Assumption 3.4(i) is satisfied, $E(\psi_N|\mu_{V_N}, \mu_{U_N}) - \psi$ converges in probability to 0 as $N \to \infty$. This result is proved by rewriting

$$\begin{aligned} E(\psi_N|\mu_{V_N}, \mu_{U_N}) &= \frac{1}{N/2}\sum_{k=1}^{N/2} \tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right), F_{V_N}^{-1}\left(\frac{2k}{N}\right), F_{U_N}^{-1}\left(\frac{k}{N/2}\right)\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} \tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{U_N}^{-1}\left(\frac{i}{N}\right)\right) + R, \end{aligned} \tag{B.2}$$

where F_{V_N} and F_{U_N} are the cdfs representing the empirical measures μ_{V_N} and μ_{U_N} respectively, and R is a remainder term. We then show that R and

$$\frac{1}{N}\sum_{i=1}^{N} \tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{U_N}^{-1}\left(\frac{i}{N}\right)\right) - \psi$$

converge in probability to zero. Lemma C.9(i) shows that $\psi_N - E(\psi_N|\mu_{V_N}, \mu_{U_N})$ converges in probability to zero by bounding its variance by $\frac{1}{J}4\|\Psi\|_\infty^2$. Since $\psi_N - \psi = (\psi_N - E(\psi_N|\mu_{V_N}, \mu_{U_N})) + (E(\psi_N|\mu_{V_N}, \mu_{U_N}) - \psi)$ is the sum of two terms that converge in probability to 0, the result follows directly from Slutsky's theorem.

Proof of Part (ii): Lemma C.8(ii) shows if Assumption 3.4(ii) is satisfied, then for any bounded μ_X-Donsker class Γ_X and for any bounded μ_Z-Donsker class Γ_Z,

$$\left[\sqrt{N}(E(\psi_N|\mu_{V_N}, \mu_{U_N}) - \psi), \sqrt{N}(\mu_{X_N} - \mu_X)(\gamma_X), \sqrt{N/2}(\mu_{Z_N} - \mu_Z)(\gamma_Z)\right]$$

indexed by $\gamma_X \in \Gamma_X$ and $\gamma_Z \in \Gamma_Z$ is asymptotically equivalent to

$$\left[\begin{array}{c} \sqrt{N}\int_0^1 \nabla\tilde{\psi}_q\left(q,q,q\right)\cdot\left[\begin{array}{c} \left(\mu_{(X,\varepsilon)_N}-\mu_{X,\varepsilon}\right)\left(1\left\{h\left(x;\theta_0\right)+\varepsilon\leq F_V^{-1}\left(q_X\right)\right\}\right) \\ \left(\mu_{(X,\varepsilon)_N}-\mu_{X,\varepsilon}\right)\left(1\left\{h\left(x;\theta_0\right)+\varepsilon\leq F_V^{-1}\left(q_X\right)\right\}\right) \\ \left(\mu_{(Z,\eta)_N}-\mu_{Z,\eta}\right)\left(1\left\{g\left(z;\theta_0\right)+\eta\leq F_U^{-1}\left(q_Z\right)\right\}\right) \end{array}\right] dq \\ \sqrt{N}\left(\mu_{X_N}-\mu_X\right)\left(\gamma_X\right) \\ \sqrt{N/2}\left(\mu_{Z_N}-\mu_Z\right)\left(\gamma_Z\right) \end{array}\right], \quad \text{(B.3)}$$

which converges weakly to a mean-zero Gaussian process with a covariance kernel V'. This covariance kernel is derived by using equation (B.2) to show that $\sqrt{N}R$ converges in probability to zero, and then analyzing

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{U_N}^{-1}\left(\frac{i}{N}\right)\right)-\psi\right)$$

using Taylor approximations. Since $\left\|\nabla\tilde{\psi}_q\right\|_\infty < \infty$, the expression in (B.3) is a sum of $\mu_{X,\varepsilon}$- and $\mu_{Z,\eta}$- Donsker classes because we have added a finite number of sums of i.i.d. random variables to Γ_X and Γ_Z. Let $\Gamma_{X,\varepsilon}$ and $\Gamma_{Z,\eta}$ be the index sets for this empirical process. Lemma C.10 shows that if Assumption 3.4(ii) is satisfied, then for any bounded $\mu_{X,\varepsilon}$-Donsker class $\Gamma_{X,\varepsilon}$ and for any bounded $\mu_{Z,\eta}$-Donsker class $\Gamma_{Z,\eta}$,

$$\left[\sqrt{N}\left(\psi_N - E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right),\sqrt{N}\left(\mu_{(X,\varepsilon)_N}-\mu_{X,\varepsilon}\right)\left(\gamma_{X,\varepsilon}\right),\sqrt{N/2}\left(\mu_{(Z,\eta)_N}-\mu_{Z,\eta}\right)\left(\gamma_{Z,\eta}\right)\right]$$

indexed by $\gamma_{X,\varepsilon} \in \Gamma_{X,\varepsilon}$ and $\gamma_{Z,\eta} \in \Gamma_{Z,\eta}$ converges weakly to a mean-zero Gaussian process with a covariance kernel V''. To prove this result, we first compute the joint moment generating function for particular elements, $\gamma_{X,\varepsilon}$ and $\gamma_{Z,\eta}$, to show that it approaches the moment generating function of a mean-zero normal random variable, and derive the covariance $V''\left(\gamma_{X,\varepsilon},\gamma_{Z,\varepsilon}\right)$. We then verify equicontinuity of the process to show weak convergence.

Therefore, applying this result to the process indexed by $\Gamma_{X,\varepsilon}$ and $\Gamma_{Z,\eta}$, we have that the

process

$$\begin{bmatrix} \sqrt{N}\left(E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)-\psi\right)+\sqrt{N}\left(\psi_N-E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right) \\ \sqrt{N}\left(\mu_{X_N}-\mu_X\right)(\gamma_X) \\ \sqrt{N/2}\left(\mu_{Z_N}-\mu_Z\right)(\gamma_Z) \end{bmatrix}$$

indexed by $\gamma_X \in \Gamma_X$ and $\gamma_Z \in \Gamma_Z$ converges weakly to a mean zero Gaussian process with covariance kernel V.

We now compute V. Note that $V(\gamma_\Psi,\gamma_Z) = V'(\gamma_\Psi,\gamma_Z)+\sqrt{2}V''(\gamma_\Psi,\gamma_Z)$ and $V(\gamma_\Psi,\gamma_X) = V'(\gamma_\Psi,\gamma_X)+2V''(\gamma_\Psi,\gamma_X)$ since covariance is bilinear. $V(\gamma_\Psi,\gamma_\Psi) = V'(\gamma_\Psi,\gamma_\Psi)+2V''(\gamma_\Psi,\gamma_\Psi)$ since $Cov\left(X-E\left[X|\mathcal{I}\right],E\left[X|\mathcal{I}\right]-E\left[X\right]\right)=0$ for any sigma-field $\mathcal{I}$ by the law of iterated expectations. Finally, by definition, $V(\gamma_X,\gamma_Z)=0$, $V(\gamma_X,\gamma_X')=V'(\gamma_X,\gamma_X')$ and $V(\gamma_Z,\gamma_Z')=V'(\gamma_Z,\gamma_Z')$. The remaining elements are $V(\gamma_\Psi,\gamma_X)=V'(\gamma_\Psi,\gamma_X)+2V''(\gamma_\Psi,\gamma_X)$, $V(\gamma_\Psi,\gamma_\Psi)=V'(\gamma_\Psi,\gamma_\Psi)+2V''(\gamma_\Psi,\gamma_\Psi)$ and $V(\gamma_\Psi,\gamma_Z)=V'(\gamma_\Psi,\gamma_Z)+\sqrt{2}V''(\gamma_\Psi,\gamma_Z)$, where V' and V'' are as defined in Lemmata C.8 and C.9 respectively.

B.2.2 Proof Sketch for Proposition 3.4

Consider a sequence of measures (μ_{X_N},μ_{Z_N}) and scalars $h_N \to 0$ such that $\frac{1}{h_N}(\mu_{X_N}-\mu_X,\mu_{Z_N}-\mu_Z)$ converges to $G=(G_X,G_Z)$ uniformly in L_∞^Γ, where G is bounded and uniformly continuous. The Hadamard derivative is the limit of

$$\frac{1}{h_N}\left[\tilde{\psi}^\delta\left[\mu_X,\mu_Z\right](\theta)-\tilde{\psi}^\delta\left[\mu_{X_N},\mu_{Z_N}\right](\theta)\right] \tag{B.4}$$

$$= \frac{1}{h_N}\left[\int_\delta^{1-\delta}\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq \tag{B.5}$$

$$-\int_\delta^{1-\delta}\frac{\int\Psi(x_1,x_2,z)\,\phi_{\varepsilon,N}(q,x_1;\theta)\,\phi_{\varepsilon,N}(q,x_2;\theta)\,\phi_{\eta,N}(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}{\int\phi_{\varepsilon,N}(q,x_1;\theta)\,\phi_{\varepsilon,N}(q,x_2;\theta)\,\phi_{\eta,N}(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}dq\Bigg]$$

where

$$\phi_\eta(q,z;\theta)=f_\eta\left(F^{-1}_{u;\theta,\mu_Z}(q)-g(z;\theta)\right) \qquad \phi_{\eta,N}(q,z;\theta)=f_\eta\left(F^{-1}_{N,U;\theta,\mu_{Z_N}}(q)-g(z;\theta)\right)$$
$$\phi_\varepsilon(q,x;\theta)=f_\varepsilon\left(F^{-1}_{v;\theta,\mu_X}(q)-h(x;\theta)\right) \qquad \phi_{\varepsilon,N}(q,x;\theta)=f_\varepsilon\left(F^{-1}_{N,V;\theta,\mu_{X_N}}(q)-h(x;\theta)\right).$$

in terms of G_X and G_Z. The detailed calculations are presented in Appendix C.2.2.

Here, we illustrate the basic ideas of the argument and the components of the derivative by computing the limit of the following simplified expression:

$$\frac{1}{h_N}\left[\int_\delta^{1-\delta}\frac{\int\Psi(x)\,\phi_\varepsilon(q,x;\theta)\,d\mu_X}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}dq-\int_\delta^{1-\delta}\frac{\int\Psi(x)\,\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}{\int\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}dq\right].$$

We first rewrite the difference

$$\begin{aligned}
&\int_\delta^{1-\delta}\frac{\int\Psi(x)\,\phi_\varepsilon(q,x;\theta)\,d\mu_X}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}dq-\int_\delta^{1-\delta}\frac{\int\Psi(x)\,\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}{\int\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}dq\\
=&\int_\delta^{1-\delta}\frac{\int\Psi(x)\,\phi_\varepsilon(q,x;\theta)\,(d\mu_X-d\mu_{X_N})}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}dq+\int_\delta^{1-\delta}\frac{\int\Psi(x)\,\phi_\varepsilon(q,x;\theta)\,d\mu_{X_N}}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}dq\\
&-\int_\delta^{1-\delta}\frac{\int\Psi(x)\,\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}\frac{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}{\int\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}dq\\
=&\int_\delta^{1-\delta}\frac{\int\Psi(x)\,\phi_\varepsilon(q,x;\theta)\,(d\mu_X-d\mu_{X_N})}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}dq+\\
&\int_\delta^{1-\delta}\frac{\int\Psi(x)\,(\phi_\varepsilon(q,x;\theta)-\phi_{\varepsilon,N}(q,x;\theta))\,d\mu_{X_N}}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}dq\\
&+\int_\delta^{1-\delta}\frac{\int\Psi(x)\,\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}\times\left(1-\frac{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}{\int\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}\right)dq\\
=&\int_\delta^{1-\delta}T_1(q)+T_2(q)+T_3(q)\,dq
\end{aligned}$$

To obtain the limit of $\frac{1}{h_N}\int_\delta^{1-\delta}T_1(q)\,dq$, note that $\frac{1}{h_N}(\mu_{X_N}-\mu_X)$ converges uniformly to $G_X\in L_\infty^{\Gamma_X}$. Therefore,

$$\begin{aligned}
&\frac{1}{h_N}\int_\delta^{1-\delta}T_1(q)\,dq=\frac{1}{h_N}\int_\delta^{1-\delta}\frac{\int\Psi(x)\,\phi_\varepsilon(q,x;\theta)\,(d\mu_X-d\mu_{X_N})}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}dq\\
&\to\int_\delta^{1-\delta}\frac{\int\Psi(x)\,\phi_\varepsilon(q,x;\theta)\,dG_X}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}dq.
\end{aligned}$$

To obtain the limit of $\frac{1}{h_N}T_2(q)$, note that

$$\begin{aligned}
\frac{1}{h_N}T_2(q)&=\frac{\int\Psi(x)\,\frac{1}{h_N}(\phi_\varepsilon(q,x;\theta)-\phi_{\varepsilon,N}(q,x;\theta))\,d\mu_{X_N}}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}\\
&=\frac{1}{h_N}\left(F_{V;\theta}^{-1}(q)-F_{N,V;\theta}^{-1}(q)\right)\frac{\int\Psi(x)\,f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x;\theta)\right)d\mu_{X_N}}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}+o(1)\\
&=\frac{1}{h_N}\left(F_{V;\theta}^{-1}(q)-F_{N,V;\theta}^{-1}(q)\right)\frac{\int\Psi(x)\,f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x;\theta)\right)d\mu_X}{\int\phi_\varepsilon(q,x;\theta)\,d\mu_X}+o(1)
\end{aligned}$$

where the second equality follows from a Taylor expansion and the dominated convergence theorem (since f'_ε is bounded), and the last equality follows from the fact that $d\mu_{X_N} - d\mu_X \to 0$ and uniform bounds over $q \in (\delta, 1-\delta)$ on the remaining terms. We then show that

$$\frac{1}{h_N}\left(F^{-1}_{V;\theta}(q) - F^{-1}_{N,V;\theta}(q)\right) \to$$
$$\frac{1}{f_{V;\theta}\left(F^{-1}_{V;\theta}(q)\right)} \int G_X\left(1\left\{h(x;\theta) + \varepsilon \leq F^{-1}_{V;\theta}(q)\right\}\right) dF_\varepsilon = G^q_V(\theta)$$

uniformly in $q \in (\delta, 1-\delta)$ to obtain the limit

$$\frac{1}{h_N}\int_\delta^{1-\delta} T_2(q)\, dq \to \int_\delta^{1-\delta} G^q_V(\theta) \frac{\int \Psi(x) f'_\varepsilon\left(F^{-1}_{V;\theta}(q) - h(x;\theta)\right) d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\, d\mu_X} dq.$$

Finally, we rewrite

$$\begin{aligned}
T_3(q) &= \frac{\int \Psi(x)\, \phi_{\varepsilon,N}(q,x;\theta)\, d\mu_{X_N}}{\int \phi_\varepsilon(q,x;\theta)\, d\mu_X} \times \left(1 - \frac{\int \phi_\varepsilon(q,x;\theta)\, d\mu_X}{\int \phi_{\varepsilon,N}(q,x;\theta)\, d\mu_{X_N}}\right) \\
&= \frac{\int \Psi(x)\, \phi_{\varepsilon,N}(q,x;\theta)\, d\mu_{X_N}}{\int \phi_{\varepsilon,N}(q,x;\theta)\, d\mu_X} \times \left(\frac{\int \phi_{\varepsilon,N}(q,x;\theta)\, d\mu_X - \int \phi_\varepsilon(q,x;\theta)\, d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\, d\mu_{X_N}}\right) \\
&= \frac{\int \Psi(x)\, \phi_{\varepsilon,N}(q,x;\theta)\, d\mu_{X_N}}{\int \phi_{\varepsilon,N}(q,x;\theta)\, d\mu_X} \times \left(-\tilde{T}_1(q) - \tilde{T}_2(q)\right) \\
&= \frac{\int \Psi(x)\, \phi_\varepsilon(q,x;\theta)\, d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\, d\mu_X} \times \left(-\tilde{T}_1(q) - \tilde{T}_2(q)\right) \\
&\quad + \left(\frac{\int \Psi(x)\, \phi_{\varepsilon,N}(q,x;\theta)\, d\mu_{X_N}}{\int \phi_{\varepsilon,N}(q,x;\theta)\, d\mu_X} - \frac{\int \Psi(x)\, \phi_\varepsilon(q,x;\theta)\, d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\, d\mu_X}\right) \times \left(-\tilde{T}_1(q) - \tilde{T}_2(q)\right)
\end{aligned}$$

where $\tilde{T}_1(q) = T_1(q)$ and $\tilde{T}_2(q) = T_2(q)$ evaluated at $\Psi(x) = 1$. Since $\frac{1}{h_N}\left(-\tilde{T}_1(q) - \tilde{T}_2(q)\right)$ is finite, the second term is negligible. Hence,

$$\frac{1}{h_N} T_3(q) \to -$$
$$\frac{\int \Psi(x)\, \phi_\varepsilon(q,x;\theta)\, d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\, d\mu_X}\left(\frac{\int \phi_\varepsilon(q,x;\theta)\, dG_X}{\int \phi_\varepsilon(q,x;\theta)\, d\mu_X} + G^q_V(\theta)\frac{\int f'_\varepsilon\left(F^{-1}_{V;\theta}(q) - h(x;\theta)\right) d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\, d\mu_X}\right).$$

The limit of $\frac{1}{h_N}\int_\delta^{1-\delta} T_1(q) + T_2(q) + T_3(q)\, dq$ given by the expressions above yields the Hadamard derivative of interest. Online Appendix C.2.2 uses a dominated convergence argument to ensure that $T_1(q) + T_2(q) + T_3(q)$ converges uniformly in q.

Appendix C

Second Appendix to Chapter 3: Proof Details

C.1 Proofs: Identification

C.1.1 Non Identification in Data from One-to-One Matches

In this section, we show that a model with unobservables on one side of the market can rationalize any data from a large one-to-one matching market under the following condition:

Axiom C.1 *The primitives h, g, $F_{X,\varepsilon}$, $F_{Z,\eta}$ are such that*

(i) $F_{h(X)|V}(\bar{h}, v) = \mathbb{P}\left(h(X) \leq \bar{h} | h(X) + \varepsilon = v\right) = \gamma(\kappa\bar{h} - v)$ for some function γ and constant κ, (ii) $F_V^{-1} \circ F_U$ is a linear function, (iii) The functions h, g and $f_{h(X)|V}$ are twice continuously differentiable, and (iv) ε and η are independent of X and Z respectively.

As is evident, these conditions are satisfied in Example 3.1. The joint distribution F_{XZ} produced by the model in the example is identical to one produced by the following transformation model, $X = \frac{1}{\kappa^{1/2}} Z + \eta_1$, where $\eta_1 \sim N\left(0, 1 - \frac{1}{\kappa}\right)$.

Proposition C.1 *Under Assumptions C.1 and 3.2, any joint distribution F_{XZ} can be rationalized in a matching model with $\varepsilon \equiv 0$.*

Proof. The proof proceeds by rewriting the matching model with $\varepsilon \equiv 0$ in terms of the transformation model of Chiappori and Komunjer (2008). We will then use Chiappori and Komunjer (2008) Proposition 2, which states that the transformation model is correctly specified.

In the matching model, quantiles of $h(X) + \varepsilon$ are matched with quantiles of $g(Z) + \eta$. We will use Proposition 2 of Chiappori and Komunjer (2008) to show that there exist increasing functions $\bar{\Gamma}$, $\bar{g}$, $F_{\bar{\eta}}$ such that the the transformation model

$$h(X) = \bar{\Gamma}\left(\bar{g}(Z) + \bar{\eta}\right)$$

rationalizes any joint distribution F_{XZ} from a matching model satisfying Assumptions 3.1 - 3.2 and Condition C.1. This is model is equivalent to a matching model with $\bar{h} = \bar{\Gamma}^{-1} \circ h$, $\varepsilon \equiv 0$, and $F_{\bar{\eta}}$, $\bar{g}$. In what follows, we will treat X and Z as known scalars with $h(\cdot)$ and $g(\cdot)$ as increasing functions of them respectively. This simplification is without loss of generality since our we show that a positive monotone transformation $h(\cdot)$ and $g(\cdot)$ exists that yields an identical joint distribution F_{XZ}.

Since X and Z are unidimensional, Assumption A3 of Chiappori and Komunjer (2008) is then equivalent to independence of ε and η from X and Z respectively, as maintained under the hypotheses of Proposition C.1.

Let the probability that a firm with observable trait z is matched with workers with $h(X)$ at most $\bar{h}$ be denoted $\Phi\left(\bar{h}, z\right) = F_{h(X)|Z}\left(\bar{h}, z\right)$. Note that

$$\begin{aligned}
\Phi\left(\bar{h}, z\right) &= \int F_{h(X)|V}\left(\bar{h}, F_V^{-1} F_U\left(g(z) + \eta\right)\right) dF_\eta \\
&= \int \gamma\left(\kappa\bar{h} - A\left(g(z) + \eta\right)\right) dF_\eta,
\end{aligned}$$

for some constant A. The first equality is derived from the quantile-quantile matching of workers and firms and the second equality follows from Conditions C.1 (i) and C.1 (ii).

First, we ensure that Φ has continuous third order partial derivatives $\partial^3\Phi(\bar{h}, z)/\partial\bar{h}\partial^2 z$ and $\partial^3\Phi(\bar{h}, z)/\partial^2\bar{h}\partial z$, and that $\partial\Phi(\bar{h}, z)/\partial\bar{h} > 0$. Conditions C.1 (iii) guarantees the existence of the required partial derivatives. Further, since $F_{h(X)|V}\left(\bar{h}, v\right)$ is strictly increasing in $\bar{h}$, we

have that $\partial\Phi(\bar{h},z)/\partial\bar{h} > 0$.

We now verify that $\Phi\left(\bar{h},z\right)$ satisfies Condition C in Chiappori and Komunjer (2008), i.e.

$$\frac{\partial^2}{\partial\bar{h}\partial z_k}\left(\log\left|\frac{\partial\Phi\left(\bar{h},z\right)/\partial\bar{h}}{\partial\Phi\left(\bar{h},z\right)/\partial z_1}\right|\right)=0.$$

The partial derivatives of $\Phi\left(\bar{h},z\right)$ with respect to $\bar{h}$ and z_1 are given by:

$$\begin{aligned}\frac{\partial\Phi\left(\bar{h},z\right)}{\partial\bar{h}} &= \kappa\int\gamma'\left(\kappa\bar{h}-A\left(g\left(z\right)+\eta\right)\right)dF_\eta\\ \frac{\partial\Phi\left(\bar{h},z\right)}{\partial z_1} &= -A\frac{\partial g\left(z\right)}{\partial z_1}\int\gamma'\left(\bar{h}-A\left(g\left(z\right)+\eta\right)\right)dF_\eta.\end{aligned}$$

Note that γ' exists since the existence of densities f_X, f_ε and differentiability of $h\left(\cdot\right)$ implies that the derivatives of $F_{h(X)|V}\left(\bar{h},v\right)$ exist.

Using the expressions above, rewrite

$$\frac{\partial^2}{\partial\bar{h}\partial z_k}\left(\log\left|\frac{\partial\Phi\left(\bar{h},z\right)/\partial\bar{h}}{\partial\Phi\left(\bar{h},z\right)/\partial z_1}\right|\right)=\frac{\partial^2}{\partial\bar{h}\partial z_k}\left(\log|\kappa|-\log\left|A\frac{\partial g\left(z\right)}{\partial z_1}\right|\right)=0.$$

The last equality follows since $\log\left|\frac{\partial g(z)}{\partial z_1}\right|$ and $\log|\kappa|$ do not depend on $\bar{h}$.

We now show that equations (4) and (5) in Chiappori and Komunjer (2008) are satisfied. Since $F_{h(X)|V}\left(\bar{h},A\left(g\left(z\right)+\eta\right)\right)$ is a cdf, it is bounded, $\lim_{\bar{h}\to-\infty}F_{X|V}\left(\bar{h},A\left(g\left(z\right)+\eta\right)\right)=0$ and $\lim_{\bar{h}\to\infty}F_{h(X)|V}\left(\bar{h},A\left(g\left(z\right)+\eta\right)\right)=1$ for each z and η. Hence, $\lim_{\bar{h}\to-\infty}\Phi\left(\bar{h},z\right)=0$ and $\lim_{\bar{h}\to\infty}\Phi\left(\bar{h},z\right)=1$.

To verify (5), note that

$$\begin{aligned}&\int_0^{\bar{h}}\frac{\partial\Phi\left(a,z\right)/\partial x}{\partial\Phi\left(a,z\right)/\partial z_1}\frac{\partial\Phi\left(0,z\right)/\partial z_1}{\partial\Phi\left(0,z\right)/\partial x}da\\ =&\int_0^{\bar{h}}\frac{\kappa}{-A\partial g\left(z\right)/\partial z_1}\frac{-A\partial g\left(z\right)/\partial z_1}{\kappa}da\\ =&\int_0^{\bar{h}}1da=\bar{h}.\end{aligned}$$

Equation (5) of Chiappori and Komunjer (2008) follows since $h\left(X\right)$ has full support on $\mathbb{R}$.

By Proposition 2 of Chiappori and Komunjer (2008), there exist $\bar{\Gamma}$, $\bar{g}$, $F_{\bar{\eta}}$ that rationalize Φ. ■

C.1.2 Preliminaries

Since $h(X)$ and $g(Z)$ admit bounded continuous densities and are identified up to positive monotone transformation, it is without loss to treat x and z as single dimensional variable that are uniformly distributed on $[0,1]$. Proposition 3.1 implies that this simplification is without loss of generality.

Let $v = h(x) + \varepsilon$, where $h(x)$ is strictly increasing with $h(\bar{x}) = 0$, $h'(\bar{x}) = 1$ and let ε be median zero with density f_ε. For quantile $\tau \in [0,1]$, let $f_{\tau|X}(\tau, x) = \frac{f_\varepsilon\left(F_V^{-1}(\tau) - h(x)\right)}{f_V\left(F_V^{-1}(\tau)\right)}$ be the density on $v = F_V^{-1}(\tau)$ given x, where $F_V(v) = \int F_\varepsilon(v - h(x))\, dF_X$.

Lemma C.1 *The function $h(x)$ and the density f_ε are identified from $f_{\tau|x}(\tau)$ if $h(x)$ is differentiable and ε has full support on $\mathbb{R}$.*

Proof. Let $\phi(x, x')$ be the probability that $h(x) + \varepsilon > h(x') + \varepsilon'$ given x and x'. $\phi(x, x')$ is identified from $f_{\tau|x}(\tau)$ since it can be written as

$$\phi(x, x') = \int_0^1 \int_{\tau > \tau'} f_{\tau|X}(\tau, x)\, f_{\tau|X}(\tau', x')\, d\tau d\tau'.$$

However, $\phi(x, x')$ can also be written in terms of the primitives $h(\cdot)$ and f_ε as

$$\phi(x, x') = \int F_\varepsilon\left(h(x) + \varepsilon - h(x')\right) f_\varepsilon(\varepsilon)\, d\varepsilon.$$

Taking the derivative with respect to x and x', we get

$$\begin{aligned} \frac{\partial \phi(x, x')}{\partial x} &= h'(x) \int f_\varepsilon\left(h(x) + \varepsilon - h(x')\right) f_\varepsilon(\varepsilon)\, d\varepsilon \\ \frac{\partial \phi(x, x')}{\partial x'} &= -h'(x') \int f_\varepsilon\left(h(x) + \varepsilon - h(x')\right) f_\varepsilon(\varepsilon)\, d\varepsilon. \end{aligned}$$

The ratio $\frac{\partial \phi(x,x')}{\partial x} / \frac{\partial \phi(x,x')}{\partial x'}$ is identified and is equal to $-\frac{h'(x)}{h'(x')}$.Since $h'(\bar{x}) = 1$, $h'(x)$ can be determined everywhere. The boundary condition $h(\bar{x}) = 0$ provides the unique solution to the resulting differential equation determining $h(\cdot)$.

We now need to show that F_ε is identified. Let $R_x(t) = \mathbb{P}\left(h(x) + \varepsilon \leq F_V^{-1}(t) \,|x\right)$. $R_x(t)$ is known since it is equal to $\int_0^t f_{\tau|x}(\tau)\, d\tau$. Since F_V^{-1} is continuous and ε admits a full support density, $R_x(t)$ is continuous and strictly increasing in t. Let τ^* be the median rank

of $\bar{x}$, i.e. $R_{\bar{x}}(t^*) = \frac{1}{2}$. Since ε is median-zero, $h(\bar{x}) = 0$ and $\mathbb{P}\left(h(\bar{x}) + \varepsilon \leq F_V^{-1}(t^*) \,|\bar{x}\right) = \frac{1}{2}$, we have that $F_V^{-1}(t^*) = 0$. For any x, $R_x(t^*) \in (0,1)$ is therefore the probability that $h(x) + \varepsilon \leq 0$ given x, i.e. $R_x(t^*) = F_\varepsilon(-h(x))$. Since $h(x)$ and $R_x(t^*)$ are known and have full support on $\mathbb{R}$, F_ε is identified. ■

Lemma C.2 *Suppose f_ε has a non-vanishing characteristic function and $h(X)$ has full support on $\mathbb{R}$. For any function $m(v)$, we have that $\int f_\varepsilon(v - h(x))\, m(v)\, dv = 0$ for all x implies that $m(v) = 0$ a.e. Further, if $h(\cdot)$ is differentiable and strictly increasing, then for any function $m(x)$, $\int f_\varepsilon(v - h(x))\, m(x)\, dx = 0$ for all v implies that $m(x) = 0$ a.e.*

Proof. Note that

$$\int f_\varepsilon(v - h(x))\, m(v)\, dv = \int f_\varepsilon(\varepsilon)\, m(h(x) + \varepsilon)\, d\varepsilon$$

is a convolution of $m(\cdot)$ with $-\varepsilon$. Since f_ε has a non-vanishing characteristic, so does $f_{-\varepsilon}$. Therefore, completeness follows from Mattner (1993), Theorem 2.1. Similarly, by a change of variables, $h(x) = h$, we have that

$$\int f_\varepsilon(v - h(x))\, m(x)\, dx = \int f_\varepsilon(v - h)\, M(h)\, dh,$$

where $M(h) = \frac{m\left(h^{-1}(h)\right)}{h'(h^{-1}(h))}$. Since f_ε has a non-vanishing characteristic, $\int f_\varepsilon(v - h)\, M(h)\, dh = 0$ implies that $M(h) = 0$ for all h. Since h is strictly increasing, this implies that $m(x) = 0$ for all x. ■

For a function m, define the operator $L_{x|q1} : L_1([0,1]) \to L_1([0,1])$ as $L_{x|q}(m) = \int f(x|q)\, m(q)\, dq$ where $f(x|q)$ is the conditional density of X given $Q = q$.

Lemma C.3 *$L_{x_1|q}$ is injective if (i) f_ε has a non-vanishing characteristic function (ii) F_V is continuous and strictly increasing and (iii) $h(X)$ has full support on $\mathbb{R}$. Further, $L_{x|q}$ is bounded, and $L_{x|q}^{-1}$ is exists and is densely defined.*

Proof. We first rewrite the operator $L_{x|q}$ as a convolution:

$$\begin{aligned}\int_0^1 f(x|q)\, m(q)\, dq &= \int f(x|v)\, m(F_V(v))\, f_V(v)\, dv \\ &= \int f(v|x)\, m(F_V(v))\, dv \\ &= \int f_\varepsilon(v - h(x))\, M(v)\, dv\end{aligned}$$

where the first equality follows from a change of variables, the second equality uses Bayes' rule $f_{x|v}(x|v) = \frac{f_{v|x}(v|x) f_x(x)}{f_V(v)}$ and the fact that the distribution of x is normalized to uniform $[0,1]$, and the third equality uses the fact that $f_{v|x}(v|x) = f_\varepsilon(v - h(x))$ and sets $M(v) = m(F_V(v))$. By Lemma C.2, $M(v) = 0$ for almost all v. Since F_V is bijective, we have that $m(q) = 0$ for almost all q. Therefore, $f(x|q)$ is complete, and as noted in Hu and Schennach (2008), hence that $L_{x|q}$ is injective.

Note that the (operator) norm of $L_{x|q}$ is at most

$$\begin{aligned}&\sup_{m \in \mathcal{L}_1([0,1])} \frac{1}{\|m\|_1} \int_0^1 \int \frac{f_\varepsilon(v - h(x))}{f_V(v)} |m(F_V(v))|\, f_V(v)\, dv dx \\ = &\sup_{m \in \mathcal{L}_1([0,1])} \frac{1}{\|m\|_1} \int \left(\int_0^1 \frac{f_\varepsilon(v - h(x))}{f_V(v)} dx \right) |m(F_V(v))|\, f_V(v)\, dv \\ = &\frac{1}{\|m\|_1} \int_0^1 |m(q)|\, dq = 1\end{aligned}$$

where we use a change in the order of integration by Fubini's theorem, the identity $f_V(v) = \int_0^1 f_\varepsilon(v - h(x))\, dx$, and a change in variables $q = F_V(v)$.

As argued in the proof of HS, Lemma 1, to show that $L_{x|q}^{-1}$ is densely defined, it is sufficient to show that the adjoint $L_{x|q}^\dagger$ is injective. Note that for any y in the space of bounded functions (dual of the domain of $L_{x|q}$), it must be that

$$\begin{aligned}\left\langle m, L_{x|q}^\dagger y \right\rangle = \left\langle L_{x|q} m, y \right\rangle &= \int_0^1 \int_0^1 f_{x|q}(x|q)\, m(q)\, dq y(x)\, dx \\ &= \int_0^1 \int_0^1 f_{x|q}(x|q)\, y(x)\, dx m(q)\, dq,\end{aligned}$$

where we changed the order of integration using Fubini's theorem. Therefore, $L_{x|q}^\dagger y(q) = \int_0^1 f_{x|q}(x|q)\, y(x)\, dx$. Since $f_{x|q}(x|q) = \frac{1}{f_V\left(F_V^{-1}(q)\right)} f_\varepsilon\left(F_V^{-1}(q) - h(x)\right)$ and $f_V(v) > 0$,

$L^{\dagger}_{x|q} y(q) = 0$ for all $q \in [0,1]$, implies that $\int_0^1 f_\varepsilon(v - h(x)) y(x) dx = 0$ for all $v \in \mathbb{R}$. By Lemma C.2, $y(x) = 0$ for almost all x. ■

Define the operator $L_{x_1|x_2} : L_1([0,1]) \to L_1([0,1])$ as $L_{x_1|x_2} m(x_1) = \int f_{x_1|x_2}(x_1|x_2) m(x_2) dx_2$ for any function $m \in L_1([0,1])$.

Lemma C.4 *$L_{x_1|x_2}$ is injective if (i) f_ε has a non-vanishing characteristic function (ii) F_V is continuous and strictly increasing and (iii) $h(x)$ has full support on $\mathbb{R}$. Further, $L_{x_1|x_2}$ is bounded, and $L^{-1}_{x_1|x_2}$ exists and is densely defined.*

Proof. Note that

$$\begin{aligned}
\int_0^1 f_{x_1|x_2}(x_1|x_2) m(x_2) dx_2 &= \int_0^1 \left(\int_0^1 f(x_1, q|x_2) dq \right) m(x_2) dx_2 \\
&= \int_0^1 \left(\int_0^1 f_{x_1|q}(x_1|q) f_{q|x_2}(q|x_2) dq \right) m(x_2) dx_2 \\
&= \int_0^1 \left(\int f_{x_1|v}(x_1|v) f_{v|x_2}(v|x_2) dv \right) m(x_2) dx_2 \\
&= \int \int_0^1 \frac{f_\varepsilon(v - h(x_1))}{f_V(v)} f_{v|x_2}(v|x_2) m(x_2) dx_2 dv \\
&= \int \frac{f_\varepsilon(v - h(x_1))}{f_V(v)} \left(\int_0^1 f_\varepsilon(v - h(x_2)) m(x_2) dx_2 \right) dv
\end{aligned}$$

where we use (i) $f(x_1, q|x_2) = f(x_1|q, x_2) f_{q|x_2}(q|x_2) = f_{x_1|q}(x_1|q) f_{q|x_2}(q|x_2)$, (ii) a change of variables $F_V(v) = q$, (iii) $f_{x_1|v}(x_1|v) = \frac{f_\varepsilon(v-h(x_1))}{f_V(v)}$, $f_{q|x_2}(F_V(v)|x_2) = \frac{1}{f_V(v)} f_{v|x_2}(v|x_2)$ and (iv) a change in the order of integration by Fubini's theorem.

By Lemma C.2, $f_\varepsilon(v - h(x_1))$ is complete, and consequently, $\int_0^1 f_{x_1|x_2}(x_1|x_2) m(x_2) dx_2 = 0$ implies $\frac{1}{f_V(v)} \int_0^1 f_\varepsilon(v - h(x_2)) m(x_2) dx_2 = 0$ for almost all v. Since $f_V(v) > 0$, we have that $\int_0^1 f_\varepsilon(v - h(x_2)) m(x_2) dx_2 = 0$ for almost all v. A second application of Lemma C.2 implies that $m(x) = 0$ for all x (Lemma C.2). Hence, $f_{x_1|x_2}(x_1|x_2)$ is complete, and as noted in Hu and Schennach (2008), completeness implies injectivity.

Note that the (operator) norm of $L_{x_1|x_2}$ is not greater than

$$\begin{aligned}
&\sup_{m\in\mathcal{L}_1([0,1])}\frac{1}{\|m\|_1}\int_0^1\int f_\varepsilon\left(v-h\left(x_1\right)\right)\frac{1}{f_V\left(v\right)}\left(\int_0^1 f_\varepsilon\left(v-h\left(x_2\right)\right)\left|m\left(x_2\right)\right|dx_2\right)dvdx_1\\
&=\sup_{m\in\mathcal{L}_1([0,1])}\frac{1}{\|m\|_1}\int_0^1\int\left(\frac{1}{f_V\left(v\right)}\int_0^1 f_\varepsilon\left(v-h\left(x_1\right)\right)dx_1\right)f_\varepsilon\left(v-h\left(x_2\right)\right)dv\left|m\left(x_2\right)\right|dx_2\\
&=\sup_{m\in\mathcal{L}_1([0,1])}\frac{1}{\|m\|_1}\int_0^1\int f_\varepsilon\left(v-h\left(x_2\right)\right)dv\left|m\left(x_2\right)\right|dx_2=1,
\end{aligned}$$

where we use (i) change in the order of integration by Fubini's theorem, and (ii) the identity $f_V\left(v\right)=\int_0^1 f_\varepsilon\left(v-h\left(x_1\right)\right)dx_1$.

As argued in the proof of HS, Lemma 1, to show that $L_{x_1|x_2}^{-1}$ is densely defined, it is sufficient to show that the adjoint $L_{x_1|x_2}^{\dagger}$ is injective. Note that for any y in the space of bounded functions (dual of the domain of $L_{x|q}$), it must be that

$$\begin{aligned}
&\left\langle m,L_{x_1|x_2}^{\dagger}y\right\rangle=\left\langle L_{x_1|x_2}m,y\right\rangle=\\
&\int_0^1\left(\int_0^1\left(\int_0^1 f_{x_1|q}\left(x_1|q\right)f_{q|x_2}\left(q|x_2\right)dq\right)m\left(x_2\right)dx_2\right)y\left(x_1\right)dx_1\\
&=\int_0^1\left(\int_0^1\left(\int_0^1 f_{x_1|q}\left(x_1|q\right)f_{q|x_2}\left(q|x_2\right)dq\right)y\left(x_1\right)dx_1\right)m\left(x_2\right)dx_2
\end{aligned}$$

where we changed the order of integration using Fubini's theorem. Hence,

$$L_{x_1|x_2}^{\dagger}y\left(x_2\right)=\int_0^1\left(\int_0^1 f_{x_1|q}\left(x_1|q\right)f_{q|x_2}\left(q|x_2\right)dq\right)y\left(x_1\right)dx_1.$$

The arguments made above for $L_{x_1|x_2}$ also imply that $L_{x_1|x_2}^{\dagger}$ is injective. ■

For a function m, define the operator $L_{z;x_1|x_2}$ as

$$L_{z;x_1|x_2}m\left(x_1\right)=\int_0^1\int_0^1 f_{z|q}\left(z|q\right)f_{x_1|q}\left(x_1|q\right)f_{q|x_2}\left(q|x_2\right)m\left(x_2\right)dqdx_2$$

where $f_{z|q}\left(z|q\right)$, $f_{x_1|q}\left(x_1|q\right)$, and $f_{q|x_2}\left(q|x_2\right)$ are the conditional densities of the respective random variables

Lemma C.5 *For any Borel set $\Lambda\subseteq\mathbb{R}$, let $P\left(\Lambda\right)=L_{x_1|q}I_\Lambda L_{x_1|q}^{-1}$ and $I_\Lambda m\left(q\right)=1\left(f_{z|q}\left(z|q\right)\in\Lambda\right)m\left(q\right)$. If (i) $L_{x_1|q}$ is a bounded invertible operator (ii) $L_{x_1|q}^{-1}$ is densely defined and (iii) $L_{x_1|x_2}^{-1}$ exists and is densely defined, then P is the unique projection-valued measure for which*

$T = L_{z;x_1|x_2} L_{x_1|x_2}^{-1} = \int \lambda P(d\lambda)$.

Proof. Hu and Schennach (2008) (henceforth HS) show that $L_{z;x_1|x_2} = L_{x_1|q} \Delta_{z;q} L_{x_1|q}^{-1} L_{x_1|x_2}$, and therefore $T = L_{z;x_1|x_2} L_{x_1|x_2}^{-1} = L_{x_1|q} \Delta_{z;q} L_{x_1|q}^{-1}$ where $\Delta_{z;q} m(q) = f_{z|q}(z|q) m(q)$. This allows them to re-write $T = \int \lambda P(d\lambda)$.

HS show that $T = L_{z;x_1|x_2} L_{x_1|x_2}^{-1}$ has a unique resolution of the identity by appealing to Theorem XV.4.5 in Dunford and Schwartz (1971) since, under Assumption HS.1 $f_{z|q}(z|q)$, is bounded, and therefore T is a bounded operator. However, $f_{z|q}(z|q)$ may not be bounded in our case. We therefore appeal to Corollary XVIII.14 in Dunford and Schwartz (1971), which shows an analogous result for unbounded scalar type operators. The result applies to our model if the projection-valued measure P is strongly countably additive, thereby satisfying Definition XVIII.10 in Dunford and Schwartz (1971)

To complete the proof, we need to show that if Λ_i is a countable sequence of disjoint Borel sets, then

$$P\left(\cup_{i=1}^{n} \Lambda_i\right) = L_{x_1|q} I_{\cup_{i=1}^{n} \Lambda_i} L_{x_1|q}^{-1} = L_{x_1|q} \left(\sum_{i=1}^{n} I_{\Lambda_i} \right) L_{x_1|q}^{-1}$$

converges to $P\left(\cup_{i=1}^{\infty} \Lambda_i\right) = L_{x_1|q} I_{\cup_{i=1}^{\infty} \Lambda_i} L_{x_1|q}^{-1}$ in the strong operator topology. Equivalently, we need to show that for any integrable function m, $\left\| L_{x_1|q} I_{\cup_{i=1}^{n} \Lambda_i} L_{x_1|q}^{-1} m - L_{x_1|q} I_{\cup_{i=1}^{\infty} \Lambda_i} L_{x_1|q}^{-1} m \right\| \to 0$. Since $L_{x_1|q}^{-1}$ is densely defined, it is enough to show this for any m in the domain of $L_{x_1|q}^{-1}$.

Let $\tilde{m} = L_{x_1|q}^{-1} m$, and note that

$$\left(I_{\cup_{i=1}^{n} \Lambda_i} - I_{\cup_{i=1}^{\infty} \Lambda_i} \right) \tilde{m}(q) = \sum_{i=n+1}^{\infty} 1\left(f_{z|q}(z|q) \in \Lambda_i \right) \tilde{m}(q)$$

converges pointwise to zero since $f_{z|q}(z|q)$ can live in only one set Λ_i. Since $\tilde{m}$ is in the domain of $L_{x|q}$, $\tilde{m}$ is integrable. By the Dominated Convergence Theorem, we have that $\left\| \left(I_{\cup_{i=1}^{n} \Lambda_i} - I_{\cup_{i=1}^{\infty} \Lambda_i} \right) \tilde{m} \right\| \to 0$. Now, note that

$$\begin{aligned} \left\| L_{x_1|q} I_{\cup_{i=1}^{n} \Lambda_i} L_{x_1|q}^{-1} m - L_{x_1|q} I_{\cup_{i=1}^{\infty} \Lambda_i} L_{x_1|q}^{-1} m \right\| &= \left\| L_{x_1|q} \left(I_{\cup_{i=1}^{n} \Lambda_i} - I_{\cup_{i=1}^{\infty} \Lambda_i} \right) L_{x_1|q}^{-1} m \right\| \\ &\leq \left\| L_{x_1|q} \right\| \left\| \left(I_{\cup_{i=1}^{n} \Lambda_i} - I_{\cup_{i=1}^{\infty} \Lambda_i} \right) L_{x_1|q}^{-1} m \right\| \end{aligned}$$

where $\left\|L_{x_1|q}\right\|$ is finite since $L_{x_1|q}$ is bounded. Further, since $\tilde{m} = L_{x_1|q}^{-1} m$, the right-hand side converges to zero. ■

Our final preliminary result will use the following condition on $f_{x|q}(x|q)$:

Condition C.1 *(i) For all $q_1^* \neq q_2^* \in [0,1]$, the set $\left\{x : f_{x|q}(x|q_1^*) \neq f_{x|q}(x|q_2^*)\right\}$ has positive probability under f_x. (ii) $f_{x|q}(x|q)$ continuously differentiable in q. (iii) for all $q \in (0,1)$, there exists an x, such that $\frac{\partial f_{x|q}(x|q)}{\partial q} \neq 0$. (iv) $f_q(q) = 1\{q \in [0,1]\}$. (v) $\int_0^1 f_{x|q}(x|q)\, m(q)\, dq = 0$ for all x, implies that $m(q) = 0$. (vi) $f(x) = \int_0^1 f_{x|q}(x|q)\, dq$.*

Lemma C.6 *Consider two conditional densities $\tilde{f}_{\tilde{q}}(x|\tilde{q})$ and $f_q(x|q)$ satisfying Condition C.1. If there exists a bijection $Q : [0,1] \to [0,1]$ such that $\tilde{f}_{\tilde{q}}(x|Q(q)) = f_q(x|q)$ then Q is the identity.*

Proof. We will show that if there exists a reindexing $\tilde{f}_{x|\tilde{q}}(x|\tilde{q})$ of q via a bijection $Q : [0,1] \to [0,1]$ such that $\tilde{f}_{x|\tilde{q}}(x|Q(q)) = f_{x|q}(x|q)$ (Assumption 4 in Hu and Schennach (2008) requires $Q(q)$ to be injective on $[0,1]$ and the support assumption in the hypothesis implies surjectivity). If $f_q(q) = 1$ and $f_{\tilde{q}}(\tilde{q}) = 1$, then $Q(\cdot)$ is the identity.

By the assumptions of the theorem,

$$f(x) = \int_0^1 f_{x|q}(x|q)\, dq = \int_0^1 \tilde{f}_{x|\tilde{q}}(x|q)\, dq = \int_0^1 f_{x|q}(x|Q(q))\, dq.$$

A change of variables, $q' = Q(q)$ yields that

$$\int_0^1 f_{x|q}(x|Q(q))\, dq = \int_0^1 f_{x|q}(x|q')\, dQ^{-1}(q') = \int_0^1 f_{x|q}(x|q')\, \frac{1}{Q'(Q^{-1}(q'))} dq'.$$

The second inequality follows from the inverse function theorem. Differentiability of Q follows from the implicit function theorem: $Q(q)$ is defined implicitly from $\tilde{f}_{x|\tilde{q}}(x|Q(q)) - f_{x|q}(x|q) = 0$, where for every Q there exists an x such that $\tilde{f}_{x|\tilde{q}}(x|Q(q))$ has a non-zero derivative under the hypotheses of the theorem. Hence,

$$\int_0^1 f_{x|q}(x|q)\, dq - \int_0^1 f_{x|q}(x|Q(q))\, dq = 0 \Rightarrow \int_0^1 f_{x|q}(x|q) \left(1 - \frac{1}{Q'(Q^{-1}(q))}\right) dq = 0.$$

By Lemma C.3, for all $q \in [0,1]$, $\left(1 - \frac{1}{Q'(Q^{-1}(q))}\right) = 0$. Therefore $Q(\cdot)$ is the identity since $Q'(q) = 1$. ■

C.2 Detailed Proofs: Estimation

C.2.1 Lemmata Used in Proposition 3.3

We will use the following lemmata for the result, which are stated, for each dimension of ψ. We omit the dimension index for notational simplicity. Define $q_{N,V}(q) = F_V\left(F_{V_N}^{-1}(q)\right)$ and $q_{N,U}(q) = F_U\left(F_{U_N}^{-1}(q)\right)$.

Lemma C.7 *Suppose that f_V and f_U are continuous, and Γ_X and Γ_Z are respectively μ_X and μ_Z Donsker. The stochastic process defined by*

$$\begin{bmatrix} \sqrt{N}\left(q_{N,V}(q_X) - q_X\right) \\ \sqrt{N/2}\left(q_{N,U}(q_Z) - q_Z\right) \\ \sqrt{N}\left(\mu_{X_N} - \mu_X\right)(\gamma_X) \\ \sqrt{N/2}\left(\mu_{Z_N} - \mu_Z\right)(\gamma_Z) \end{bmatrix}$$

indexed by $q_X, q_Z \in [0,1]$, $\gamma_X \in \Gamma_X$ and $\gamma_Z \in \Gamma_Z$ is asymptotically equivalent to the empirical process

$$\begin{bmatrix} \sqrt{N}\left(\mu_{(X,\varepsilon)_N} - \mu_{X,\varepsilon}\right)\left(1\left\{h(x;\theta_0) + \varepsilon \leq F_V^{-1}(q_X)\right\}\right) \\ \sqrt{N/2}\left(\mu_{(Z,\eta)_N} - \mu_{Z,\eta}\right)\left(1\left\{g(z;\theta_0) + \eta \leq F_U^{-1}(q_Z)\right\}\right) \\ \sqrt{N}\left(\mu_{X_N} - \mu_X\right)(\gamma_X) \\ \sqrt{N/2}\left(\mu_{Z_N} - \mu_Z\right)(\gamma_Z) \end{bmatrix},$$

which converges weakly to the mean-zero Gaussian process with covariance kernel given by

$$\begin{aligned}
\Omega(q_X, q_Z) &= \Omega(q_Z, \gamma_X) = \Omega(q_X, \gamma_Z) = \Omega(\gamma_X, \gamma_Z) = 0 \\
\Omega(q_Z, \gamma_Z) &= \mu_{Z,\eta}\left(\gamma_Z 1\left\{g(z;\theta_0) + \eta \leq F_U^{-1}(q_Z)\right\}\right) - \\
&\quad \mu_Z(\gamma_Z)\,\mu_{Z,\eta}\left(1\left\{g(z;\theta_0) + \eta \leq F_U^{-1}(q_Z)\right\}\right) \\
\Omega(q_X, \gamma_X) &= \mu_{X,\varepsilon}\left(\gamma_X 1\left\{h(x;\theta_0) + \varepsilon \leq F_V^{-1}(q_X)\right\}\right) - \\
&\quad \mu_X(\gamma_X)\,\mu_{X,\varepsilon}\left(1\left\{h(x;\theta_0) + \varepsilon \leq F_V^{-1}(q_X)\right\}\right) \\
\Omega(\gamma_X, \gamma_X') &= \mu_X(\gamma_X\gamma_X') - \mu_X(\gamma_X)\,\mu_X(\gamma_X') \\
\Omega(\gamma_Z, \gamma_Z') &= \mu_Z(\gamma_Z\gamma_Z') - \mu_Z(\gamma_Z)\,\mu_Z(\gamma_Z') \\
\Omega(q_X, q_X') &= \mu_{X,\varepsilon}\left(1\left\{h(x;\theta_0) + \varepsilon \leq F_V^{-1}(q_X)\right\}1\left\{h(x;\theta_0) + \varepsilon \leq F_V^{-1}(q_X')\right\}\right) \\
&\quad -\mu_{X,\varepsilon}\left(1\left\{h(x;\theta_0) + \varepsilon \leq F_V^{-1}(q_X)\right\}\right)\mu_{X,\varepsilon}\left(1\left\{h(x;\theta_0) + \varepsilon \leq F_V^{-1}(q_X')\right\}\right) \\
\Omega(q_Z, q_Z') &= \mu_{Z,\eta}\left(1\left\{g(z;\theta_0) + \eta \leq F_U^{-1}(q_Z)\right\}1\left\{g(z;\theta_0) + \eta \leq F_U^{-1}(q_Z')\right\}\right) \\
&\quad -\mu_{Z,\eta}\left(1\left\{g(z;\theta_0) + \eta \leq F_U^{-1}(q_Z)\right\}\right)\mu_{Z,\eta}\left(1\left\{g(z;\theta_0) + \eta \leq F_U^{-1}(q_Z')\right\}\right),
\end{aligned}$$

where $q_X, q_X' \in [0,1]$, $q_Z, q_Z' \in [0,1]$, $\gamma_X, \gamma_X' \in \Gamma_X$ *and* $\gamma_Z, \gamma_Z' \in \Gamma_Z$.

Proof. Note that $F_{V_N}\left(F_V^{-1}(.)\right)$ is the empirical cumulative distribution function of a uniformly distributed random variable, and $q_{N,V} : [0,1] \to [0,1]$ is its associated quantile function. Therefore, $F_{V_N}\left(F_V^{-1}(.)\right)$ satisfied the assumptions for theorem 4 of Csorgo and Revesz (1978). This theorem implies that

$$\sqrt{N}\sup_{q\in[0,1]}\left|\left[F_{V_N}\left(F_V^{-1}(q)\right) - F_V\left(F_V^{-1}(q)\right)\right] - \left[F_V\left(F_{V_N}^{-1}(q)\right) - F_V\left(F_V^{-1}(q)\right)\right]\right| = o_p(1).$$

Since $q_{N,V}(q) = F_V\left(F_{V_N}^{-1}(q)\right)$ and $F_V\left(F_V^{-1}(q)\right) = q$, we therefore have that

$$\sqrt{N}\sup_{q\epsilon[0,1]}\left|\left[F_{V_N}\left(F_V^{-1}(q)\right) - F_V\left(F_V^{-1}(q)\right)\right] - \left[q_{N,V}(q) - q\right]\right| = o_p(1).$$

By an identical argument,

$$\sqrt{\frac{N}{2}} \sup_{q \epsilon [0,1]} \left| \left[F_{U_N} \left(F_U^{-1}(q) \right) - F_U \left(F_U^{-1}(q) \right) \right] - \left[q_{N,U}(q) - q \right] \right| = o_p(1).$$

Note that

$$F_{V_N} \left(F_V^{-1}(q) \right) - F_V \left(F_V^{-1}(q) \right) = \left(\mu_{(X,\varepsilon)_N} - \mu_{X,\varepsilon} \right) \left(1 \left\{ h(x;\theta_0) + \varepsilon \leq F_V^{-1}(q) \right\} \right)$$

and

$$F_{U_N} \left(F_U^{-1}(q) \right) - F_U \left(F_U^{-1}(q) \right) = \left(\mu_{(Z,\eta)_N} - \mu_{Z,\eta} \right) \left(1 \left\{ g(z;\theta_0) + \eta \leq F_U^{-1}(q) \right\} \right).$$

The result therefore follows from the functional central limit theorem, since the first and last two components of

$$\begin{bmatrix} \sqrt{N} \left(\mu_{(X,\varepsilon)_N} - \mu_{X,\varepsilon} \right) \left(1 \left\{ h(x;\theta_0) + \varepsilon \leq F_V^{-1}(q_X) \right\} \right) \\ \sqrt{N} \left(\mu_{X_N} - \mu_X \right)(\gamma_X) \\ \sqrt{N/2} \left(\mu_{(Z,\eta)_N} - \mu_{Z,\eta} \right) \left(1 \left\{ g(z;\theta_0) + \eta \leq F_U^{-1}(q_Z) \right\} \right) \\ \sqrt{N/2} \left(\mu_{Z_N} - \mu_Z \right)(\gamma_Z) \end{bmatrix}$$

are two independent empirical processes index by $\mu_{X,\varepsilon}$ and $\mu_{Z,\eta}$ Donsker classes. ■

Lemma C.8 *(i) If Assumption 3.4(i) is satisfied, $E(\psi_N | \mu_{V_N}, \mu_{U_N}) - \psi$ converges in probability to 0 as $N \to \infty$.*

(ii) If Assumption 3.4(ii) is satisfied, then for any bounded μ_X-Donsker class Γ_X and bounded μ_Z-Donsker class Γ_Z,

$$\begin{bmatrix} \sqrt{N} \left(E(\psi_N | \mu_{V_N}, \mu_{U_N}) - \psi \right) \\ \sqrt{N} \left(\mu_{X_N} - \mu_X \right)(\gamma_X) \\ \sqrt{N/2} \left(\mu_{Z_N} - \mu_Z \right)(\gamma_Z) \end{bmatrix}$$

is asymptotically equivalent to the process

$$\left[\begin{array}{c} \sqrt{N}\int_0^1 \nabla\tilde{\psi}_q\left(q,q,q\right)\cdot\left[\begin{array}{c}\left(\mu_{(X,\varepsilon)_N}-\mu_{X,\varepsilon}\right)\left(1\left\{h\left(x;\theta_0\right)+\varepsilon\leq F_V^{-1}\left(q_X\right)\right\}\right)\\ \left(\mu_{(X,\varepsilon)_N}-\mu_{X,\varepsilon}\right)\left(1\left\{h\left(x;\theta_0\right)+\varepsilon\leq F_V^{-1}\left(q_X\right)\right\}\right)\\ \left(\mu_{(Z,\eta)_N}-\mu_{Z,\eta}\right)\left(1\left\{g\left(z;\theta_0\right)+\eta\leq F_U^{-1}\left(q_Z\right)\right\}\right)\end{array}\right]dq \\ \sqrt{N}\left(\mu_{X_N}-\mu_X\right)\left(\gamma_X\right) \\ \sqrt{N/2}\left(\mu_{Z_N}-\mu_Z\right)\left(\gamma_Z\right)\end{array}\right],$$

which converges weakly to a mean-zero Gaussian process with covariance kernel given by

$$\begin{aligned}
V'\left(\gamma_X,\gamma_Z\right) &= 0\\
V'\left(\gamma_\Psi,\gamma_Z\right) &= \sqrt{2}\int_0^1\tilde{\psi}_{q,3}\left(q_Z,q_Z,q_Z\right)\Omega\left(q_Z,\gamma_Z\right)dq_Z\\
V'\left(\gamma_\Psi,\gamma_X\right) &= \int_0^1\left(\tilde{\psi}_{q,1}\left(q_X,q_X,q_X\right)+\tilde{\psi}_{q,2}\left(q_X,q_X,q_X\right)\right)\Omega\left(q_X,\gamma_X\right)dq_X\\
V'\left(\gamma_\Psi,\gamma_\Psi\right) &= \int_0^1\int_0^1\left(\tilde{\psi}_{q,1}\left(q_X,q_X,q_X\right)+\tilde{\psi}_{q,2}\left(q_X,q_X,q_X\right)\right)\\
&\quad\left(\tilde{\psi}_{q,1}\left(q'_X,q'_X,q'_X\right)+\tilde{\psi}_{q,2}\left(q'_X,q'_X,q'_X\right)\right)\Omega\left(q_X,q'_X\right)dq_Xdq'_X\\
&\quad+2\int_0^1\int_0^1\tilde{\psi}_{q,3}\left(q_Z,q_Z,q_Z\right)\tilde{\psi}_{q,3}\left(q'_Z,q'_Z,q'_Z\right)\Omega\left(q_Z,q'_Z\right)dq_Zdq'_Z\\
V'\left(\gamma_X,\gamma'_X\right) &= \Omega\left(\gamma_X,\gamma'_X\right)\\
V'\left(\gamma_Z,\gamma'_Z\right) &= \Omega\left(\gamma_Z,\gamma'_Z\right),
\end{aligned}$$

where $\gamma_X,\gamma'_X\in\Gamma_X$, $\gamma_Z,\gamma'_Z\in\Gamma_Z$ *and* γ_Ψ *indexes* $\sqrt{N}\left(E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)-\psi\right)$.

Proof. The quantity $E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)$ can be computed by using the fact that for all $1\leq k\leq J$, the k'th most desirable firm is occupied by the $2k$-th and the $(2k-1)$-th most desirable workers. By definition, the conditional expectation of $\Psi\left(x_1,x_2,z\right)$ given μ_{V_N}, μ_{U_N} for the k'th desirable job is $\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right),F_{V_N}^{-1}\left(\frac{2k}{N}\right),F_{U_N}^{-1}\left(\frac{k}{N/2}\right)\right)$ where F_{V_N} and F_{U_N} are the cdfs representing the empirical measures μ_{V_N} and μ_{U_N} respectively. Therefore,

$$\begin{aligned}
E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right) &= \frac{1}{N/2}\sum_{k=1}^{N/2}\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right),F_{V_N}^{-1}\left(\frac{2k}{N}\right),F_{U_N}^{-1}\left(\frac{k}{N/2}\right)\right)\\
&= \frac{1}{N}\sum_{i=1}^{N}\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{U_N}^{-1}\left(\frac{i}{N}\right)\right)+R. \qquad \text{(C.1)}
\end{aligned}$$

where

$$
\begin{aligned}
R &= \frac{1}{N/2}\sum_{k=1}^{N/2}\left[\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right),F_{V_N}^{-1}\left(\frac{2k}{N}\right),F_{U_N}^{-1}\left(\frac{k}{N/2}\right)\right)\right. \\
&\quad -\frac{1}{2}\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right),F_{V_N}^{-1}\left(\frac{2k-1}{N}\right),F_{U_N}^{-1}\left(\frac{2k-1}{N}\right)\right) \\
&\quad \left.-\frac{1}{2}\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k}{N}\right),F_{V_N}^{-1}\left(\frac{2k}{N}\right),F_{U_N}^{-1}\left(\frac{2k}{N}\right)\right)\right]. \qquad \text{(C.2)}
\end{aligned}
$$

Our proof of part (i) proceeds by showing that $R \to 0$ and that

$$
\frac{1}{N}\sum_{i=1}^{N}\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{U_N}^{-1}\left(\frac{i}{N}\right)\right)-\psi \to 0. \qquad \text{(C.3)}
$$

The proof of part (ii) is analogous. It characterizes the limit distribution of

$$
\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{U_N}^{-1}\left(\frac{i}{N}\right)\right)-\psi\right)
$$

under stronger assumptions.

Proof of Part (i): We begin by bounding the absolute value of R in equation (C.2) using the triangle inequality as:

$$
\begin{aligned}
|R| &\leq \frac{1}{N}\sum_{k=1}^{N/2}\left|\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right),F_{V_N}^{-1}\left(\frac{2k}{N}\right),F_{U_N}^{-1}\left(\frac{2k}{N}\right)\right)\right. \\
&\quad \left.-\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k}{N}\right),F_{V_N}^{-1}\left(\frac{2k}{N}\right),F_{U_N}^{-1}\left(\frac{2k}{N}\right)\right)\right| \\
&\quad +\frac{1}{N}\sum_{k=1}^{N/2}\left|\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right),F_{V_N}^{-1}\left(\frac{2k}{N}\right),F_{U_N}^{-1}\left(\frac{2k}{N}\right)\right)\right. \\
&\quad \left.-\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right),F_{V_N}^{-1}\left(\frac{2k-1}{N}\right),F_{U_N}^{-1}\left(\frac{2k}{N}\right)\right)\right|.
\end{aligned}
$$

For any $\delta \in \left(0, \frac{1}{2}\right)$, we have that:

$$\begin{aligned}
|R| &\leq \frac{1}{N} \sum_{\lceil \delta N/2 \rceil < k < \lfloor (1-\delta)N/2 \rfloor} \left| \tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right), F_{V_N}^{-1}\left(\frac{2k}{N}\right), F_{U_N}^{-1}\left(\frac{2k}{N}\right)\right)\right. \\
&\quad \left. -\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k}{N}\right), F_{V_N}^{-1}\left(\frac{2k}{N}\right), F_{U_N}^{-1}\left(\frac{2k}{N}\right)\right)\right| \\
&\quad + \frac{1}{N} \sum_{\lceil \delta N/2 \rceil < k < \lfloor (1-\delta)N/2 \rfloor} \left| \tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right), F_{V_N}^{-1}\left(\frac{2k}{N}\right), F_{U_N}^{-1}\left(\frac{2k}{N}\right)\right)\right. \\
&\quad \left. -\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right), F_{V_N}^{-1}\left(\frac{2k-1}{N}\right), F_{U_N}^{-1}\left(\frac{2k}{N}\right)\right)\right| \\
&\quad + 4\delta \left\| \Psi \right\|_\infty \\
&= \tilde{R} + 4\delta \left\| \Psi \right\|_\infty . \qquad \text{(C.4)}
\end{aligned}$$

Since $\tilde{\psi}$ is Lipschitz continuous

$$\tilde{R} \leq \sup_{\lceil J\delta \rceil < k < \lfloor J(1-\delta) \rfloor} |\tilde{\psi}|_{LC} \left[2 \left| F_{V_N}^{-1}\left(\frac{2k-1}{N}\right) - F_{V_N}^{-1}\left(\frac{2k}{N}\right)\right|\right], \qquad \text{(C.5)}$$

where $|\tilde{\psi}|_{LC}$ denotes the Lipschitz constant. By Example 3.9.21 in van der Vaart and Wellner (2000), for all $\lceil \delta N/2 \rceil < k < \lfloor (1-\delta) N/2 \rfloor$ $\left| F_{V_N}^{-1}\left(\frac{2k-1}{N}\right) - F_{V_N}^{-1}\left(\frac{2k}{N}\right)\right|$ converges in probability to 0 uniformly in k (Assumption 3.4(i)b. implies that f_V is continuous with full support). Therefore, since $\tilde{R} \geq 0$, it converges in probability to 0.

Now, we show that the difference in equation (C.3) converges in probability to 0. Note that F_{U_N} is constant on each interval $[\frac{k-1}{N/2}, \frac{k}{N/2})$ and F_{V_N} is constant on $[\frac{i-1}{N}, \frac{i}{N})$. Hence,

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} \tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{U_N}^{-1}\left(\frac{i}{N}\right)\right) - \psi \\
&= \int_0^1 \tilde{\psi}\left(F_{V_N}^{-1}(q), F_{V_N}^{-1}(q), F_{U_N}^{-1}(q)\right) dq - \int_0^1 \tilde{\psi}\left(F_V^{-1}(q), F_V^{-1}(q), F_U^{-1}(q)\right) dq \\
&= \int_\delta^{1-\delta} \left[\tilde{\psi}\left(F_{V_N}^{-1}(q), F_{V_N}^{-1}(q), F_{U_N}^{-1}(q)\right) - \tilde{\psi}\left(F_V^{-1}(q), F_V^{-1}(q), F_U^{-1}(q)\right)\right] dq \\
&\quad + \left(\int_0^\delta + \int_{1-\delta}^1 \right) \left[\tilde{\psi}\left(F_{V_N}^{-1}(q), F_{V_N}^{-1}(q), F_{U_N}^{-1}(q)\right) - \tilde{\psi}\left(F_V^{-1}(q), F_V^{-1}(q), F_U^{-1}(q)\right)\right] dq \\
&= T_1 + T_2 \qquad \text{(C.6)}
\end{aligned}$$

where $\delta \in \left(0, \frac{1}{2}\right)$.

We now bound T_1 and T_2 in terms of δ. Since $\|\Psi\|_\infty < \infty$, $|T_2| \leq 4\delta \|\Psi\|_\infty$. To bound T_1, note that

$$\begin{aligned}
|T_1| &= \left| \int_\delta^{1-\delta} \left[\tilde{\psi}\left(F_{V_N}^{-1}(q), F_{V_N}^{-1}(q), F_{U_N}^{-1}(q) \right) - \tilde{\psi}\left(F_V^{-1}(q), F_V^{-1}(q), F_U^{-1}(q) \right) \right] dq \right| \\
&\leq \int_\delta^{1-\delta} \left| \tilde{\psi}\left(F_{V_N}^{-1}(q), F_{V_N}^{-1}(q), F_{U_N}^{-1}(q) \right) - \tilde{\psi}\left(F_V^{-1}(q), F_V^{-1}(q), F_U^{-1}(q) \right) \right| dq \\
&\leq \sup_{q \in [\delta, 1-\delta]} \left| \tilde{\psi}\left(F_{V_N}^{-1}(q), F_{V_N}^{-1}(q), F_{U_N}^{-1}(q) \right) - \tilde{\psi}\left(F_V^{-1}(q), F_V^{-1}(q), F_U^{-1}(q) \right) \right| \\
&\leq |\tilde{\psi}|_{LC} \sup_{q \in [\delta, 1-\delta]} \left| \left(F_{V_N}^{-1}(q), F_{V_N}^{-1}(q), F_{U_N}^{-1}(q) \right) - \left(F_V^{-1}(q), F_V^{-1}(q), F_U^{-1}(q) \right) \right|. \qquad \text{(C.7)}
\end{aligned}$$

Combining equations (C.1) - (C.7) and the bound on T_2, we have that

$$\begin{aligned}
& |E(\psi_N | \mu_{V_N}, \mu_{U_N}) - \psi| \\
\leq\ & \left| \frac{1}{N} \sum_{i=1}^N \tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{U_N}^{-1}\left(\frac{i}{N}\right) \right) - \psi \right| + |R| \\
\leq\ & |T_1| + |T_2| + \left|\tilde{R}\right| + 4\delta \|\Psi\|_\infty \\
\leq\ & |\tilde{\psi}|_{LC} \sup_{q \in [\delta, 1-\delta]} \left| \left(F_{V_N}^{-1}(q), F_{V_N}^{-1}(q), F_{U_N}^{-1}(q) \right) - \left(F_V^{-1}(q), F_V^{-1}(q), F_U^{-1}(q) \right) \right| + \\
& 8\delta \|\Psi\|_\infty + o_p(1)
\end{aligned}$$

since $|T_2| \leq 4\delta \|\Psi\|_\infty$ and $\left|\tilde{R}\right| = o_p(1)$.

We now show that $|E(\psi_N | \mu_{V_N}, \mu_{U_N}) - \psi| \to 0$ in probability as $N \to \infty$. Fix $\varepsilon > 0$ and choose $\delta = \frac{\varepsilon}{16\|\Psi\|_\infty}$. By Example 3.9.21 in van der Vaart and Wellner (2000),

$$\sup_{q \in [\delta, 1-\delta]} \left| \left(F_{V_N}^{-1}(q), F_{V_N}^{-1}(q), F_{U_N}^{-1}(q) \right) - \left(F_V^{-1}(q), F_V^{-1}(q), F_U^{-1}(q) \right) \right|$$

converges in probability to 0 (Assumption 3.4(i)b. implies that f_V and f_U are continuous with full support). Hence, for sufficiently large N we have

$$P\left(|\tilde{\psi}|_{LC} \sup_{q \in [\delta, 1-\delta]} \left| \left(F_{V_N}^{-1}(q), F_{V_N}^{-1}(q), F_{U_N}^{-1}(q) \right) - \left(F_V^{-1}(q), F_V^{-1}(q), F_U^{-1}(q) \right) \right| > \frac{\varepsilon}{2} \right) < \varepsilon.$$

This implies $P\left(\left|E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)-\psi\right|>\varepsilon\right)<\varepsilon$, proving the desired convergence in probability to 0.

Proof of Part (ii): Let $q_{N,V}(q)=F_V\left(F_{V_N}^{-1}(q)\right)$, $q_{N,U}(q)=F_U\left(F_{U_N}^{-1}(q)\right)$, and

$$\tilde{\psi}_q(q_1,q_2,q_3)=\tilde{\psi}\left(F_V^{-1}(q_1),F_V^{-1}(q_2),F_U^{-1}(q_3)\right).$$

Equation (C.2) can be rewritten by using this notation as

$$\begin{aligned}R &= \frac{1}{N/2}\sum_{k=1}^{N/2}\left[\tilde{\psi}_q\left(q_{N,V}\left(\frac{2k-1}{N}\right),q_{N,V}\left(\frac{2k}{N}\right),q_{N,U}\left(\frac{k}{N/2}\right)\right)\right.\\ &\quad -\frac{1}{2}\tilde{\psi}_q\left(q_{N,V}\left(\frac{2k-1}{N}\right),q_{N,V}\left(\frac{2k-1}{N}\right),q_{N,U}\left(\frac{k}{N/2}\right)\right)\\ &\quad \left.-\frac{1}{2}\tilde{\psi}_q\left(q_{N,V}\left(\frac{2k}{N}\right),q_{N,V}\left(\frac{2k}{N}\right),q_{N,U}\left(\frac{k}{N/2}\right)\right)\right].\end{aligned}$$

By the triangle inequality and the assumption that $\tilde{\psi}_q$ has a bounded derivative,

$$|R|\leq\frac{1}{N}\sum_{k=1}^{N/2}\left\|\nabla\tilde{\psi}_q\right\|_\infty 2\left|q_{N,V}\left(\frac{2k-1}{N}\right)-q_{N,V}\left(\frac{2k}{N}\right)\right|.$$

Since $q_{N,V}(q)$ is monotonic in q and has range $[0,1]$, we have that

$$\begin{aligned}&\sum_{k=1}^{N/2}\left|q_{N,V}\left(\frac{2k-1}{N}\right)-q_{N,V}\left(\frac{2k}{N}\right)\right|\\ =\ &\left|\sum_{k=1}^{N/2}q_{N,V}\left(\frac{2k-1}{N}\right)-q_{N,V}\left(\frac{2k}{N}\right)\right|\\ \leq\ &1.\end{aligned}$$

Therefore, since $\left\|\nabla\tilde{\psi}_q\right\|_\infty<\infty$,

$$\sqrt{N}\,|R|\leq\frac{1}{\sqrt{N}}\left\|\nabla\tilde{\psi}_q\right\|_\infty\to 0. \tag{C.8}$$

Now, we compute the limit distribution of

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{U_N}^{-1}\left(\frac{i}{N}\right)\right)-\psi\\ =\ &\frac{1}{N}\sum_{i=1}^{N}\tilde{\psi}_q\left(q_{N,V}\left(\frac{i}{N}\right),q_{N,V}\left(\frac{i}{N}\right),q_{N,U}\left(\frac{i}{N}\right)\right)-\psi\\ =\ &\int_0^1\tilde{\psi}_q\left(q_{N,V}(q),q_{N,V}(q),q_{N,U}(q)\right)dq-\int_0^1\tilde{\psi}_q(q,q,q)\,dq.\end{aligned}$$

By Taylor's theorem,

$$\begin{aligned}
&\tilde{\psi}_q\left(q_{N,V}\left(q\right), q_{N,V}\left(q\right), q_{N,U}\left(q\right)\right) - \tilde{\psi}_q\left(q,q,q\right) \\
= \;& \tilde{\psi}_{q,1}\left(q,q,q\right)\left(q_{N,V}\left(q\right)-q\right) + \tilde{\psi}_{q,2}\left(q,q,q\right)\left(q_{N,V}\left(q\right)-q\right) + \tilde{\psi}_{q,3}\left(q,q,q\right)\left(q_{N,U}\left(q\right)-q\right) + R_q.
\end{aligned}$$

Since,

$$\sup_q \left|R_q\right| = o\left(\sup_q \left\|q_{N,V}\left(q\right)-q\right\| + \sup_q \left\|q_{N,V}\left(q\right)-q\right\| + \sup_q \left\|q_{N,U}\left(q\right)-q\right\|\right),$$

we have that $\sqrt{N}\sup_q \left|R_q\right| \to_p 0$. Therefore,

$$\begin{aligned}
&\sqrt{N}\left(E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)-\psi\right) \\
= \;& \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{U_N}^{-1}\left(\frac{i}{N}\right)\right)-\psi\right) + o_p\left(1\right) \\
= \;& \sqrt{N}\int_0^1 \nabla\tilde{\psi}_q\left(q,q,q\right)\cdot\begin{bmatrix} q_{N,V}\left(q\right)-q \\ q_{N,V}\left(q\right)-q \\ q_{N,U}\left(q\right)-q \end{bmatrix} dq + o_p\left(1\right),
\end{aligned}$$

showing the required asymptotic equivalence. Lemma C.7 characterizes the limit distribution of

$$\begin{bmatrix} \sqrt{N}\left(q_{N,V}\left(q_X\right)-q_X\right) \\ \sqrt{N/2}\left(q_{N,U}\left(q_Z\right)-q_Z\right) \\ \sqrt{N}\left(\mu_{X_N}-\mu_X\right)\left(\gamma_X\right) \\ \sqrt{N/2}\left(\mu_{Z_N}-\mu_Z\right)\left(\gamma_Z\right) \end{bmatrix}$$

indexed by $q_X, q_Z \in [0,1]$, $\gamma_X \in \Gamma_X$ and $\gamma_Z \in \Gamma_Z$. Therefore,

$$\begin{bmatrix} \sqrt{N}\left(E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)-\psi\right) \\ \sqrt{N}\left(\mu_{X_N}-\mu_X\right)\left(\gamma_X\right) \\ \sqrt{N/2}\left(\mu_{Z_N}-\mu_Z\right)\left(\gamma_Z\right) \end{bmatrix}$$

converges to a mean-zero Gaussian process with covariance kernel V'. ■

Lemma C.9 *(i)* $\psi_N - E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)$ *converges in probability to* 0 *if* $\left\|\Psi\right\|_\infty < \infty$.

(ii) Suppose Assumption 3.4(ii)b is satisfied. For any bounded functions $\gamma_{X,\varepsilon}$ on the domain of (X,ε) and $\gamma_{Z,\eta}$ on the domain of (Z,η),

$$\sqrt{N/2}\left[\psi_N - E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right), \left(\mu_{(X,\varepsilon)_N} - \mu_{X,\varepsilon}\right)(\gamma_Z)\ , \left(\mu_{(Z,\eta)_N} - \mu_{Z,\eta}\right)(\gamma_Z)\right]$$

converges to a multivariate normal distribution with mean 0 and covariance kernel $V''(\gamma_X,\gamma_Z) =$

$$\begin{bmatrix} \int_0^1 var_{q,\Psi}(q,q,q)\,dq & \int_0^1 cov_q\left(\Psi,\gamma_{X,\varepsilon}|q,q,q\right)dq & \int_0^1 cov_q\left(\Psi,\gamma_{Z,\eta}|q,q,q\right)dq \\ \int_0^1 cov_q\left(\Psi,\gamma_{X,\varepsilon}|q,q,q\right)dq & \frac{1}{2}Var\left(\gamma_{X,\varepsilon}\right) & 0 \\ \int_0^1 cov_q\left(\Psi,\gamma_{Z,\eta}|q,q,q\right)dq & 0 & Var\left(\gamma_{Z,\eta}\right) \end{bmatrix}.$$

Proof. Let $v^{(k)}$ and $u^{(k)}$ be k'th order statistics of worker and firm desirability and let $\left(X^{(k)},\varepsilon^{(k)}\right)$ and $\left(Z^{(k)},\eta^{(k)}\right)$ be the corresponding observations drawn from $\mu_{X,\varepsilon|v^{(k)}}$ and $\mu_{Z,\eta|u^{(k)}}$ respectively. We will use $J = N/2$ in this proof. Rewrite:

$$\psi_N - E\left(\psi_N|\mu_{U_N},\mu_{V_N}\right) = \frac{1}{J}\left(\sum_{k=1}^{J}\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(i)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right).$$

Proof of Part (i): The conditional variance of $\psi_N - E\left(\psi_N|\mu_{U_N},\mu_{V_N}\right)$ given (μ_{V_N},μ_{U_N}) is

$$\begin{aligned} & \frac{1}{J^2}E\left(\left(\sum_{i=1}^{J}\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)^2\middle|\,\mu_{V_N},\mu_{U_N}\right) \\ = & \frac{1}{J^2}\sum_{i=1}^{J}E\left(\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)^2\middle|\,\mu_{V_N},\mu_{U_N}\right) \\ \leq & \frac{1}{J}4\left\|\Psi\right\|_\infty^2, \end{aligned}$$

where the first equality follows from conditional independence.

However, since $\psi_N - E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)$ is by definition mean zero, it follows that the unconditional variance of $\psi_N - E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)$ is bounded above by $\frac{1}{J}4\left\|\Psi\right\|_\infty^2$, by the law of total variance. By Chebychev's inequality, $\sqrt{J}\left(\psi_N - E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right) = O_p(1)$ and thus $\psi_N - E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right) = o_p(1)$.

Proof of Part (ii): We will show that the random variables

$$\begin{aligned}
\psi_N - E\left(\psi_N | \mu_{V_N}, \mu_{U_N}\right) &= \frac{1}{J}\left(\sum_{k=1}^{J} \Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\right), \\
\left(\mu_{(X,\varepsilon)_N} - \mu_{X,\varepsilon}\right)(\gamma_{X,\varepsilon}) &= \frac{1}{2J}\sum_{k=1}^{2J} \gamma_{X,\varepsilon}\left(X^{(k)}, \varepsilon^{(k)}\right) - E\left(\gamma_{X,\varepsilon}\right), \text{ and} \\
\left(\mu_{(Z,\eta)_N} - \mu_{Z,\eta}\right)(\gamma_Z, \eta) &= \frac{1}{J}\sum_{k=1}^{J} \gamma_{Z,\eta}\left(Z^{(k)}, \eta^{(k)}\right) - E\left(\gamma_Z, \eta\right), \qquad \text{(C.9)}
\end{aligned}$$

are jointly asymptotically normal. The latter two random variables are jointly asymptotically normal by the standard CLT. We will characterize the joint limiting distribution of these three random variables by calculating their joint moment generating function and comparing it with the moment generating function of a normal random variable. We do this by computing the limiting variance-covariance matrices of the first random variable with each of the other two (note that the second and third random variables are independent), and then using a Taylor expansion of the moment generating function to show that the leading terms match the moment generating function of a normal random variable and that higher order terms are asymptotically negligible.

The sample variances of $\gamma_{X,\varepsilon}$ and $\gamma_{Z,\eta}$ and their covariance converge in probability to $Var\left(\gamma_{X,\varepsilon}\right)$, $Var\left(\gamma_{Z,\eta}\right)$ and 0 by the standard law of large numbers.

To show that the sample variances converge, we show that the second moment of the sample variances (of the random variables above) converge to 0. If these variance of the sample variances converge to 0, then the relevant sample variances will converge in probability (by Chebychev's inequality). To bound the variance of the first sample variance, by the law of total variance, let rewrite

$$\begin{aligned}
& Var\left(\frac{1}{J}\sum_{k=1}^{J}\left[\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\right]^2\right) \\
= \; & \frac{1}{J^2} E\left(\sum_{k=1}^{J} Var\left[\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) \middle| \mu_{V_N}, \mu_{U_N}\right]^2\right) \\
& + Var\left(\frac{1}{J}\sum_{k=1}^{J} E\left[\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) \middle| \mu_{V_N}, \mu_{U_N}\right]^2\right) \\
= \; & T_1 + T_2.
\end{aligned}$$

To bound the variance of the sample covariance of the first and second random variables, rewrite

$$\begin{aligned}
& Var\left(\frac{1}{J}\sum_{k=1}^{J}\left[\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right]\right. \\
& \left.\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right) \\
= \; & \frac{1}{J^2}E\left(\sum_{k=1}^{J}Var\left\{\left[\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right]\right.\right. \\
& \left.\left.\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right|\mu_{V_N},\mu_{U_N}\right\}\right) \\
& +Var\left(\frac{1}{J}\sum_{k=1}^{J}E\left\{\left[\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right]\right.\right. \\
& \left.\left.\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right|\mu_{V_N},\mu_{U_N}\right\}\right) \\
= \; & R_1+R_2.
\end{aligned}$$

To bound the variance of the sample covariance of the first and third random variables, let

$$v(k)=\left[\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right]\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]$$

and rewrite

$$\begin{aligned}
& Var\left(\frac{1}{J}\sum_{k=1}^{J}\left[\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right]\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\right) \\
& = \frac{1}{J^2}E\sum_{k=1}^{J}Var(v(k)|\mu_{V_N},\mu_{U_N}) \\
& +Var\left(\frac{1}{J}\sum_{k=1}^{J}Ev(k)|\mu_{V_N},\mu_{U_N}\right) \\
& = V_1+V_2.
\end{aligned}$$

Note that T_1, R_1 and V_1 are the sum of J bounded terms divided by J^2 and hence converge in probability to 0. To show that T_2, R_2 and V_2 converge in probability to 0, we

compute the relevant conditional expectations. For T_2, we have that

$$\frac{1}{J}\sum_{k=1}^{J} E\left(\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\middle|\, \mu_{V_N}, \mu_{U_N}\right)^2 = 0$$

since

$$\begin{aligned} & E\left(\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\middle|\, \mu_{V_N}, \mu_{U_N}\right) \\ = \;& 0 \end{aligned}$$

since the conditional expectation $\tilde{\psi}$ satisfies this by definition.

For later calculations, it will be useful to compute the variance of $\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right)$ conditional on μ_{V_N}, μ_{U_N}.

$$\begin{aligned} & \frac{1}{J}\sum_{k=1}^{J}\left[E\left(\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right)^2\middle|\, \mu_{V_N}, \mu_{U_N}\right) - E\left(\tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\middle|\, \mu_{V_N}, \mu_{U_N}\right)^2\right] \\ = \;& \frac{1}{J}\sum_{k=1}^{J} var\left(\Psi|v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) \\ = \;& \frac{1}{2J}\sum_{k=1}^{J}\left(var\left(\Psi|v^{(2k)}, v^{(2k)}, u^{(k)}\right) + var\left(\Psi|v^{(2k-1)}, v^{(2k-1)}, u^{(k)}\right)\right) \\ & -\frac{1}{2J}\sum_{k=1}^{J}\left(var\left(\Psi|v^{(2k)}, v^{(2k)}, u^{(k)}\right) - var\left(\Psi|v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\right) \\ & -\frac{1}{2J}\sum_{k=1}^{J}\left(var\left(\Psi|v^{(2k-1)}, v^{(2k-1)}, u^{(k)}\right) - var\left(\Psi|v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\right) \end{aligned} \tag{C.10}$$

The first term in the summation is

$$\begin{aligned} & \frac{1}{2J}\left(\sum_{k=1}^{J} var\left(\Psi|v^{(2k)}, v^{(2k)}, u^{(k)}\right) + \sum_{k=1}^{J} var\left(\Psi|v^{(2k-1)}, v^{(2k-1)}, u^{(k)}\right)\right) \\ = \;& \int_0^1 var_q\left(\Psi|q_{N,V}(q), q_{N,V}(q), q_{N,U}(q)\right) dq. \end{aligned}$$

Since $\|\Psi\|_\infty < \infty$, by the dominated convergence theorem,

$$\begin{aligned} & \frac{1}{2J}\left(\sum_{k=1}^{J} var\left(\Psi|v^{(2k)}, v^{(2k)}, u^{(k)}\right) + \sum_{k=1}^{J} var\left(\Psi|v^{(2k-1)}, v^{(2k-1)}, u^{(k)}\right)\right) \\ \to \;& \int_0^1 var_q\left(\Psi|q, q, q\right) dq \end{aligned} \tag{C.11}$$

almost surely. Note that

$$\begin{aligned}
& \left| \frac{1}{2J} \sum_{k=1}^{J} var\left(\Psi | v^{(2k)}, v^{(2k)}, u^{(k)}\right) - var\left(\Psi | v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) \right| \\
\leq & \left| \frac{1}{2J} \sum_{k=1}^{J} E\left(\Psi^2 | v^{(2k)}, v^{(2k)}, u^{(k)}\right) - E\left(\Psi^2 | v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) \right| \\
& + \left| \frac{1}{2J} \sum_{k=1}^{J} E\left(\Psi | v^{(2k)}, v^{(2k)}, u^{(k)}\right)^2 - E\left(\Psi | v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)^2 \right| \\
\leq & \left| \frac{1}{2J} \sum_{k=1}^{J} \int_0^{\|\Psi\|_\infty^2} \left(P\left(\Psi^2 \geq c | v^{(2k)}, v^{(2k)}, u^{(k)}\right) - P\left(\Psi^2 \geq c | v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) \right) dc \right| \\
& + \left| \frac{\|\Psi\|_\infty}{J} \sum_{k=1}^{J} \int_0^{\|\Psi\|_\infty} \left(P\left(\Psi \geq c | v^{(2k)}, v^{(2k)}, u^{(k)}\right) - P\left(\Psi \geq c | v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) \right) dc \right| \\
\to & \ 0
\end{aligned} \tag{C.12}$$

since

$$P\left(\Psi^2 \geq c | v_1, v_2, u\right) \quad = \quad \int 1\left\{\Psi\left(x_1, x_2, z\right)^2 \geq c\right\} d\mu_{X_1|v} d\mu_{X_2|v_2} d\mu_{Z|u}$$

is continuous in v_1, v_2 and u (implied by Assumption 3.4(ii)b), and $\|\Psi\|_\infty < \infty$.

Therefore, by equations (C.10), (C.11) and (C.12),

$$\frac{1}{J} \sum_{k=1}^{J} E\left(\left(\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) \right)^2 \middle| \mu_{V_N}, \mu_{U_N} \right) \to \int_0^1 var\left(\psi | q, q, q\right) dq \tag{C.13}$$

almost surely.

Similarly, for R_2,

$$\begin{aligned}
& \frac{1}{J}\sum_{k=1}^{J} E\left\{\left(\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\right)\right. \\
& \left.\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)}, \varepsilon^{(2k-1)}\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)}, \varepsilon^{(2k)}\right) - E\left(\gamma_{X,\varepsilon}\right)\right] \middle| \mu_{V_N}, \mu_{U_N}\right\} \\
= & \frac{1}{J}\sum_{k=1}^{J} E\left\{\left(\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\right)\right. \\
& \left(\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)}, \varepsilon^{(2k-1)}\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)}, \varepsilon^{(2k)}\right)\right. \\
& \left.\left.-\frac{1}{2}E\left[\gamma_{X,\varepsilon}\left(X^{(2k-1)}, \varepsilon^{(2k-1)}\right) + \gamma_{X,\varepsilon}\left(X^{(2k)}, \varepsilon^{(2k)}\right) \middle| \mu_{V_N}, \mu_{U_N}\right]\right) \middle| \mu_{V_N}, \mu_{U_N}\right\} \\
& + \frac{1}{J}\sum_{k=1}^{J} E\left\{\left(\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\right) \middle| \mu_{V_N}, \mu_{U_N}\right\} \times \\
& \left(\frac{1}{2}E\left[\gamma_{X,\varepsilon}\left(X^{(2k-1)}, \varepsilon^{(2k-1)}\right) + \gamma_{X,\varepsilon}\left(X^{(2k)}, \varepsilon^{(2k)}\right) \middle| \mu_{V_N}, \mu_{U_N}\right] - E\left(\gamma_{X,\varepsilon}\right)\right) \\
= & \frac{1}{J}\sum_{k=1}^{J} cov\left(\Psi, \gamma_{X,\varepsilon} | v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) \\
& + \frac{1}{J}\sum_{k=1}^{J} 0 \times \left(\frac{1}{2}E\left[\gamma_{X,\varepsilon}\left(X^{(2k-1)}, \varepsilon^{(2k-1)}\right) + \gamma_{X,\varepsilon}\left(X^{(2k)}, \varepsilon^{(2k)}\right) \middle| \mu_{V_N}, \mu_{U_N}\right] - E\left(\gamma_{X,\varepsilon}\right)\right) \\
= & \int_0^1 cov_q\left(\Psi, \gamma_{X,\varepsilon} | q, q, q\right) dq + o\left(1\right) \qquad \text{(C.14)}
\end{aligned}$$

where the last equality follows from arguments identical to showing equation (C.13) and

$$\begin{aligned}
& cov\left(\Psi, \gamma_{X,\varepsilon} | v_1, v_2, u\right) = \\
& \int \Psi\left(x_1, x_2, z\right)\left(\frac{1}{2}\gamma_{X,\varepsilon}\left(x_1, \varepsilon_1\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(x_2, \varepsilon_2\right)\right) d\mu_{(X_1,\varepsilon_1)|v_1} d\mu_{(X_2,\varepsilon_2)|v_2} d\mu_{Z|u} - \\
& \frac{1}{2}\int \Psi\left(x_1, x_2, z\right) d\mu_{X_1|v_1} d\mu_{X_2|v_2} d\mu_{Z|u} \int \left(\gamma_{X,\varepsilon}\left(x_1, \varepsilon_1\right) + 2\gamma_{X,\varepsilon}\left(x_2, \varepsilon_2\right)\right) d\mu_{(X_1,\varepsilon_1)|v_1} d\mu_{(X_2,\varepsilon_2)|v_2}.
\end{aligned}$$

Similarly, for V_2, we have that

$$\begin{aligned}
& \frac{1}{J}\sum_{k=1}^{J} E(v(k) | \mu_{V_N}, \mu_{U_N}) \qquad \text{(C.15)} \\
= & \int_0^1 cov_q\left(\Psi, \gamma_{Z,\eta} | q, q, q\right) dq + o\left(1\right). \qquad \text{(C.16)}
\end{aligned}$$

Note that these three calculations imply T_2, R_2, and V_2 are variances of bounded random variables which converge in probability, and hence converge to 0. It follows that the sample

variances converge in probability to their mean, which we now compute.

By the law of iterated expectations and arguments identical to showing equation (C.13),

$$\begin{aligned} & E\frac{1}{J}\sum_{k=1}^{J}\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)^2 \\ = \; & E\frac{1}{J}\sum_{k=1}^{J}E\left\{\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)^2\middle|\mu_{V_N},\mu_{U_N}\right\} \\ = \; & \int_0^1 var\left(\Psi|q,q,q\right)dq+o\left(1\right) \end{aligned} \tag{C.17}$$

and

$$\begin{aligned} & E\frac{1}{J}\sum_{k=1}^{J}\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)* \qquad (C.18) \\ & \left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right] \\ & =\int_0^1 cov_q\left(\Psi,\gamma_{X,\varepsilon}|\left(q,q,q\right)\right)dq+o\left(1\right) \qquad (C.19) \end{aligned}$$

and

$$\begin{aligned} & E\frac{1}{J}\sum_{k=1}^{J}\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right] \\ = \; & \int_0^1 cov_q\left(\Psi,\gamma_{Z,\eta}|q,q,q\right)dq+o\left(1\right) \end{aligned} \tag{C.20}$$

This characterizes the asymptotic variance of the random variables in equation (C.9).

We now characterize the limiting distribution by computing the limit of the moment generating function. For arbitrary $C_1, C_2, C_3 > 0$ we must compute

$$E\left(\exp C_1\left[\psi_N-E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right]+C_2\left(\mu_{(X,\varepsilon)_N}-\mu_{X,\varepsilon}\right)\left(\gamma_{X,\varepsilon}\right)+C_3\left(\mu_{(Z,\eta)_N}-\mu_{Z,\eta}\right)\left(\gamma_{Z,\eta}\right)\right).$$

$$
\begin{aligned}
& E\exp\langle(C_1\sqrt{J}\left[\psi_N - E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right] + C_2\sqrt{J}\left(\mu_{(X,\varepsilon)_N} - \mu_{X,\varepsilon}\right)(\gamma_{X,\varepsilon}) + \\
& C_3\sqrt{J}\left(\mu_{(Z,\eta)_N} - \mu_{Z,\eta}\right)(\gamma_{Z,\eta}))\rangle \\
= \; & E\exp\frac{1}{\sqrt{J}}\left(\sum_{k=1}^{J} C_1\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\right. \\
& + \sum_{k=1}^{J} C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right) - E\left(\gamma_{X,\varepsilon}\right)\right] \\
& \left. + \sum_{k=1}^{J} C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right) - E\left(\gamma_{Z,\eta}\right)\right]\right) \\
= \; & E\Pi_{k=1}^{J}\exp\frac{1}{\sqrt{J}}\left(C_1\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\right. \\
& + C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right) - E\left(\gamma_{X,\varepsilon}\right)\right] \\
& \left. + C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right) - E\left(\gamma_{Z,\eta}\right)\right]\right)
\end{aligned}
$$

By the Law of Iterated Expectations, this equals

$$
\begin{aligned}
& E\left[E\left[\Pi_{k=1}^{J}\exp\frac{1}{\sqrt{J}}\left(C_1\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\right.\right. \\
& + C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right) - E\left(\gamma_{X,\varepsilon}\right)\right] \\
& \left.\left. + C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right) - E\left(\gamma_{Z,\eta}\right)\right]\right)\Big|\,\mu_{V_N},\mu_{U_N}\right]\Big] \\
= \; & E\Pi_{k=1}^{J}E\left[\exp\frac{1}{\sqrt{J}}\left(C_1\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\right.\right. \\
& + C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right) - E\left(\gamma_{X,\varepsilon}\right)\right] \\
& \left.\left. + C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right) - E\left(\gamma_{Z,\eta}\right)\right]\right)\Big|\,\mu_{V_N},\mu_{U_N}\right] \\
= \; & E\exp\sum_{k=1}^{J}\log E\left[\exp\frac{1}{\sqrt{J}}\left(C_1\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\right.\right. \\
& + C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right) - E\left(\gamma_{X,\varepsilon}\right)\right] \\
& \left.\left. + C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right) - E\left(\gamma_{Z,\eta}\right)\right]\right)\Big|\,\mu_{V_N},\mu_{U_N}\right]
\end{aligned}
$$

where the first equality follows from conditional independence of the terms k and $l \neq k$. Replacing the inner $\exp(x)$ by its Taylor expansion $\exp(x) = 1 + x + \frac{1}{2}x^2 + R(x)$ yields the

expression

$$
\begin{aligned}
& E\exp\left(\left[C_1\sqrt{J}\left[\psi_N - E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right] + C_2\sqrt{J}\left(\mu_{(X,\varepsilon)_N} - \mu_{X,\varepsilon}\right)\left(\gamma_{X,\varepsilon}\right)\right.\right. \\
& \left.\left. + C_3\sqrt{J}\left(\mu_{(Z,\eta)_N} - \mu_{Z,\eta}\right)\left(\gamma_{Z,\eta}\right)\right]\right) \\
= \; & E\exp\sum_{k=1}^{J}\log E\left[1 + \frac{1}{\sqrt{J}}\left(C_1\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\right.\right. \\
& + C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right) - E\left(\gamma_{X,\varepsilon}\right)\right] \\
& \left.\left. + C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right) - E\left(\gamma_{Z,\eta}\right)\right]\right) \\
& + \frac{1}{2J}\left(C_1\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\right. \\
& + C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right) - E\left(\gamma_{X,\varepsilon}\right)\right] \\
& \left.\left. + C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right) - E\left(\gamma_{Z,\eta}\right)\right]\right)^2 + \frac{R_k}{J^{\frac{3}{2}}}\middle|\,\mu_{V_N},\mu_{U_N}\right] \\
= \; & E\exp\sum_{k=1}^{J}\log E\left[1+\right. \\
& + \frac{C_2}{\sqrt{J}}\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)}\right) - E\left(\gamma_{X\varepsilon}\right)\right] \\
& + \frac{C_3}{\sqrt{J}}\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right) - E\left(\gamma_{Z,\eta}\right)\right] \\
& + \frac{1}{2J}\left(C_1\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\right. \\
& + C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right) + \frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},,\varepsilon^{(2k)}\right) - E\left(\gamma_{X,\varepsilon}\right)\right] \\
& \left.\left. + C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right) - E\left(\gamma_{Z,\eta}\right)\right]\right)^2 + \frac{R_k}{J^{\frac{3}{2}}}\middle|\,\mu_{V_N},\mu_{U_N}\right]
\end{aligned}
$$

where the first term $E\left[\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\middle|\,\mu_{V_N},\mu_{U_N}\right] = 0$ by the definition of $\tilde{\psi}$.

Since γ_X, γ_Z and Ψ are bounded, we approximate $\log(1+x)$ by its Taylor expansion $\log(1+x) = x - \frac{1}{2}x^2 + r(x)$ and keep track only of terms J^{-1} and lower (note that R_k is

bounded as well). The above equation simplifies to

$$
\begin{aligned}
&E\exp\sum_{k=1}^{J}E\left\{\frac{1}{\sqrt{J}}C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right.\\
&+\frac{1}{\sqrt{J}}C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\\
&+\frac{1}{2J}\left[C_1\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\right.\\
&+C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\\
&+C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\Big]^2\\
&-\frac{1}{2J}\left(C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right)^2\\
&-\frac{1}{2J}\left(C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\right)^2\\
&-\frac{1}{2J}2C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]*\\
&C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\Big|\,\mu_{V_N},\mu_{U_N}\Big\}\\
&+o\left(J^{-1}\right)
\end{aligned}
$$

Since
$\frac{1}{J}\sum_{k=1}^{J}\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]$
converges in probability to 0, we can rewrite this as

$$
\begin{aligned}
&E\exp\sum_{k=1}^{J}E\left\{\frac{1}{\sqrt{J}}C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right.\\
&+\frac{1}{\sqrt{J}}C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\\
&+\frac{1}{2J}\left[C_1\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\right.\\
&+C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\\
&+C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\Big]^2\\
&-\frac{1}{2J}\left(C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right)^2\\
&-\frac{1}{2J}\left(C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\right)^2\Big|\,\mu_{V_N},\mu_{U_N}\Big\}+o\left(1\right)
\end{aligned}
$$

By the variance computations in equations (C.17), (C.19) and (C.20),

$$\sum_{k=1}^{J}\frac{1}{2J}\left[C_1\left(\Psi\left(X^{(2k-1)},X^{(2k)},Z^{(k)}\right)-\tilde{\psi}\left(v^{(2k-1)},v^{(2k)},u^{(k)}\right)\right)\right.$$
$$+C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]$$
$$\left.+C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\right]^2$$

converges in probability to

$$\begin{aligned}V_1 &= \frac{C_1^2}{2}\int_0^1 var_q\left(q,q,q\right)dq+C_1C_2\int_0^1 cov_q\left(\Psi,f|q,q,q\right)dq\\ &\quad+C_1C_3\int_0^1 cov_q\left(\Psi,g|q,q,q\right)dq\\ &\quad+\frac{C_2^2}{2}\frac{1}{2}Var\left(\gamma_{X,\varepsilon}\right)+\frac{C_3^2}{2}Var\left(\gamma_{Z,\eta}\right).\end{aligned}$$

Therefore,

$$E\exp\left(\left[C_1\sqrt{J}\left[\psi_N-E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right]+C_2\sqrt{J}\left(\mu_{\left(X,\varepsilon\right)_N}-\mu_{X,\varepsilon}\right)\left(\gamma_{X,\varepsilon}\right)\right.\right.$$
$$\left.\left.+C_3\sqrt{J}\left(\mu_{\left(Z,\eta\right)_N}-\mu_{\left(Z,\eta\right)}\right)\left(\gamma_{Z,\eta}\right)\right]\right)=$$
$$\exp\left(V_1\right)E\exp\left\{\frac{1}{\sqrt{J}}\sum_{k=1}^{J}E\left[C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right.\right.$$
$$\left.+C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\Big|\,\mu_{V_N},\mu_{U_N}\right]$$
$$-\frac{1}{2J}\sum_{k=1}^{J}E\left[C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right.$$
$$\left.\left.+C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\Big|\,\mu_{V_N},\mu_{U_N}\right]^2\right\}+o\left(1\right).$$

Since convergence in distribution implies convergence of moment generating functions and

$$\frac{1}{2J}\sum_{k=1}^{J}E\left[C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right.$$
$$\left.+C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\Big|\,\mu_{V_N},\mu_{U_N}\right]^2$$

converges in probability to

$$\frac{1}{2}C_2^2\frac{1}{2}Var\left(\gamma_{X,\varepsilon}\right)+\frac{1}{2}C_3^2Var\left(\gamma_{Z,\eta}\right),$$

we can rewrite,

$$\begin{aligned}
&E\exp\left(\left[C_1\sqrt{J}\left[\psi_N-E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right]\right.\right.\\
&\left.\left.+C_2\sqrt{J}\left(\mu_{(X,\varepsilon)_N}-\mu_{X,\varepsilon}\right)\left(\gamma_{X,\varepsilon}\right)+C_3\sqrt{J}\left(\mu_{(Z,\eta)_N}-\mu_{Z,\eta}\right)\left(\gamma_{Z,\eta}\right)\right]\right)=\exp\left(V_1\right)*\\
&E\exp\left(\sum_{k=1}^{J}E\left[\frac{1}{\sqrt{J}}\left(C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right.\right.\right.\\
&\left.\left.\left.+C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\right)\middle|\mu_{V_N},\mu_{U_N}\right]-\frac{C_2^2}{2}Var\left(\gamma_{X,\varepsilon}\right)-\frac{C_3^2}{2}Var\left(\gamma_{Z,\eta}\right)\Bigg)\\
&+o\left(1\right)\\
=\;&\exp\left(V_1\right)\exp\left(-\frac{1}{2}C_2^2\frac{1}{2}Var\left(\gamma_{X,\varepsilon}\right)-\frac{1}{2}C_3^2Var\left(\gamma_{Z,\eta}\right)\right)\times\\
&E\exp\sum_{k=1}^{J}E\left[\frac{1}{\sqrt{J}}\left(C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right.\right.\\
&\left.\left.+C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\right)\middle|\mu_{V_N},\mu_{U_N}\right]+o\left(1\right).
\end{aligned}$$

By the Levy continuity theorem and the equality
$E\exp\left(tX\right)=\exp\left(E\left[X\right]'t+\frac{1}{2}t'V\left(X\right)^{-1}t\right)$ for normally distributed random variables, the product of the second and third terms,

$$\begin{aligned}
&\exp\left(-\frac{1}{2}C_2^2\frac{1}{2}Var\left(\gamma_{X,\varepsilon}\right)-\frac{1}{2}C_3^2Var\left(\gamma_{Z,\eta}\right)\right)\times\\
&E\exp\sum_{k=1}^{J}E\left[\frac{1}{\sqrt{J}}\left(C_2\left[\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k-1)},\varepsilon^{(2k-1)}\right)+\frac{1}{2}\gamma_{X,\varepsilon}\left(X^{(2k)},\varepsilon^{(2k)}\right)-E\left(\gamma_{X,\varepsilon}\right)\right]\right.\right.\\
&\left.\left.+C_3\left[\gamma_{Z,\eta}\left(Z^{(k)},\eta^{(k)}\right)-E\gamma_{Z,\eta}\right]\right)\middle|\mu_{V_N},\mu_{U_N}\right]
\end{aligned}$$

converges to 1. Hence,

$$\begin{aligned}
&E\exp\left(\left[C_1\sqrt{J}\left[\psi_N-E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right]+C_2\sqrt{J}\left(\mu_{(X,\varepsilon)_N}-\mu_{X,\varepsilon}\right)\left(\gamma_{X,\varepsilon}\right)\right.\right.\\
&\left.\left.+C_3\sqrt{J}\left(\mu_{(Z,\eta)_N}-\mu_{Z,\eta}\right)\left(\gamma_{Z,\eta}\right)\right]\right)
\end{aligned}$$

converges in probability to $\exp(V_1)$. Therefore, by Levy continuity,

$$\begin{bmatrix} \sqrt{J}\left(\psi_N - E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right) \\ \sqrt{J}\left(\mu_{(X,\varepsilon)_N} - \mu_{X,\eta}\right)(\gamma_{X,\varepsilon}) \\ \sqrt{J}\left(\mu_{(Z,\eta)_N} - \mu_{Z,\eta}\right)(\gamma_{Z,\eta}) \end{bmatrix}$$

converges in distribution to a mean-zero normal with covariance

$$V''(\gamma_X,\gamma_Z) = \begin{bmatrix} \int_0^1 \sigma^2_{q,\Psi}(q,q,q)\,dq & \int_0^1 cov_q\left(\Psi,\gamma_{X,\varepsilon}|q,q,q\right)dq & \int_0^1 cov_q\left(\Psi,\gamma_{Z,\eta}|q,q,q\right)dq \\ \int_0^1 cov_q\left(\Psi,\gamma_{X,\varepsilon}|q,q,q\right)dq & \frac{1}{2}Var\left(\gamma_{X,\varepsilon}\right) & 0 \\ \int_0^1 cov_q\left(\Psi,\gamma_{Z,\eta}|q,q,q\right)dq & 0 & Var\left(\gamma_{Z,\eta}\right) \end{bmatrix}.$$

■

Lemma C.10 *Suppose Assumption 3.4(ii) is satisfied. For any $\mu_{X,\varepsilon}-$ Donsker class $\Gamma_{X,\varepsilon}$ of bounded functions on (X,ε) and $\mu_{Z,\eta}$ Donsker class $\Gamma_{Z,\eta}$ of bounded functions on (Z,η),*

$$\begin{bmatrix} \sqrt{N/2}\left(\psi_N - E\left(\psi_N|\mu_{U_N},\mu_{V_N}\right)\right) \\ \sqrt{N/2}\left(\mu_{(X,\varepsilon)_N} - \mu_{X,\varepsilon}\right)(\gamma_{X,\varepsilon}) \\ \sqrt{N/2}\left(\mu_{(Z,\eta)_N} - \mu_{Z,\eta}\right)(\gamma_{Z,\eta}) \end{bmatrix}$$

indexed by $\gamma_{X,\varepsilon} \in \Gamma_{X,\varepsilon}$ and $\gamma_{Z,\eta} \in \Gamma_{Z,\eta}$ converges to a Gaussian process whose covariance kernel characterized by V''.

Proof. Let $\gamma_{X,\varepsilon}$ be a linear combination of a finite number of elements of $\Gamma_{X,\varepsilon}$ and $\gamma_{Z,\eta}$ be a linear combination of a finite number of elements of $\Gamma_{Z,\eta}$. By Lemma C.9,

$$\sqrt{N/2}\left(\psi_N - E\left(\psi_N|\mu_{U_N},\mu_{V_N}\right),\left(\mu_{(X,\varepsilon)_N} - \mu_{X,\varepsilon}\right)(\gamma_{X,\varepsilon}),\left(\mu_{(Z,\eta)_N} - \mu_{Z,\eta}\right)(\gamma_{Z,\eta})\right)$$

converges in distribution to $N\left(0,V''\left(\gamma_{X,\varepsilon},\gamma_{Z,\eta}\right)\right)$. Let H_N be the stochastic process $\sqrt{N/2}\left(\psi_N - E\left(\psi_N|\mu_{U_N},\mu_{V_N}\right)\right)$ jointly with the empirical processes on $\Gamma_{X,\varepsilon}$ and $\Gamma_{Z,\eta}$. We index these stochastic processes with $\gamma \in \Gamma$ where we endow Γ with the L_2 metric. By the Cramer-Wold device, the finite dimensional distributions of H_N converge to a Gaussian process whose covariance kernel is defined as follows:

For two elements of $\Gamma_{X,\varepsilon}$ and $\Gamma_{Z,\eta}$, the covariance kernel is that of the associated empirical processes. The covariance of an element of $\Gamma_{X,\varepsilon}$ and an element of $\Gamma_{Z,\eta}$ is 0. The covariance of $\gamma_{X,\varepsilon} \in \Gamma_{X,\varepsilon}$ with $\sqrt{N/2}\left(\psi_N - E\left(\psi_N|\mu_{U_N}, \mu_{V_N}\right)\right)$ is $\int_0^1 cov_q\left(\Psi, \gamma_{X,\varepsilon}|q,q,q\right) dq$. The covariance of $\gamma_{Z,\eta} \in \Gamma_{Z,\eta}$ with the term $\sqrt{N}\left(\psi_N - E\left(\psi_N|\mu_{U_N}, \mu_{V_N}\right)\right)$ is $\int_0^1 cov_q\left(\Psi, \gamma_{Z,\eta}|q,q,q\right) dq$. The variance of $\sqrt{N/2}\left(\psi_N - E\left(\psi_N|\mu_{U_N}, \mu_{V_N}\right)\right)$ is $\int_0^1 \sigma^2_{q,\Psi}\left(q,q,q\right) dq$.

We now verify equicontinuity to show weak convergence of H_N. We prove this directly using equicontinuity properties of the empirical processes on $\Gamma_{X,\varepsilon}$ and $\Gamma_{Z,\varepsilon}$. Denote

$$\begin{aligned} Var_{V,Z}\left(m_V, m_Z\right) &= \int Var\left(\Psi\left(X_1, X_2, Z\right)|v, Z = z\right) dm_Z dm_V \\ Var_{U,X}\left(m_U, m_X\right) &= \int Var\left(\Psi\left(X_1, X_2, Z\right)|u, X_1 = x_1, X_2 = x_2\right) dm_X dm_X dm_U. \end{aligned}$$

Let $Var\left(\Psi\left(X_1, X_2, Z\right)|v_1, v_2, Z = z\right) = Var\left(v_1, v_2, z\right)$. Consider the quantity $\frac{1}{N/2}\sum_{i=1}^{N/2} Var\left(v^{(2i-1)}, v^{(2i)}, z^{(i)}\right)$, and note that since Var is bounded and uniformly continuous, it is equal to

$$\begin{aligned} & \frac{1}{N}\sum_{i=1}^{N/2}\left[Var\left(v^{(2i-1)}, v^{(2i-1)}, z^{(i)}\right) + Var\left(v^{(2i)}, v^{(2i)}, z^{(i)}\right)\right] + o\left(1\right) \\ = \; & Var_{V,Z}\left(\mu_{V_N}, \mu_{Z_N}\right) + o\left(1\right) \\ = \; & Var_{V,Z}\left(\mu_V, \mu_Z\right) + o\left(1\right) \end{aligned}$$

An identical argument implies

$$\frac{1}{N/2}\sum_{i=1}^{N/2} Var\left(x^{(2i-1)}, x^{(2i)}, u^{(i)}\right) = Var_{U,X}\left(\mu_U, \mu_X\right) + o\left(1\right).$$

Since $\mu_{X|v}$ and $\mu_{Z|u}$ are not degenerate, $Var_{V,Z}\left(\mu_V, \mu_Z\right)$ and $Var_{U,X}\left(\mu_U, \mu_X\right)$ are strictly positive. Hence, $\lim_N \sup_{\mu_{V_N}, \mu_{Z_N}} Var_{V,Z}\left(\mu_{V_N}, \mu_{Z_N}\right)$ and $\lim_N \sup_{\mu_{U_N}, \mu_{X_N}} Var_{U,X}\left(\mu_{U_N}, \mu_{X_N}\right)$ are strictly positive. Hence, for large enough N, there is a $\delta > 0$ such that a $\delta-$ball around $H_N\left(\gamma\right) = \sqrt{N/2}\left(\psi_N - E\left(\psi_N|\mu_{V_N}, \mu_{U_N}\right)\right)$ contains no other element $H_N\left(\gamma'\right)$ for $\gamma' \neq \gamma$. Pick $\delta > 0$ such that the δ ball around $\sqrt{N/2}\left(\psi_N - E\left(\psi_N|\mu_{V_N}, \mu_{U_N}\right)\right)$ is a singleton. Hence,

if $B_{\Gamma_{X,\varepsilon}}(\gamma,\delta) = B(\gamma,\delta) \cap \Gamma_{X,\varepsilon}$, and $B_{\Gamma_{Z,\eta}}(\gamma,\delta) = B(\gamma,\delta) \cap \Gamma_{Z,\eta}$,

$$\begin{aligned}
&\sup_{\gamma\in\Gamma} \sup_{\gamma'\in B(\gamma,\delta)} \left|H_N(\gamma) - H_N(\gamma')\right| \\
\leq\ &\sup_{\gamma_{X,\varepsilon}\in\Gamma_{X,\varepsilon}} \sup_{\gamma'_{X,\varepsilon}\in B_{\Gamma_{X,\varepsilon}}(\gamma_{X,\varepsilon},\delta)} \left|H_N(\gamma_{X,\varepsilon}) - H_N(\gamma'_{X,\varepsilon})\right| + \\
&\sup_{\gamma_{Z,\eta}\in\Gamma_{Z,\eta}} \sup_{\gamma'_{X,\varepsilon}\in B_{\Gamma_{X,\varepsilon}}\left(\gamma_{Z,\eta},\delta\right)} \left|H_N(\gamma_{Z,\eta}) - H_N(\gamma'_{X,\varepsilon})\right| \\
&+ \sup_{\gamma_{X,\varepsilon}\in\Gamma_{X,\varepsilon}} \sup_{\gamma'_{Z,\eta}\in B_{\Gamma_{Z,\eta}}(\gamma_{X,\varepsilon},\delta)} \left|H_N(\gamma_{X,\varepsilon}) - H_N\left(\gamma'_{Z,\eta}\right)\right| + \\
&\sup_{\gamma_{Z,\eta}\in\Gamma_{Z,\eta}} \sup_{\gamma'_{Z,\eta}\in B_{\Gamma_{Z,\eta}}\left(\gamma_{Z,\eta},\delta\right)} \left|H_N(\gamma_{Z,\eta}) - H_N\left(\gamma'_{Z,\eta}\right)\right|
\end{aligned}$$

since $B(\gamma,\delta) = \{\gamma\}$ if $H_N(\gamma) = \sqrt{N/2}\,(\psi_N - E(\psi_N|\mu_{U_N},\mu_{V_N}))$.

For a fixed $\varepsilon,\eta > 0$, there is (by definition of stochastic equicontinuity) there exists $\delta > 0$ such that

$$\limsup_{N\to\infty} P\left(\sup_{\gamma_{X,\varepsilon}\in\Gamma_{X,\varepsilon}} \sup_{\gamma'_{X,\varepsilon}\in B_{\Gamma_{X,\varepsilon}}(\gamma_{X,\varepsilon},\delta)} \left|H_N(\gamma_{X,\varepsilon}) - H_N(\gamma'_{X,\varepsilon})\right| > \frac{\varepsilon}{6}\right) < \frac{\eta}{6}$$

and

$$\limsup_{N\to\infty} P\left(\sup_{\gamma_{Z,\eta}\in\Gamma_{Z,\eta}} \sup_{\gamma'_{Z,\eta}\in B_{\Gamma_{Z,\eta}}\left(\gamma_{Z,\eta},\delta\right)} \left|H_N(\gamma_{Z,\eta}) - H_N\left(\gamma'_{Z,\eta}\right)\right| > \frac{\varepsilon}{6}\right) < \frac{\eta}{6}.$$

Now we show that

$$\limsup_{N\to\infty} P\left(\sup_{\gamma_{X,\varepsilon}\in\Gamma_{X,\varepsilon}} \sup_{\gamma'_{Z,\eta}\in B_{\Gamma_{Z,\eta}}(\gamma_{X,\varepsilon},\delta)} \left|H_N(\gamma_{X,\varepsilon}) - H_N\left(\gamma'_{Z,\eta}\right)\right| > \frac{\varepsilon}{3}\right) < \frac{\eta}{6}.$$

Note that independence of empirical processes on $\Gamma_{X,\varepsilon}$ and $\Gamma_{Z,\eta}$ implies that $B_{\Gamma_{Z,\eta}}(\gamma_{X,\varepsilon},\delta)$ is nonempty only if $\gamma_{X,\varepsilon}$ has L^2 norm less than δ. If this is the case, every element of

$B_{\Gamma_{Z,\eta}}(\gamma_{X,\varepsilon}, \delta)$ also has L^2 norm less than δ. Therefore,

$$
\begin{aligned}
& P\left(\sup_{\gamma_{X,\varepsilon}\in\Gamma_{X,\varepsilon}} \sup_{\gamma'_{Z,\eta}\in B_{\Gamma_{Z,\eta}}(\gamma_{X,\varepsilon},\delta)} \left|H_N(\gamma_{X,\varepsilon}) - H_N\left(\gamma'_{Z,\eta}\right)\right| > \frac{\varepsilon}{3}\right) \\
\leq\ & P\left(\sup_{\gamma_{X,\varepsilon}\in\Gamma_{X,\varepsilon}} \sup_{\gamma'_{Z,\eta}\in B_{\Gamma_{Z,\eta}}(\gamma_{X,\varepsilon},\delta)} |H_N(\gamma_{X,\varepsilon})| + \left|H_N\left(\gamma'_{Z,\eta}\right)\right| > \frac{\varepsilon}{3}\right) \\
\leq\ & P\left(\sup_{\gamma_{Z,\eta}\in\Gamma_{Z,\eta}} \sup_{\gamma'_{Z,\eta}\in B_{\Gamma_{Z,\eta}}(\gamma_{Z,\eta},\delta)} \left|H_N(\gamma_{Z,\eta}) - H_N\left(\gamma'_{Z,\eta}\right)\right| + \right. \\
& \left. \sup_{\gamma_{X,\varepsilon}\in\Gamma_{X,\varepsilon}} \sup_{\gamma'_{X,\varepsilon}\in B_{\Gamma_{X,\varepsilon}}(\gamma_{X,\varepsilon},\delta)} \left|H_N(\gamma_{X,\varepsilon}) - H_N(\gamma'_{X,\varepsilon})\right| > \frac{\varepsilon}{3}\right) \\
\leq\ & P\left(\sup_{\gamma_{Z,\eta}\in\Gamma_{Z,\eta}} \sup_{\gamma'_{Z,\eta}\in B_{\Gamma_{Z,\eta}}(\gamma_{Z,\eta},\delta)} \left|H_N(\gamma_{Z,\eta}) - H_N\left(\gamma'_{Z,\eta}\right)\right| > \frac{\varepsilon}{6}\right) \\
& + P\left(\sup_{\gamma_{X,\varepsilon}\in\Gamma_{X,\varepsilon}} \sup_{\gamma'_{X,\varepsilon}\in B_{\Gamma_{X,\varepsilon}}(\gamma_{X,\varepsilon},\delta)} \left|H_N(\gamma_{X,\varepsilon}) - H_N(\gamma'_{X,\varepsilon})\right| > \frac{\varepsilon}{6}\right)
\end{aligned}
$$

where the second inequality follows from the triangle inequality since a constant 0 function is an element of both $\Gamma_{X,\varepsilon}$ and $\Gamma_{Z,\eta}$. By the same argument

$$
\begin{aligned}
& P\left(\sup_{\gamma_{Z,\eta}\in\Gamma_{Z,\eta}} \sup_{\gamma'_{X,\varepsilon}\in B_{\Gamma_{X,\varepsilon}}(\gamma_{X,\varepsilon},\delta)} \left|H_N(\gamma_{Z,\eta}) - H_N(\gamma'_{X,\varepsilon})\right| > \frac{\varepsilon}{3}\right) \\
\leq\ & P\left(\sup_{\gamma_{Z,\eta}\in\Gamma_{Z,\eta}} \sup_{\gamma'_{Z,\eta}\in B_{\Gamma_{Z,\eta}}(\gamma_{Z,\eta},\delta)} \left|H_N(\gamma_{Z,\eta}) - H_N\left(\gamma'_{Z,\eta}\right)\right| > \frac{\varepsilon}{6}\right) \\
& + P\left(\sup_{\gamma_{X,\varepsilon}\in\Gamma_{X,\varepsilon}} \sup_{\gamma'_{X,\varepsilon}\in B_{\Gamma_{X,\varepsilon}}(\gamma_{X,\varepsilon},\delta)} \left|H_N(\gamma_{X,\varepsilon}) - H_N(\gamma'_{X,\varepsilon})\right| > \frac{\varepsilon}{6}\right)
\end{aligned}
$$

and thus

$$\begin{aligned}
&\limsup_{n\to\infty} P\left(\sup_{\gamma\in\Gamma}\sup_{\gamma'\in B(\gamma,\delta)}\left|H_N(\gamma)-H_N(\gamma')\right|>\varepsilon\right)\\
\leq\ & 3\limsup_{n\to\infty} P\left(\sup_{\gamma_{X,\varepsilon}\in\Gamma_{X,\varepsilon}}\sup_{\gamma'_{X,\varepsilon}\in B_{\Gamma_{X,\varepsilon}}(\gamma_{X,\varepsilon},\delta)}\left|H_N(\gamma_{X,\varepsilon})-H_N(\gamma'_{X,\varepsilon})\right|>\frac{\varepsilon}{6}\right)\\
&+3\limsup_{n\to\infty} P\left(\sup_{\gamma_{Z,\eta}\in\Gamma_{Z,\eta}}\sup_{\gamma'_{Z,\eta}\in B_{\Gamma_{Z,\eta}}(\gamma_{Z,\eta},\delta)}\left|H_N(\gamma_{Z,\eta})-H_N\left(\gamma'_{Z,\eta}\right)\right|>\frac{\varepsilon}{6}\right)\\
<\ & \eta.
\end{aligned}$$

This proves stochastic equicontinuity of $H_N(\gamma)$ and hence weak convergence to the Gaussian process defined above. ■

C.2.2 Proof of Proposition 3.4

We will show that the Hadamard derivative of $\psi^\delta : L_\infty^\Gamma \to L_\infty^\Theta$ evaluated at (μ_X, μ_Z) in the direction (G_X, G_Z) is

$$\begin{aligned}
&\nabla_{(G_X,G_Z)}\tilde{\psi}^\delta[\mu_X,\mu_Z](\theta)=\\
&\int_\delta^{1-\delta} G_U^q(\theta)\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,f'_\eta\left(F^{-1}_{U;\theta}(q)-h(z;\theta)\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}\\
&+G_V^q(\theta)\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,f'_\varepsilon\left(F^{-1}_{V;\theta}(q)-h(x_2;\theta)\right)\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}\\
&+G_V^q(\theta)\frac{\int\Psi(x_1,x_2,z)\,f'_\varepsilon\left(F^{-1}_{V;\theta}(q)-h(x_1;\theta)\right)\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq\\
&+\int_\delta^{1-\delta}\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,dG_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq\\
&+\int_\delta^{1-\delta}\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_X dG_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq\\
&+\int_\delta^{1-\delta}\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}dG_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq\\
&+\int_\delta^{1-\delta}\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}L_G[\mu_X,\mu_Z](\theta,q)\,dq.
\end{aligned}$$

where

$$\begin{aligned} G_V^q(\theta) &= \frac{1}{f_{V;\theta}\left(F_{V,\theta}^{-1}(q)\right)} \int G_X\left(1\left\{h(x;\theta)+\varepsilon \leq F_{V;\theta}^{-1}(q)\right\}\right) dF_\varepsilon, \\ G_U^q(\theta) &= \frac{1}{f_{U;\theta}\left(F_{U,\theta}^{-1}(q)\right)} \int G_Z\left(1\left\{g(z;\theta)+\eta \leq F_{U;\theta}^{-1}(q)\right\}\right) dF_\eta, \end{aligned}$$

and $L_G[\mu_X, \mu_Z](\theta, q)$ is the negative of

$$\begin{aligned} & G_U^q(\theta) \frac{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,f_\eta'\left(F_{U;\theta}^{-1}(q)-g(z;\theta)\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\ & + G_V^q(\theta) \frac{\int \phi_\varepsilon(q,x_1;\theta)\,f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x_2;\theta)\right)\phi_\eta(q,z;\theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\ & + G_V^q(\theta) \frac{\int f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x_1;\theta)\right)\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\ & + \frac{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\, dG_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\ & + \frac{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\, d\mu_X dG_{X_2} d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\ & + \frac{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\, d\mu_{X_1} d\mu_{X_2} dG_{X_3}}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} \end{aligned}$$

and

$$\begin{aligned} \phi_\eta(q,z;\theta) &= f_\eta\left(F_{u;\theta,\mu_Z}^{-1}(q)-g(z;\theta)\right) \\ \phi_\varepsilon(q,x;\theta) &= f_\varepsilon\left(F_{v;\theta,\mu_X}^{-1}(q)-h(x;\theta)\right). \end{aligned}$$

Proof. Let

$$\begin{aligned} \phi_{\eta,N}(q,z;\theta) &= f_\eta\left(F_{N,U;\theta,\mu_{Z_N}}^{-1}(q)-g(z;\theta)\right) \\ \phi_{\varepsilon,N}(q,x;\theta) &= f_\varepsilon\left(F_{N,V;\theta,\mu_{X_N}}^{-1}(q)-h(x;\theta)\right) \end{aligned}$$

where $F_{N,U;\theta}(u) = \int F_\eta(u-g(Z;\theta))\, d\mu_{Z_N}$ and $F_{N,V;\theta}(v) = \int F_\varepsilon(v-h(X;\theta))\, d\mu_{X_N}$.

Consider a sequence of measures (μ_{X_N}, μ_{Z_N}) and a sequence of scalars $h_N \to 0$ such that $\frac{1}{h_N}(\mu_{X_N}-\mu_X, \mu_{Z_N}-\mu_Z)$ converges to $G = (G_X, G_Z)$ uniformly in L_∞^Γ, where G is bounded

and uniformly continuous. We can rewrite

$$\int_{\delta}^{1-\delta} \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq$$
$$- \int_{\delta}^{1-\delta} \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon,N}\left(q, x_1; \theta\right) \phi_{\varepsilon,N}\left(q, x_2; \theta\right) \phi_{\eta,N}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon,N}\left(q, x_1; \theta\right) \phi_{\varepsilon,N}\left(q, x_2; \theta\right) \phi_{\eta,N}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}} dq =$$
$$\int_{\delta}^{1-\delta} \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) \left(d\mu_{X_1} d\mu_{X_2} d\mu_Z - d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}\right)}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq$$
$$+ \int_{\delta}^{1-\delta} \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq$$
$$- \int_{\delta}^{1-\delta} \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon,N}\left(q, x_1; \theta\right) \phi_{\varepsilon,N}\left(q, x_2; \theta\right) \phi_{\eta,N}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \times$$
$$\frac{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon,N}\left(q, x_1; \theta\right) \phi_{\varepsilon,N}\left(q, x_2; \theta\right) \phi_{\eta,N}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}} dq =$$
$$\int_{\delta}^{1-\delta} \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) \left(d\mu_{X_1} d\mu_{X_2} d\mu_Z - d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}\right)}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq$$
$$+ \left[\int_{\delta}^{1-\delta} \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \right.$$
$$\left. - \int_{\delta}^{1-\delta} \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon,N}\left(q, x_1; \theta\right) \phi_{\varepsilon,N}\left(q, x_2; \theta\right) \phi_{\eta,N}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \right]$$
$$+ \int_{\delta}^{1-\delta} \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon,N}\left(q, x_1; \theta\right) \phi_{\varepsilon,N}\left(q, x_2; \theta\right) \phi_{\eta,N}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \times$$
$$\left(1 - \frac{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon,N}\left(q, x_1; \theta\right) \phi_{\varepsilon,N}\left(q, x_2; \theta\right) \phi_{\eta,N}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}\right) dq =$$
$$T_1 + T_2 + T_3 =$$
$$\int_{\delta}^{1-\delta} T_1\left(q\right) dq + \int_{\delta}^{1-\delta} T_2\left(q\right) dq + \int_{\delta}^{1-\delta} T_3\left(q\right) dq \tag{C.21}$$

To compute $T_1(q)$ note that

$$\begin{aligned}
& d\mu_{X_1} d\mu_{X_2} d\mu_Z - d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N} \\
= & \left(d\mu_{X_1} - d\mu_{X_{N,1}}\right) d\mu_{X_2} d\mu_Z + d\mu_{X_{N,1}} d\mu_{X_2} d\mu_Z - d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N} \\
= & \left(d\mu_{X_1} - d\mu_{X_{N,1}}\right) d\mu_{X_2} d\mu_Z + d\mu_{X_{N,1}} \left(d\mu_{X_2} - d\mu_{X_{N,2}}\right) d\mu_Z \\
& + d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_Z - d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N} \\
= & \left(d\mu_{X_1} - d\mu_{X_{N,1}}\right) d\mu_{X_2} d\mu_Z + d\mu_{X_{N,1}} \left(d\mu_{X_2} - d\mu_{X_{N,2}}\right) d\mu_Z + d\mu_{X_{N,1}} d\mu_{X_{N,2}} \left(d\mu_Z - d\mu_{Z_N}\right) \\
= & \left(d\mu_{X_1} - d\mu_{X_{N,1}}\right) d\mu_{X_2} d\mu_Z + d\mu_X \left(d\mu_{X_2} - d\mu_{X_{N,2}}\right) d\mu_Z \\
& + \left(d\mu_{X_{N,1}} - d\mu_X\right) \left(d\mu_{X_2} - d\mu_{X_{N,2}}\right) d\mu_Z + d\mu_{X_1} d\mu_{X_2} \left(d\mu_Z - d\mu_{Z_N}\right) \\
& + \left(d\mu_{X_{N,1}} - d\mu_{X_1}\right) d\mu_{X_2} \left(d\mu_Z - d\mu_{Z_N}\right) + d\mu_{X_{N,1}} \left(d\mu_{X_{N,2}} - d\mu_{X_2}\right) \left(d\mu_Z - d\mu_{Z_N}\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
T_1(q) = & \frac{\int \Psi(x_1, x_2, z) \phi_\varepsilon(q, x_1; \theta) \phi_\varepsilon(q, x_2; \theta) \phi_\eta(q, z; \theta) \left(d\mu_{X_1} - d\mu_{X_{N,1}}\right) d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q, x_1; \theta) \phi_\varepsilon(q, x_2; \theta) \phi_\eta(q, z; \theta) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\
& + \frac{\int \Psi(x_1, x_2, z) \phi_\varepsilon(q, x_1; \theta) \phi_\varepsilon(q, x_2; \theta) \phi_\eta(q, z; \theta) d\mu_X \left(d\mu_{X_2} - d\mu_{X_{N,2}}\right) d\mu_Z}{\int \phi_\varepsilon(q, x_1; \theta) \phi_\varepsilon(q, x_2; \theta) \phi_\eta(q, z; \theta) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\
& + \frac{\int \Psi(x_1, x_2, z) \phi_\varepsilon(q, x_1; \theta) \phi_\varepsilon(q, x_2; \theta) \phi_\eta(q, z; \theta) d\mu_{X_1} d\mu_{X_2} \left(d\mu_Z - d\mu_{Z_N}\right)}{\int \phi_\varepsilon(q, x_1; \theta) \phi_\varepsilon(q, x_2; \theta) \phi_\eta(q, z; \theta) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\
& + \frac{R(q)}{\int \phi_\varepsilon(q, x_1; \theta) \phi_\varepsilon(q, x_2; \theta) \phi_\eta(q, z; \theta) d\mu_{X_1} d\mu_{X_2} d\mu_Z}
\end{aligned} \tag{C.22}$$

where $R(q)$ is equal to

$$\begin{aligned}
& \int \Psi(x_1, x_2, z) \phi_\varepsilon(q, x_1; \theta) \phi_\varepsilon(q, x_2; \theta) \phi_\eta(q, z; \theta) \left(d\mu_{X_{N,1}} - d\mu_X\right) \left(d\mu_{X_2} - d\mu_{X_{N,2}}\right) d\mu_Z \\
& + \int \Psi(x_1, x_2, z) \phi_\varepsilon(q, x_1; \theta) \phi_\varepsilon(q, x_2; \theta) \phi_\eta(q, z; \theta) \left(d\mu_{X_{N,1}} - d\mu_{X_1}\right) d\mu_{X_2} \left(d\mu_Z - d\mu_{Z_N}\right) \\
& + \int \Psi(x_1, x_2, z) \phi_\varepsilon(q, x_1; \theta) \phi_\varepsilon(q, x_2; \theta) \phi_\eta(q, z; \theta) d\mu_{X_{N,1}} \left(d\mu_{X_{N,2}} - d\mu_{X_2}\right) \left(d\mu_Z - d\mu_{Z_N}\right) \\
& = R_1(q) + R_2(q) + R_3(q).
\end{aligned}$$

Now we show that each of $\frac{1}{h_N} R_1$, $\frac{1}{h_N} R_2$ and $\frac{1}{h_N} R_3$ are negligible. To show that $\frac{1}{h_N} R_1(q)$ is

negligible, we rewrite it as

$$\begin{aligned}
&\frac{1}{h_N} R_1(q) \\
= &\int \Psi(x_1, x_2, z)\, \phi_\varepsilon(q, x_1; \theta)\, \phi_\varepsilon(q, x_2; \theta)\, \phi_\eta(q, z; \theta)\, dG_{X_1} \left(d\mu_{X_2} - d\mu_{X_{N,2}}\right) d\mu_Z + \\
&\int \Psi(x_1, x_2, z)\, \phi_\varepsilon(q, x_1; \theta)\, \phi_\varepsilon(q, x_2; \theta)\, \phi_\eta(q, z; \theta) \\
&\left(\frac{1}{h_N}\left(d\mu_{X_{N,1}} - d\mu_X\right) - dG_{X_1}\right)\left(d\mu_{X_2} - d\mu_{X_{N,2}}\right) d\mu_Z \\
= &\, S_1(q) + S_2(q)
\end{aligned}$$

and show that S_1 and S_2 are negligible. Note that

$$\begin{aligned}
&\sup_{q\in(\delta,1-\delta),\theta} |S_2(q)| \leq \int \sup_{q\in(\delta,1-\delta),\theta} \\
&\left|\Psi(x_1, x_2, z)\, \phi_\varepsilon(q, x_1; \theta)\, \phi_\varepsilon(q, x_2; \theta)\, \phi_\eta(q, z; \theta) \left(\frac{1}{h_N}\left(d\mu_{X_{N,1}} - d\mu_X\right) - dG_{X_1}\right)\right| \\
&\left(d\mu_{X_2} + d\mu_{X_{N,2}}\right) d\mu_Z \\
\leq &\, 2\|\Psi\|_\infty \|f_\varepsilon\|_\infty \left\|f_\eta\right\|_\infty \sup_{q\in(\delta,1-\delta),\theta} \left|\phi_\varepsilon(q, x_1; \theta) \left(\frac{1}{h_N}\left(d\mu_{X_{N,1}} - d\mu_X\right) - dG_{X_1}\right)\right| \\
= &\, o(1).
\end{aligned}$$

since

$$\phi_\varepsilon(q, x_2; \theta) = f_\varepsilon\left(F_V^{-1}(q) - h(x; \theta)\right)$$

indexed by q, θ is a sub-class of Γ. Turning to S_1, note that

$$S_1(q) = \int \tilde{\Psi}(x_1, x_2, z, q, \theta)\, dG_{X_1} \left(d\mu_{X_2} - d\mu_{X_{N,2}}\right) d\mu_Z$$

where $\tilde{\Psi}(x_1, x_2, z, q, \theta) = \Psi(x_1, x_2, z)\, \phi_\varepsilon(q, x_1; \theta)\, \phi_\varepsilon(q, x_2; \theta)\, \phi_\eta(q, z; \theta)$ and $\int \tilde{\Psi}(x_1, x_2, z, q, \theta)\, dG_{X_1}$ is a bounded uniformly continuous function of (x_2, z, q, θ) for $q \in (\delta, 1-\delta)$. For any $\varepsilon > 0$, fix a compact set $\bar{\chi} = 1\{x : c_1 \leq x \leq c_2\}$ for $c_1, c_2 \in \mathbb{R}^{k_X}$,

such that $\mu_X(\chi \backslash \bar{\chi}) \leq \varepsilon$. By the triangle inequality,

$$\begin{aligned} & \left| \int \int \tilde{\Psi}(x_1, x_2, z, q, \theta) \, dG_{X_1} \left(d\mu_{X_2} - d\mu_{X_{N,2}} \right) \right| \\ \leq & \|G\|_\infty \left(\mu_X(\chi \backslash \bar{\chi}) + \mu_{X_N}(\chi \backslash \bar{\chi}) \right) + \left| \int_{\bar{\chi}} \int \tilde{\Psi}(x_1, x_2, z, q, \theta) \, dG_{X_1} \left(d\mu_{X_2} - d\mu_{X_{N,2}} \right) \right|. \end{aligned}$$

Since G is uniformly continuous, there exists a collection $\chi^1, \ldots, \chi^M$ of subsets $\chi^i = \left\{ x : c_1^i \leq x \leq c_2^i \right\}$ containing points $x^1, \ldots, x^M$ that cover $\bar{\chi}$ such that

$$\int_{\bar{\chi}} \int \tilde{\Psi}(x_1, x_2, z, q, \theta) \, dG_{X_1} \left(d\mu_{X_2} - d\mu_{X_{N,2}} \right) - \sum_{i=1}^{M} \int \tilde{\Psi}\left(x_1, x^i, z, q, \theta \right) dG_{X_1} \left(\mu_{X_2} - \mu_{X_{N,2}} \right)(\chi_i)$$

has absolute value strictly less than ε. Note that $1\left\{ x \in \chi^i \right\} \in \Gamma_X$. By the triangle inequality,

$$\begin{aligned} & \left| \int_{\bar{\chi}} \int \tilde{\Psi}(x_1, x_2, z, q, \theta) \, dG_{X_1} \left(d\mu_{X_2} - d\mu_{X_{N,2}} \right) \right| \\ < & \varepsilon + \left| \sum_{i=1}^{M} \int \tilde{\Psi}\left(x_1, x^i, z, q, \theta \right) dG_{X_1} \left(\mu_{X_2}(\chi_i) - \mu_{X_{N,2}}(\chi_i) \right) \right| \\ \leq & \varepsilon + M \|G\|_\infty \left\| \mu_{X_2} - \mu_{X_{N,2}} \right\|_\infty, \end{aligned}$$

where $\left\| \mu_{X_2} - \mu_{X_{N,2}} \right\|_\infty = \sup_{\gamma_X \in \Gamma_X} \left| \left(\mu_{X_2} - \mu_{X_{N,2}} \right)(\gamma_X) \right|$. Thus,

$$\begin{aligned} & \left| \int \int \tilde{\Psi}(x_1, x_2, z, q, \theta) \, dG_{X_1} \left(d\mu_{X_2} - d\mu_{X_{N,2}} \right) \right| \\ \leq & \|G\|_\infty \left(\mu_X(\chi \backslash \bar{\chi}) + \mu_{X_N}(\chi \backslash \bar{\chi}) \right) + \varepsilon + M \|G\|_\infty \left\| d\mu_{X_2} - d\mu_{X_{N,2}} \right\|_\infty. \end{aligned}$$

Since $\limsup_{N \to \infty} \left\| d\mu_{X_2} - d\mu_{X_{N,2}} \right\|_\infty = 0$, we have that

$$\begin{aligned} & \lim_{N \to \infty} \sup_{z,q} \left| \int \int \tilde{\Psi}(x_1, x_2, z, q, \theta) \, dG_{X_1} \left(d\mu_{X_2} - d\mu_{X_{N,2}} \right) \right| \\ \leq & 2 \|G\|_\infty \mu_X(\chi \backslash \bar{\chi}) + \varepsilon \\ \leq & (2 \|G\|_\infty + 1) \varepsilon \end{aligned}$$

Since this inequality holds for all $\varepsilon > 0$,

$$\lim_{N \to \infty} \sup_{z,q} \left| \int \int \tilde{\Psi}(x_1, x_2, z, q, \theta) \, dG_{X_1} \left(d\mu_{X_2} - d\mu_{X_{N,2}} \right) \right| = 0$$

Thus,

$$\begin{aligned}\sup_{q\in(\delta,1-\delta),\theta}|S_1(q,N,\theta)| &\leq \sup_{q\in(\delta,1-\delta),\theta}\left|\int\tilde{\Psi}(x_1,x_2,z,q,\theta)\,dG_{X_1}\left(d\mu_{X_2}-d\mu_{X_{N,2}}\right)d\mu_Z\right| \\ &\leq \sup_{z,q}\left|\int\int\tilde{\Psi}(x_1,x_2,z,q,\theta)\,dG_{X_1}\left(d\mu_{X_2}-d\mu_{X_{N,2}}\right)\right|\to 0.\end{aligned}$$

Hence,

$$\begin{aligned}\sup_{q\in(\delta,1-\delta),\theta}\left|\frac{1}{h_N}R_1(q,N,\theta)\right| &\leq \sup_{q\in(\delta,1-\delta),\theta}|S_1(q,N,\theta)|+\sup_{q\in(\delta,1-\delta),\theta}|S_2(q,N,\theta)| \\ &\to 0\end{aligned}$$

Identical arguments show that $R_2\to 0$ and $R_3\to 0$. Lemma C.11 implies that

$$\inf_{q\in(\delta,1-\delta),\theta\in\Theta}\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z>0.$$

Therefore,

$$\frac{R(q)}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}\to 0.$$

It follows that equation (C.22) can be re-written as

$$\begin{aligned}\frac{1}{h_N}T_1 &= \int_\delta^{1-\delta}\frac{1}{h_N}T_1(q)\,dq \\ &= \int_\delta^{1-\delta}\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,dG_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq \\ &\quad+\int_\delta^{1-\delta}\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_X dG_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq \\ &\quad+\int_\delta^{1-\delta}\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}dG_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq \\ &\quad+o(1).\end{aligned}\tag{C.23}$$

To compute the limit of T_2, rewrite

$$\begin{aligned}T_2(q) &= \frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \\ &-\frac{\int\Psi(x_1,x_2,z)\,\phi_{\varepsilon,N}(q,x_1;\theta)\,\phi_{\varepsilon,N}(q,x_2;\theta)\,\phi_{\eta,N}(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}\end{aligned}$$

by observing that

$$
\begin{aligned}
& \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) - \phi_{\varepsilon, N}\left(q, x_1; \theta\right) \phi_{\varepsilon, N}\left(q, x_2; \theta\right) \phi_{\eta, N}\left(q, z; \theta\right) \\
= \; & \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \left[\phi_{\eta}\left(q, z; \theta\right) - \phi_{\eta, N}\left(q, z; \theta\right)\right] + \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta, N}\left(q, z; \theta\right) \\
& - \phi_{\varepsilon, N}\left(q, x_1; \theta\right) \phi_{\varepsilon, N}\left(q, x_2; \theta\right) \phi_{\eta, N}\left(q, z; \theta\right) \\
= \; & \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \left[\phi_{\eta}\left(q, z; \theta\right) - \phi_{\eta, N}\left(q, z; \theta\right)\right] \\
& + \phi_{\varepsilon}\left(q, x_1; \theta\right) \left[\phi_{\varepsilon}\left(q, x_2; \theta\right) - \phi_{\varepsilon, N}\left(q, x_2; \theta\right)\right] \phi_{\eta, N}\left(q, z; \theta\right) \\
& + \left[\phi_{\varepsilon}\left(q, x_1; \theta\right) - \phi_{\varepsilon, N}\left(q, x_1; \theta\right)\right] \phi_{\varepsilon, N}\left(q, x_2; \theta\right) \phi_{\eta, N}\left(q, z; \theta\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& \frac{1}{h_N} T_2\left(q\right) = \\
& \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \frac{1}{h_N} \left[\phi_{\eta}\left(q, z; \theta\right) - \phi_{\eta, N}\left(q, z; \theta\right)\right] d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\
& + \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon}\left(q, x_1; \theta\right) \frac{1}{h_N} \left[\phi_{\varepsilon}\left(q, x_2; \theta\right) - \phi_{\varepsilon, N}\left(q, x_2; \theta\right)\right] \phi_{\eta, N}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\
& + \frac{\int \Psi\left(x_1, x_2, z\right) \frac{1}{h_N} \left[\phi_{\varepsilon}\left(q, x_1; \theta\right) - \phi_{\varepsilon, N}\left(q, x_1; \theta\right)\right] \phi_{\varepsilon, N}\left(q, x_2; \theta\right) \phi_{\eta, N}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\
& = \frac{1}{h_N} \left(F^{-1}_{U;\theta}\left(q\right) - F^{-1}_{N,U;\theta}\left(q\right)\right) x \\
& \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) f'_{\eta}\left(F^{-1}_{U;\theta}\left(q\right) - g\left(z; \theta\right)\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\
& + \frac{1}{h_N} \left(F^{-1}_{V;\theta}\left(q\right) - F^{-1}_{N,V;\theta}\left(q\right)\right) x \\
& \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon}\left(q, x_1; \theta\right) f'_{\varepsilon}\left(F^{-1}_{V;\theta}\left(q\right) - h\left(x_2; \theta\right)\right) \phi_{\eta, N}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\
& + \frac{1}{h_N} \left(F^{-1}_{V;\theta}\left(q\right) - F^{-1}_{N,V;\theta}\left(q\right)\right) x \\
& \frac{\int \Psi\left(x_1, x_2, z\right) f'_{\varepsilon}\left(F^{-1}_{V;\theta}\left(q\right) - h\left(x_1; \theta\right)\right) \phi_{\varepsilon, N}\left(q, x_2; \theta\right) \phi_{\eta, N}\left(q, z; \theta\right) d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}\left(q, x_1; \theta\right) \phi_{\varepsilon}\left(q, x_2; \theta\right) \phi_{\eta}\left(q, z; \theta\right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\
& + o\left(1\right) = \qquad \text{(C.24)}
\end{aligned}
$$

$$
K_1\left(q\right) + K_2\left(q\right) + K_3\left(q\right) + o\left(1\right), \qquad \text{(C.25)}
$$

where the equality follows from a Taylor expansion and dominated convergence theorem (since f'_ε and f'_η are bounded).

Rewrite $K_1(q)$ as

$$\begin{aligned}
&\frac{1}{h_N}\left(F^{-1}_{U;\theta}(q) - F^{-1}_{N,U;\theta}(q)\right) \\
&\frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,f'_\eta\left(F^{-1}_{U;\theta}(q) - g(z;\theta)\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \\
&+\frac{1}{h_N}\left(F^{-1}_{U;\theta}(q) - F^{-1}_{N,U;\theta}(q)\right) \\
&\frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,f'_\eta\left(F^{-1}_{U;\theta}(q) - g(z;\theta)\right)d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \\
&-\frac{1}{h_N}\left(F^{-1}_{U;\theta}(q) - F^{-1}_{N,U;\theta}(q)\right) \\
&\frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,f'_\eta\left(F^{-1}_{U;\theta}(q) - g(z;\theta)\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \\
= \;& C_1(q) + C_2(q) - C_3(q)
\end{aligned}$$

where $C_2(q) - C_3(q)$ is not greater in absolute value than

$$\begin{aligned}
&\frac{\sup_{\theta,q\in(\delta,1-\delta)}\left|\frac{1}{h_N}\left(F^{-1}_{U;\theta}(q) - F^{-1}_{N,U;\theta}(q)\right)\right|}{\inf_\theta \int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \times \\
&\left|\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,f'_\eta\left(F^{-1}_{U;\theta}(q) - g(z;\theta)\right)\right. \\
&\left.\left(d\mu_{X_1}d\mu_{X_2}d\mu_Z - d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{N,Z}\right)\right|
\end{aligned}$$

which goes to 0 uniformly in $q \in (\delta, 1-\delta)$ by the same argument used to compute the limit of $T_1(q)$. To compute the limit of $C_1(q)$, note

$$\begin{aligned}
& F^{-1}_{U;\theta}(q) - F^{-1}_{N,U;\theta}(q) \\
= \;& \frac{1}{f_{U;\theta}\left(F^{-1}_{U,\theta}(q)\right)}\left(F_{U;\theta}\left(F^{-1}_{U;\theta}(q)\right) - F_{N,U;\theta}\left(F^{-1}_{U;\theta}(q)\right)\right) + o(1) \\
= \;& \frac{1}{f_{U;\theta}\left(F^{-1}_{U,\theta}(q)\right)}\int F_\eta\left(F^{-1}_{U;\theta}(q) - g(z;\theta)\right)(d\mu_Z - d\mu_{Z_N}) + o(1) \\
= \;& \frac{1}{f_{U;\theta}\left(F^{-1}_{U,\theta}(q)\right)}\int (\mu_Z - \mu_{Z_N})\left(1\left\{g(z;\theta) + \eta \le F^{-1}_{U;\theta}(q)\right\}\right)dF_\eta + o(1)
\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{1}{h_N}\left(F_{U;\theta}^{-1}(q)-F_{N,U;\theta}^{-1}(q)\right) &\rightarrow \frac{1}{f_{U;\theta}\left(F_{U,\theta}^{-1}(q)\right)}\int G_Z\left(1\left\{g(z;\theta)+\eta\leq F_{U;\theta}^{-1}(q)\right\}\right)dF_\eta \\ &= G_U^q(\theta)\end{aligned}$$

uniformly in $q\in(\delta,1-\delta)$. Hence, $K_1(q)$ converges to

$$G_U^q(\theta)\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,f_\eta'\left(F_{U;\theta}^{-1}(q)-g(z;\theta)\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}$$

Similar arguments show that $K_2(q)$ and $K_3(q)$ respectively converge to

$$G_V^q(\theta)\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x_2;\theta)\right)\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}$$

and

$$G_V^q(\theta)\frac{\int\Psi(x_1,x_2,z)\,f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x_1;\theta)\right)\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}.$$

Consequently, equation (C.25) can be written as

$$\begin{aligned}&\frac{1}{h_n}T_2=\int_\delta^{1-\delta}\frac{1}{h_n}T_2(q)\,dq= \\ &\int_\delta^{1-\delta}G_U^q(\theta)\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,f_\eta'\left(F_{U;\theta}^{-1}(q)-g(z;\theta)\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq \\ &+\int_\delta^{1-\delta}G_V^q(\theta)\frac{\int\Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x_2;\theta)\right)\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq \\ &+\int_\delta^{1-\delta}G_V^q(\theta)\frac{\int\Psi(x_1,x_2,z)\,f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x_1;\theta)\right)\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq \\ &+o(1).\end{aligned}\tag{C.26}$$

Finally, we rewrite

$$
\begin{aligned}
& T_3(q) = \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_{Z}} \times \\
& \left(1 - \frac{\int \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_{Z}}{\int \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}\right) = \\
& \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}} \\
& \times \left(-\tilde{T}_1(q) - \tilde{T}_2(q)\right) \\
= \quad & \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_{Z}}{\int \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_{Z}} \\
& \times \left(-\tilde{T}_1(q) - \tilde{T}_2(q)\right) \\
& + \left(\frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}\right. \\
& \left. - \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_{Z}}{\int \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_{Z}}\right) \\
& \times \left(-\tilde{T}_1(q) - \tilde{T}_2(q)\right) \qquad \text{(C.27)}
\end{aligned}
$$

where $\tilde{T}_1(q) = T_1(q)$ and $\tilde{T}_2(q) = T_2(q)$ evaluated at $\Psi = 1$. Since $\sup_{\theta, q \in (\delta, 1-\delta), N} \left| \frac{1}{h_N} \tilde{T}_1(q) \right|$ and $\sup_{\theta, q \in (\delta, 1-\delta), N} \left| \frac{1}{h_N} \tilde{T}_2(q) \right|$ are finite, and

$$
\begin{aligned}
& \sup_{\theta, q \in (\delta, 1-\delta), N} \left| \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}} \right. \\
& \left. - \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_{Z}}{\int \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_{Z}} \right| \to 0,
\end{aligned}
$$

we have that

$$
\begin{aligned}
& \frac{1}{h_N} T_3(q) = \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_{Z}}{\int \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_{Z}} \times \\
& \frac{1}{h_N} \left(-\tilde{T}_1(q) - \tilde{T}_2(q)\right) + o(1).
\end{aligned}
$$

Equations (C.23) and (C.26), along with $\tilde{T}_1(q) = T_1(q)$ and $\tilde{T}_2(q) = T_2(q)$, imply that

$$\begin{aligned}
&\frac{1}{h_N}\tilde{T}_1(q) + \frac{1}{h_N}\tilde{T}_2(q) \to \\
&G_U^q(\theta)\frac{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,f'_\eta\left(F^{-1}_{U;\theta}(q) - g(z;\theta)\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \\
&+G_V^q(\theta)\frac{\int \phi_\varepsilon(q,x_1;\theta)\,f'_\varepsilon\left(F^{-1}_{V;\theta}(q) - h(x_2;\theta)\right)\phi_\eta(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \\
&+G_V^q(\theta)\frac{\int f'_\varepsilon\left(F^{-1}_{V;\theta}(q) - h(x_1;\theta)\right)\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \\
&+\frac{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,dG_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \\
&+\frac{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_X dG_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \\
&+\frac{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}dG_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \\
=\;& -L_G(\theta,q)
\end{aligned} \tag{C.28}$$

uniformly in θ and $q \in (\delta, 1-\delta)$. Equations (C.27) and (C.28) imply that

$$\frac{1}{h_N}T_3 = \int_\delta^{1-\delta} \frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} L_G(\theta,q)\,dq + o(1) \tag{C.29}$$

uniformly in θ.

Together, equations (C.23), (C.26), (C.28) and (C.29) imply that

$$\begin{aligned}
&\frac{1}{h_N}\int_\delta^{1-\delta} \frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq \\
&-\frac{1}{h_N}\int_\delta^{1-\delta} \frac{\int \Psi(x_1,x_2,z)\,\phi_{\varepsilon,N}(q,x_1;\theta)\,\phi_{\varepsilon,N}(q,x_2;\theta)\,\phi_{\eta,N}(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}{\int \phi_{\varepsilon,N}(q,x_1;\theta)\,\phi_{\varepsilon,N}(q,x_2;\theta)\,\phi_{\eta,N}(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}dq
\end{aligned}$$

converges to $Lim_{G;\delta}(\theta)$, which equals

$$\int_{\delta}^{1-\delta} G_U^q(\theta) \frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,f'_\eta\left(F^{-1}_{U;\theta}(q)-g(z;\theta)\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}$$
$$+G_V^q(\theta)\frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,f'_\varepsilon\left(F^{-1}_{V;\theta}(q)-h(x_2;\theta)\right)\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}$$
$$+G_V^q(\theta)\frac{\int \Psi(x_1,x_2,z)\,f'_\varepsilon\left(F^{-1}_{V;\theta}(q)-h(x_1;\theta)\right)\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}$$
$$+\frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,dG_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}$$
$$+\frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_X dG_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}$$
$$+\frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}dG_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}$$
$$+\frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}L_G(\theta,q)\,dq, \tag{C.30}$$

where $L_G(\theta,q)$ is defined in equation (C.28). This expression is therefore the Hadamard derivative of interest. ■

Lemma C.11 *Suppose that f_ε is bounded away from zero on every compact interval, and $h(x;\theta)$ is uniformly μ_X-integrable over $\theta \in \Theta$, then for every $q \in (0,1)$, $\inf_{\theta\in\Theta}\int f_\varepsilon\left(F^{-1}_{V;\theta}(q)-h(x;\theta)\right)d\mu_X > 0$.*

Proof. First, we show that there exists $M < \infty$, such that $\inf_{\theta\in\Theta} F^{-1}_{V;\theta}(q) > -M$ and $\sup_{\theta\in\Theta} F^{-1}_{V;\theta}(q) < M$. To do so, it is enough to show that for any $\delta > 0$, there exists M such that $\sup_\theta \mathbb{P}(|h(x;\theta)+\varepsilon| > M) < \delta$. The triangle inequality implies that $\sup_\theta \mathbb{P}(|h(x;\theta)+\varepsilon| > M) \leq \sup_{\theta\in\Theta}\mathbb{P}\left(|h(x;\theta)| > \frac{M}{2}\right) + \mathbb{P}\left(|\varepsilon| > \frac{M}{2}\right)$. For large enough M, the second term is less than $\frac{\delta}{2}$ by definition and the first term is less than $\frac{\delta}{2}$ since $h(x;\theta)$ is uniformly integrable.

Since for each q the map from θ to $F^{-1}_{V;\theta}(q)$ lives in a compact interval, $F^{-1}_{V;\theta}(q)-h(x;\theta)$ is a uniformly integrable family. Therefore, $\inf_{\theta\in\Theta}\int f_\varepsilon\left(F^{-1}_{V;\theta}(q)-h(x;\theta)\right)d\mu_X > 0$ since f_ε is bounded away from zero on any compact interval. ■

C.2.3 Proof of Proposition 3.5

Proof of Part (i): We need to show that $\sup_\theta |(\psi_N - \psi_N(\theta)) - (\psi - \psi(\theta))|$ converges in probability to zero. By the triangle inequality,

$$\sup_\theta |(\psi_N - \psi_N(\theta)) - (\psi - \psi(\theta))| \leq |\psi_N - \psi| + \sup_\theta |\psi_N(\theta) - \psi(\theta)|.$$

Proposition 3.3(i) shows that $|\psi_N - \psi|$ converges in probability to 0. We now show that the second term also converges in probability to zero.

By definition of $\psi^\delta[m_X, m_Z]$,

$$\psi_N(\theta) - \psi(\theta) = \psi^0[\mu_{X_N}, \mu_{Z_N}](\theta) - \psi^0[\mu_X, \mu_Z](\theta).$$

Further, for any $\delta \in \left(0, \frac{1}{2}\right)$, we have that

$$\left|\psi^0[\mu_{X_N}, \mu_{Z_N}](\theta) - \psi^0[\mu_X, \mu_Z](\theta)\right| \leq \left|\psi^\delta[\mu_{X_N}, \mu_{Z_N}](\theta) - \psi^\delta[\mu_X, \mu_Z](\theta)\right| + 2\|\Psi\|_\infty \delta.$$

Proposition 3.4 implies that $\psi^\delta[\mu_X, \mu_Z] : L_\infty^\Gamma \to L_\infty^\Theta$ is uniformly continuous in μ_X, μ_Z. Since Γ_X is μ_X-Glivenko Cantelli, and Γ_Z is μ_Z-Glivenko Cantelli, $\sup_\theta \left|\psi^\delta[\mu_{X_N}, \mu_{Z_N}](\theta) - \psi^\delta[\mu_X, \mu_Z](\theta)\right|$ converges in probability to zero for any $\delta \in \left(0, \frac{1}{2}\right)$ by the continuous mapping theorem. Hence, $\sup_\theta |\psi_N(\theta) - \psi(\theta)|$ converges in probability to 0.

Proof of Part (ii): Consider the process

$$\begin{bmatrix} \sqrt{N}(\psi_N - \psi(\theta_0)) \\ \sqrt{N}(\mu_{X_N} - \mu_X) \\ \sqrt{N/2}(\mu_{Z_N} - \mu_Z) \end{bmatrix},$$

where $\sqrt{N}(\mu_{X_N} - \mu_X)$ is the empirical process indexed by Γ_X and $\sqrt{N/2}(\mu_{Z_N} - \mu_Z)$ is the empirical process indexed by Γ_Z. Proposition 3.3(ii) shows that this process converges weakly to the Gaussian process, $\tilde{G} = (G_\Psi, G_X, G_Z)$, which a mean zero Gaussian process with covariance kernel V.

By the functional delta method and the Hadamard derivative derived in Proposition 3.4, we have that

$$m_N^{\delta}(\theta) = \sqrt{N}(\psi_N - \psi(\theta_0)) - \sqrt{N}\left(\psi_N^{\delta}(\theta) - \psi^{\delta}(\theta)\right)$$

converges weakly to a mean zero Gaussian process

$$G_\Psi - \nabla_{(G_X, G_Z)}\psi^{\delta}[\mu_X, \mu_Z](\theta).$$

Therefore, there exists a sequence δ_N of positive numbers decreasing to 0 such that

$$d\left(m_N^{\delta_N}(\cdot), G_\Psi - \nabla_{(G_X, G_Z)}\psi^{\delta_N}[\mu_X, \mu_Z](\cdot)\right) \to 0,$$

where d is a metric for weak convergence, and (by Assumption 3.6(ii)c.)

$$\sup_{\|\theta-\theta_0\|\leq b_N}\left|\nabla_{(G_X, G_Z)}\psi^{\delta_N}[\mu_X, \mu_Z](\theta) - \nabla_{(G_X, G_Z)}\psi^{\delta_N}[\mu_X, \mu_Z](\theta_0)\right| = o_p(1).$$

In what follows, we fix such a sequence of δ_N.

We derive the limit distribution of $m_N^0(\theta_0) = \sqrt{N}(\psi_N - \psi_N(\theta_0))$ to show Condition 1(ii) a. By the triangle inequality,

$$\begin{aligned} & d\left(m_N^0(\theta_0), G_\Psi - \nabla_{(G_X, G_Z)}\psi^0[\mu_X, \mu_Z](\theta_0)\right) \\ \leq\ & d\left(m_N^0(\theta_0), m_N^{\delta_N}(\theta_0)\right) + d\left(m_N^{\delta_N}(\theta_0), G_\Psi - \nabla_{(G_X, G_Z)}\psi^{\delta_N}[\mu_X, \mu_Z](\theta_0)\right) \\ & + d\left(G_\Psi - \nabla_{(G_X, G_Z)}\psi^{\delta_N}[\mu_X, \mu_Z](\theta_0), G_\Psi - \nabla_{(G_X, G_Z)}\psi^0[\mu_X, \mu_Z](\theta_0)\right) \end{aligned}$$

The first term converges to zero as $N \to \infty$ by Assumption 3.6(ii)b. The second term converges to zero by the choice of δ_N. The third term goes to zero since

$$\left(G_\Psi - \nabla_{(G_X, G_Z)}\psi^{\delta_N}[\mu_X, \mu_Z](\theta_0)\right) - \left(G_\Psi - \nabla_{(G_X, G_Z)}\psi^0[\mu_X, \mu_Z](\theta_0)\right)$$

converges in probability and therefore in distribution to 0 (by Assumption 3.6(ii)d). Hence, $m_N^0(\theta_0)$ converges in distribution to $G_\Psi - \nabla_{(G_X, G_Z)}\psi^0[\mu_X, \mu_Z](\theta_0)$. Note that this limiting random variable is distributed $N\left(0, \lim_{\delta\to 0} V^{\delta}\right)$ where V^{δ} is the variance

of $G_\Psi - \nabla_{(G_X,G_Z)}\psi^\delta\left[\mu_X,\mu_Z\right](\theta_0)$.

Now, we verify Condition 1(ii) b. By the triangle inequality, for any sequence $\{b_N\}$ of positive numbers converging to zero,

$$\sup_{\|\theta-\theta_0\|\leq b_N}\left|m_N^0(\theta_0)-m_N^0(\theta)\right| \leq \sup_{\|\theta-\theta_0\|\leq b_N}\left|m_N^{\delta_N}(\theta_0)-m_N^{\delta_N}(\theta)\right| + 2\sup_{\|\theta-\theta_0\|\leq b_N}\left|m_N^0(\theta)-m_N^{\delta_N}(\theta)\right|.$$

Note that, by the triangle inequality,

$$\begin{aligned} d\left(m_N^{\delta_N}(\theta_0), m_N^{\delta_N}(\theta)\right) &\leq 2d\left(m_N^{\delta_N}(\cdot), G_\Psi - \nabla_{(G_X,G_Z)}\psi^{\delta_N}\left[\mu_X,\mu_Z\right](\cdot)\right) \\ &\quad + d\left(G_\Psi - \nabla_{(G_X,G_Z)}\psi^{\delta_N}\left[\mu_X,\mu_Z\right](\theta), G_\Psi - \nabla_{(G_X,G_Z)}\psi^{\delta_N}\left[\mu_X,\mu_Z\right](\theta_0)\right) \end{aligned}$$

converges to 0 since $\sup_{\|\theta-\theta_0\|\leq b_N}\left|\nabla_{(G_X,G_Z)}\psi^{\delta_N}\left[\mu_X,\mu_Z\right](\theta) - \nabla_{(G_X,G_Z)}\psi^{\delta_N}\left[\mu_X,\mu_Z\right](\theta_0)\right| = o_p(1)$. Assumption 3.6(ii)b. implies that $2E\sup_{\|\theta-\theta_0\|\leq b_N}\left|m_N^0(\theta)-m_N^{\delta_N}(\theta)\right|$ converges to zero as $N\to\infty$. Therefore,

$$2\sup_{\|\theta-\theta_0\|\leq b_N}\left|m_N^0(\theta)-m_N^{\delta_N}(\theta)\right|$$

converges in probability to zero. Hence,

$$\begin{aligned} \sqrt{N}\left(\left(\psi(\theta_0)-\psi_N(\theta_0)\right)-\left(\psi(\theta)-\psi_N(\theta)\right)\right) &= m_N^0(\theta_0)-m_N^0(\theta) \\ \Rightarrow \sup_{\|\theta-\theta_0\|\leq b_N}\left|\sqrt{N}\left(\left(\psi(\theta_0)-\psi_N(\theta_0)\right)-\left(\psi(\theta)-\psi_N(\theta)\right)\right)\right| &= \sup_{\|\theta-\theta_0\|\leq b_N}\left|m_N^0(\theta_0)-m_N^0(\theta)\right| \\ &= o_p(1). \end{aligned}$$

C.3 Auxiliary Results on Estimation

C.3.1 Primitive conditions for Assumption 3.4(i)

Axiom C.2 *(i) $\Psi(x_1,x_2,z)$ is bounded and symmetric in x_1 and x_2*

(ii) The quantities $\int \frac{|f_\varepsilon'(v-h(x;\theta_0))|}{\int f_\varepsilon(v-h(X;\theta_0))d\mu_X}d\mu_X$ and $\int \frac{|f_\eta'(u-g(z;\theta_0))|}{\int f_\eta(u-g(Z;\theta_0))d\mu_Z}d\mu_Z$ are uniformly bounded in v and u respectively

Lemma C.12 *If Assumption C.2 is satisfied, then* $\|\nabla\tilde{\psi}\|_\infty < \infty$. *Hence,* $\tilde{\psi}(v_1, v_2, u; \theta_0)$ *is Lipschitz continuous in* v_1, v_2 *and* u.

Proof. Note that

$$\begin{aligned}
&\tilde{\psi}(v_1, v_2, u) \\
= &\int \Psi(X_1, X_2, Z)\, d\mu_{X|v_1} d\mu_{X|v_2} d\mu_{Z|u} \\
= &\int \Psi(X_1, X_2, Z)\ \tilde{f}_{v,x}(v_1, X_1)\ \tilde{f}_{v,x}(v_2, X_2)\ \tilde{f}_{u,z}(u, Z)\ d\mu_{X_1} d\mu_{X_2} d\mu_Z \\
&\text{where } \tilde{f}_{v,x}(v, x) = \frac{f_\varepsilon(v - h(x;\theta_0))}{\int f_\varepsilon(v - h(X;\theta_0))\, d\mu_X} \text{ and } \tilde{f}_{u,z}(u, z) = \frac{f_\eta(u - g(z,\theta_0))}{\int f_\eta(u - g(Z,\theta_0))\, d\mu_Z}
\end{aligned}$$

We will only show $\tilde{\psi}(v_1, v_2, u)$ has a bounded derivative with respect to v_1 as the proof for the other two arguments are identical. Note that

$$\begin{aligned}
&\frac{\partial}{\partial v} \frac{f_\varepsilon(v - h(x;\theta_0))}{\int f_\varepsilon(v - h(X;\theta_0))\, d\mu_X} \\
= &\frac{f'_\varepsilon(v - h(x;\theta_0))}{\int f_\varepsilon(v - h(X;\theta_0))\, d\mu_X} - \frac{\frac{f'_\varepsilon(v-h(x;\theta_0))}{\int f_\varepsilon(v-h(X;\theta_0))d\mu_X} \int f'_\varepsilon(v - h(X;\theta_0))\, d\mu_X}{\left(\int f_\varepsilon(u - h(X;\theta_0))\, d\mu_X\right)^2}
\end{aligned} \tag{C.31}$$

If the expression in equation (C.31) is μ_X integrable in X, then the Dominated Convergence Theorem implies that the derivative $\frac{\partial}{\partial v_1}\tilde{\psi}(v_1, v_2, u)$ exists and is given by

$$\int \Psi(X_1, X_2, Z) \frac{\partial}{\partial v_1} \tilde{f}_{v,x}(v_1, X_1)\ \tilde{f}_{v,x}(v_2, X_2)\ \tilde{f}_{u,z}(u, Z)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z.$$

To proceed, we will show that

$$\begin{aligned}
\sup_v \left| \int \left(\frac{f'_\varepsilon(v - h(x;\theta_0))}{\int f_\varepsilon(v - h(X;\theta_0))\, d\mu_X} - \frac{f_\varepsilon(v - h(x;\theta_0)) \int f'_\varepsilon(v - h(X;\theta_0))\, d\mu_X}{\left(\int f_\varepsilon(v - h(X;\theta_0))\, d\mu_X\right)^2} \right) d\mu_X \right| &< \infty \\
\sup_u \left| \int \left(\frac{f'_\eta(u - g(z;\theta_0))}{\int f_\eta(u - g(Z,\theta_0))\, d\mu_Z} - \frac{f_\eta(u - g(z,\theta_0)) \int f'_\eta(u - g(Z,\theta_0))\, d\mu_Z}{\left(\int f_\eta(u - g(Z,\theta_0))\, d\mu_Z\right)^2} \right) d\mu_Z \right| &< \infty
\end{aligned}$$

for the first expression since the proof of the other expression is identical. Note that

$$\begin{aligned}
& \sup_v \left| \int \left(\frac{f_\varepsilon'\left(v - h\left(x;\theta_0\right)\right)}{\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X} - \frac{f_\varepsilon\left(v - h\left(x;\theta_0\right)\right) \int f_\varepsilon'\left(v - h\left(X;\theta_0\right)\right) d\mu_X}{\left(\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X\right)^2} \right) d\mu_X \right| \\
\leq & \sup_v \int \left| \frac{f_\varepsilon'\left(v - h\left(x;\theta_0\right)\right)}{\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X} - \frac{f_\varepsilon\left(v - h\left(x;\theta_0\right)\right) \int f_\varepsilon'\left(v - h\left(X;\theta_0\right)\right) d\mu_X}{\left(\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X\right)^2} \right| d\mu_X \\
\leq & \sup_v \int \left| \frac{f_\varepsilon'\left(v - h\left(x;\theta_0\right)\right)}{\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X} \right| + \left| \frac{f_\varepsilon\left(v - h\left(x;\theta_0\right)\right) \int f_\varepsilon'\left(v - h\left(X;\theta_0\right)\right) d\mu_X}{\left(\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X\right)^2} \right| d\mu_X \\
\leq & \sup_v \int \frac{\left|f_\varepsilon'\left(v - h\left(x;\theta_0\right)\right)\right|}{\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X} d\mu_X + \\
& \sup_v \int \frac{f_\varepsilon\left(v - h\left(x;\theta_0\right)\right) \sup_v \int \left|f_\varepsilon'\left(v - h\left(X;\theta_0\right)\right)\right| d\mu_X}{\left(\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X\right)^2} d\mu_X \\
\leq & \sup_v \int \frac{\left|f_\varepsilon'\left(v - h\left(x;\theta_0\right)\right)\right|}{\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X} d\mu_X \\
& + \sup_v \int \frac{f_\varepsilon\left(v - h\left(x;\theta_0\right)\right)}{\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X} d\mu_X \sup_v \int \frac{\left|f_\varepsilon'\left(v - h\left(X;\theta_0\right)\right)\right|}{\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X} d\mu_X \\
\leq & \sup_v \int \frac{\left|f_\varepsilon'\left(v - h\left(x;\theta_0\right)\right)\right|}{\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X} d\mu_X \left(1 + \sup_v \int \frac{f_\varepsilon\left(v - h\left(X;\theta_0\right)\right)}{\int f_\varepsilon\left(v - h\left(X;\theta_0\right)\right) d\mu_X} d\mu_X\right) < \infty
\end{aligned}$$

by Assumption C.2 (ii).

Since $\|\Psi\|_\infty < \infty$ (Assumption C.2 (i)) and

$$\begin{aligned}
& \int \tilde{f}_{v,x}\left(v_2, X_2\right) \tilde{f}_{u,z}\left(u, Z\right) d\mu_{X_1} d\mu_Z \\
= & \int \frac{f_\varepsilon\left(v_1 - h\left(X_1;\theta_0\right)\right)}{\int f_\varepsilon\left(v_1 - h\left(X_1;\theta_0\right)\right) d\mu_{X_1}} d\mu_{X_1} \int \frac{f_\eta\left(u - g\left(z,\theta_0\right)\right)}{\int f_\eta\left(u - g\left(Z,\theta_0\right)\right) d\mu_Z} d\mu_Z \leq 1,
\end{aligned}$$

we have that

$$\begin{aligned}
& \frac{\partial}{\partial v_1} \tilde{\psi}\left(v_1, v_2, u\right) \\
& \leq \left| \int \left(\frac{f_\varepsilon'\left(v_1 - h\left(x;\theta_0\right)\right)}{\int f_\varepsilon\left(v_1 - h\left(X;\theta_0\right)\right) d\mu_X} - \frac{f_\varepsilon\left(v_1 - h\left(x;\theta_0\right)\right) \int f_\varepsilon'\left(v_1 - h\left(X;\theta_0\right)\right) d\mu_X}{\left(\int f_\varepsilon\left(v_1 - h\left(X;\theta_0\right)\right) d\mu_X\right)^2} \right) d\mu_X \right| \times \\
& \|\Psi\|_\infty \int \frac{f_\varepsilon\left(v_1 - h\left(X_1;\theta_0\right)\right)}{\int f_\varepsilon\left(v_1 - h\left(X_1;\theta_0\right)\right) d\mu_{X_1}} d\mu_{X_1} \int \frac{f_\eta\left(u - g\left(z,\theta_0\right)\right)}{\int f_\eta\left(u - g\left(Z,\theta_0\right)\right) d\mu_Z} d\mu_Z \\
& < \infty.
\end{aligned}$$

■

C.3.2 Primitive conditions for Assumption 3.6(ii)

For each x and z, define the Lipschitz constants $h_{LC}(x) = \sup_{\theta\in\Theta} \frac{|h(x;\theta)-h(x;\theta')|}{\|\theta-\theta'\|}$, and $g_{LC}(z) = \sup_{\theta\in\Theta} \frac{|g(z;\theta)-g(z;\theta')|}{\|\theta-\theta'\|}$.

Axiom C.3 *(i) $\Psi(x_1, x_2, z)$ indexed by x_2 and z is μ_X-Donsker and $\Psi(x_1, x_2, z)$ indexed by x_1 and x_2 is μ_Z-Donsker*

(ii) f_ε and f_η are bounded away from zero on any compact interval of $\mathbb{R}$, and have continuous first derivatives

(iii) there exist constants $C_1, C_2 > 0$ such that

$$\max\left\{f_\varepsilon(v),\ f_\eta(v),\ \left|f_\varepsilon'(v)\right|,\ \left|f_\eta'(v)\right|, \sup_{\theta\in\Theta} P(|h(x;\theta)| > v), \sup_{\theta\in\Theta} P(|g(z;\theta)| > v)\right\} \leq C_1 \exp(-C_2|v|)$$

(iv) $\int h_{LC}(X)^4 d\mu_X$, $\int g_{LC}(Z)^4 d\mu_Z$, and $\left\|\nabla\tilde{\psi}_q\right\|_\infty$ are finite

(v) $\Psi(x_1, x_2, z) = \sum_{k=1}^K a_k \Psi_1^k(x_1)\Psi_2^k(x_2)\Psi_z(z)$ with $\left\|\Psi_\cdot^k\right\|_\infty < \infty$ for some constants $a_1, \ldots, a_K$

(vi) $\|f_\varepsilon''\|_\infty$, $\int_{-\infty}^{\infty} |f_\varepsilon''(v)|\, dv$, $\left\|f_\eta''\right\|_\infty$, and $\int_{-\infty}^{\infty} \left|f_\eta''(v)\right| dv$ are finite

(vii) ε and η have full support on $\mathbb{R}$

Theorem C.2.2 *If Assumption C.3 is satisfied, then Assumption 3.6(ii) is satisfied.*

Proof. Assumption 3.6(ii) a. is verified by Proposition C.2.

Assumption 3.6(ii) b. is verified by Proposition C.3.

Assumption 3.6(ii) c. is verified by Proposition C.4.

Assumption 3.6(ii) d. is verified by Proposition C.5. ■

C.3.3 Donsker Properties for Γ_X and Γ_Z

For each x, define the Lipschitz constant $h_{LC}(x) = \sup_{\theta\in\Theta} \frac{|h(x;\theta)-h(x;\theta')|}{\|\theta-\theta'\|}$.

Claim C.3.3 *Suppose*

1. $\left(\int h_{LC}(x)^2 d\mu_X\right)^{1/2}$, $\|f_\varepsilon\|_\infty$ and $\|\Psi\|_\infty$ are finite

2. $\Psi(x_1, x_2, z)$ indexed by x_2 and z is μ_X-Donsker

Then, we have that

1. $F_\varepsilon(c - h(x;\theta))$ indexed by c and θ is a μ_X-Donsker class.

2. If $\int_{-\infty}^{\infty} |f'_\varepsilon(v)|\, dv < \infty$, then $f_\varepsilon(c - h(x;\theta))$ indexed by c and θ is a μ_X-Donsker class

3. If $\int_{-\infty}^{\infty} |f''_\varepsilon(v)|\, dv < \infty$, then $f'_\varepsilon(c - h(x;\theta))$ indexed by c and θ is a μ_X-Donsker class.

Proof. We only spell out the argument for the second statement since the other two are analogous, as $\int_{-\infty}^{\infty} |f_\varepsilon(v)|\, dv = 1$ by definition. Consider the class

$$f_\varepsilon(c - h(x;\theta))$$

indexed by $c \in \mathbb{R}$ and $\theta \in \Theta$. We will show that this class is Donsker by bounding its L_2-bracketing number.

Fix a partition $-\infty = c_0 < c_1 < c_2 < \ldots < c_N = \infty$. Lets compute

$$\begin{aligned}
&\sup_{\theta\in\Theta} \int [f_\varepsilon(c_n - h(x;\theta)) - f_\varepsilon(c_{n+1} - h(x;\theta))]^2 d\mu_X \\
\leq\ & 2\|f_\varepsilon\|_\infty \sup_{\theta\in\Theta} \int |f_\varepsilon(c_n - h(x;\theta)) - f_\varepsilon(c_{n+1} - h(x;\theta))|\, d\mu_X \\
\leq\ & 2\|f_\varepsilon\|_\infty \sup_{\theta\in\Theta} \int \int_{c_n}^{c_{n+1}} |f'_\varepsilon(c - h(x;\theta))|\, dc d\mu_X
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_n \sup_{\theta\in\Theta} \int [f_\varepsilon(c_n - h(x;\theta)) - f_\varepsilon(c_{n+1} - h(x;\theta))]^2 d\mu_X \\
\leq\ & 2\|f_\varepsilon\|_\infty \sup_{\theta\in\Theta} \int \sum_n \int_{c_n}^{c_{n+1}} |f'_\varepsilon(c - h(x;\theta))|\, dc d\mu_X \\
=\ & 2\|f_\varepsilon\|_\infty \sup_{\theta\in\Theta} \int \int_{-\infty}^{\infty} |f'_\varepsilon(c - h(x;\theta))|\, dc d\mu_X \\
=\ & 2\|f_\varepsilon\|_\infty \int_{-\infty}^{\infty} |f'_\varepsilon(c)|\, dc \\
=\ & \tilde{K} < \infty
\end{aligned} \tag{C.32}$$

where $\tilde{K}$ does not depend on the choice of $c_0 < c_1 < c_2 < \ldots < c_N$. Now, consider the function

$$\tilde{f}(a) = 2\|f_\varepsilon\|_\infty \int_{-\infty}^{a} |f'_\varepsilon(c)|\, dc.$$

Note that $\tilde{f}(a)$ is continuous, non-decreasing and has image $\left[0, \tilde{f}(\infty)\right]$. For any N and $n \in \{0, \ldots, N\}$ define

$$c_i = \tilde{f}^{-1}\left(\frac{n}{N}\right).$$

Then, for each n inequality in equation (C.32),

$$\sup_{\theta\in\Theta} \int \left[f_\varepsilon(c_n - h(x;\theta)) - f_\varepsilon(c_{n+1} - h(x;\theta))\right]^2 d\mu_X \leq \frac{\tilde{K}}{N}.$$

Consider an $1/\sqrt{N}$-net $\Theta \subseteq \mathbb{R}^d$, Θ_i for $i \in \{1, \ldots, D\}$. Note that $D = \left(\sqrt{N}\text{diam}(\Theta)\right)^d$. For each Θ_i and each n, define the bracket

$$\left[\inf_{\theta\in\Theta_i} \inf_{c\in[c_n,c_{n+1}]} f_\varepsilon(c - h(x;\theta)), \sup_{\theta\in\Theta_i} \sup_{c\in[c_n,c_{n+1}]} f_\varepsilon(c - h(x;\theta))\right].$$

The volume of these brackets are

$$\begin{aligned}
&\left(\int \left[\sup_{\theta\in\Theta_i} \sup_{c\in[c_n,c_{n+1}]} f_\varepsilon(c - h(x;\theta)) - \inf_{\theta\in\Theta_i} \inf_{c\in[c_n,c_{n+1}]} f_\varepsilon(c - h(x;\theta))\right]^2 d\mu_X\right)^{1/2} \\
= &\left(\int \left[f_\varepsilon\left(c^+ - h\left(x;\theta^+\right)\right) - f_\varepsilon\left(c^- - h\left(x;\theta^-\right)\right)\right]^2 d\mu_X\right)^{1/2} \\
\leq &\left(\int \left[f_\varepsilon\left(c^+ - h\left(x;\theta^+\right)\right) - f_\varepsilon\left(c^- - h\left(x;\theta^+\right)\right)\right]^2 d\mu_X\right)^{1/2} \\
&+ \left(\int \left[f_\varepsilon\left(c^- - h\left(x;\theta^+\right)\right) - f_\varepsilon\left(c^- - h\left(x;\theta^-\right)\right)\right]^2 d\mu_X\right)^{1/2} \\
\leq &\left(\frac{\tilde{K}}{N}\right)^{1/2} + \left\|f'_\varepsilon\right\|_\infty \left(\int h_{LC}(x)^2 d\mu_X\right)^{1/2} \sup_{\theta,\theta'\in\Theta_i} \left\|\theta - \theta'\right\| \\
= &\left(\frac{\tilde{K}}{N}\right)^{1/2} + \frac{\left\|f'_\varepsilon\right\|_\infty}{\sqrt{N}} \left(\int h_{LC}(x)^2 d\mu_X\right)^{1/2} = KN^{-1/2}.
\end{aligned}$$

Therefore, the ε-bracketing number is bounded by a polynomial in $1/\varepsilon$. Therefore, $\int_0^\infty \sqrt{\log \mathcal{N}(\varepsilon)}d\varepsilon$ is finite, where $\mathcal{N}(\varepsilon)$ be the ε bracketing number of this class. By van der

Vaart (2000) Theorem 2.5.6, it follows that $f_\varepsilon\left(c - h\left(x;\theta\right)\right)$ indexed by $c \in \mathbb{R}$ and $\theta \in \Theta$ is a μ_X-Donsker class. ■

Proposition C.2 *Suppose that the conditions for Claim C.3.3 hold and* $\left\|f_\eta\right\|_\infty$*, then* Γ_X *is a* μ_X-*Donsker class. Analogous conditions imply that* Γ_Z *is a* μ_Z-*Donsker class.*

Proof. We only need to show that the terms

$$\Psi\left(x_1, x_2, z\right) f_\varepsilon\left(F_{V;\theta}^{-1}\left(q\right) - h\left(x_1;\theta\right)\right) f_\varepsilon\left(F_{V;\theta}^{-1}\left(q\right) - h\left(x_2;\theta\right)\right) f_\eta\left(F_{U;\theta}^{-1}\left(q\right) - g\left(z;\theta\right)\right)$$

and

$$\Psi\left(x_1, x_2, z\right) f_\varepsilon'\left(F_{V;\theta}^{-1}\left(q\right) - h\left(x_1;\theta\right)\right) f_\varepsilon\left(F_{V;\theta}^{-1}\left(q\right) - h\left(x_2;\theta\right)\right) f_\eta\left(F_{U;\theta}^{-1}\left(q\right) - g\left(z;\theta\right)\right)$$

indexed by $\left(x_1, z, q\,, \theta\right)$ are μ_X-Donsker classes. This is because the terms $1\left\{c_1 \leq x \leq c_2\right\}$ are μ_X-Donsker since they are intersections of half-spaces, and therefore suitably measurable VC-classes. The remaining terms are μ_X-Donsker by Claim C.3.3.

Note that $f_\varepsilon\left(F_{V;\theta}^{-1}\left(q\right) - h\left(x_2;\theta\right)\right)$ indexed by $\left(q\,, \theta\right)$ is a sub-class of the μ_X-Donsker class $f_\varepsilon\left(c - h\left(x_2;\theta\right)\right)$ indexed by $\left(c\,, \theta\right)$, and is therefore μ_X-Donsker. Further, the quantities

$$\Psi\left(x_1, x_2, z\right) f_\varepsilon\left(F_{V;\theta}^{-1}\left(q\right) - h\left(x_1;\theta\right)\right) f_\eta\left(F_{U;\theta}^{-1}\left(q\right) - g\left(z;\theta\right)\right)$$

are uniformly bounded and measurable since $\left\|\Psi\right\|_\infty$, $\left\|f_\varepsilon\right\|_\infty$ and $\left\|f_\eta\right\|_\infty$ are finite. Since the product of two bounded Donsker classes is Donsker (van der Vaart (2000), example 2.10.8), we have that

$$\Psi\left(x_1, x_2, z\right) f_\varepsilon\left(F_{V;\theta}^{-1}\left(q\right) - h\left(x_1;\theta\right)\right) f_\varepsilon\left(F_{V;\theta}^{-1}\left(q\right) - h\left(x_2;\theta\right)\right) f_\eta\left(F_{U;\theta}^{-1}\left(q\right) - g\left(z;\theta\right)\right)$$

and

$$\Psi\left(x_1, x_2, z\right) f_\varepsilon'\left(F_{V;\theta}^{-1}\left(q\right) - h\left(x_1;\theta\right)\right) f_\varepsilon\left(F_{V;\theta}^{-1}\left(q\right) - h\left(x_2;\theta\right)\right) f_\eta\left(F_{U;\theta}^{-1}\left(q\right) - g\left(z;\theta\right)\right)$$

indexed by $\left(x_2, z, q\,, \theta\right)$ are μ_X-Donsker classes. ■

C.3.4 Primitive Conditions for Assumption 3.6(ii) b.

Our result verifying Assumption 3.6(ii) b. is stated in Proposition C.3 below. The main technical difficulty is solved in the following lemma. This result requires preliminaries proved below in Appendix C.3.4.

For each x and z, define the Lipschitz constants $h_{LC}(x) = \sup_{\theta\in\Theta} \frac{|h(x;\theta)-h(x;\theta')|}{\|\theta-\theta'\|}$, and $g_{LC}(z) = \sup_{\theta\in\Theta} \frac{|g(z;\theta)-g(z;\theta')|}{\|\theta-\theta'\|}$.

Lemma C.13 *Suppose that $\int h_{LC}(X)^4 \, d\mu_X$ is finite, and there exist constants C_1, $C_2 > 0$ such that*

$$\max\left\{ \left|f_\varepsilon'(v)\right|, \sup_{\theta\in\Theta} P\left(|h(x;\theta)| > v\right) \right\} \leq C_1 \exp\left(-C_2|v|\right).$$

Then, for any function $\Psi(x)$ with $\|\Psi\|_\infty < \infty$, we have that (i)

$$E \sup_\theta \int \left| \sqrt{N}\left(\mu_{X_N} - \mu_X\right)\left(\Psi(X) f_\varepsilon\left(v - h(X;\theta)\right)\right) \right| dv$$

is bounded and

(ii) for any sequence of positive numbers $\{r_N\}$ which decrease to 0 as $N \to \infty$,

$$E \sup_{\|\theta_1-\theta_2\|\leq r_N} \int \left| \sqrt{N}\left(\mu_{X_N} - \mu_X\right)\left(\Psi(X)\left[f_\varepsilon\left(v - h(X;\theta_1)\right) - f_\varepsilon\left(v - h(X;\theta_2)\right)\right]\right) \right| dv \to 0.$$

Proof. The argument combines ideas from Pollard (2002) recursive proof of Ossiander's bracketing functional central limit theorem and an application of Boucheron *et al.* (2003) (Theorem 2) concentration inequality.

Let D be the diameter of the parameter space Θ and for nonnegative integer i, let $\delta_i = D2^{-i}$. Fix a natural number i^*. Fix a δ_{i^*} net of Θ of size $N(\delta_{i^*})$ and for each $\theta \in \Theta$ let $B(\theta; i^*)$ be the center of a ball in this δ_{i^*} net which contains θ. For any nonnegative integer $i < i^*$, fix a δ_i net of Θ of size $N(\delta_i)$ and recursively define $B(\theta; i)$ to be the center of a ball in this δ_i net which contains $B(\theta; i+1)$. Note that this definition implies $d(\theta; B(\theta; i^*)) \leq \delta_{i^*}$, $d(B(\theta, i), B(\theta; i+1)) \leq \delta_i$, and that $B(\theta; i)$ takes on at most $N(\delta_i)$ distinct values. By repeated application of the triangle inequality, $d(B(\theta; i), \theta) \leq 2\delta_i$ for all θ. Let $C_\varepsilon = \int_{-\infty}^{\infty} |f_\varepsilon'(v)| \, dv$. Note that $C_\varepsilon < \infty$ by our exponential tail bound on f_ε'.

For each $i \leq i^*$, let $V_i = \sqrt{N}\frac{\delta_i}{\sqrt{\log N(\delta_i)}}$, let

$$R_i(\theta, v) = \sqrt{N}\left(\mu_{X_N} - \mu_X\right)\Psi(X)\left[f_\varepsilon\left(v - h\left(X;\theta\right)\right) - f_\varepsilon\left(v - h\left(X;B\left(\theta;i\right)\right)\right)\right]$$

and $T_i(\theta) = \left\{x : h_{LC}(x) \leq \frac{V_i}{2\delta_i}\right\}$.

To prove part (i), we separately bound $E\sup_\theta \int \left|R_0(\theta, v)\, T_0(\theta)\right| dv$ and $E\sup_\theta \int \left|R_0(\theta, v)\, T_0^c(\theta)\right| dv$. To prove part (ii), we must similarly show that $E\sup_\theta \int \left|R_i(\theta, v)\, T_i(\theta)\right| dv$ and $E\sup_\theta \int \left|R_i(\theta, v)\, T_i^c(\theta)\right| dv$ go to 0 as $i \to \infty$.

As noted by Pollard (2002),

$$R_iT_i = R_{i+1}T_{i+1} - R_{i+1}T_i^cT_{i+1} + \left(R_i - R_{i+1}\right)T_iT_{i+1} + R_iT_iT_{i+1}^c.$$

It follows that

$$\begin{aligned}
& E\sup_\theta \int \left|R_i(\theta, v)\, T_i(\theta)\right| dv \\
\leq\ & E\sup_\theta \int \left|R_{i+1}(\theta, v)\right| T_{i+1}(\theta)\, dv + E\sup_\theta \int \left|R_{i+1}(\theta, v)\right| T_i^c(\theta)\, T_{i+1}(\theta)\, dv \\
& + E\sup_\theta \int \left|\left(R_i(\theta, v) - R_{i+1}(\theta, v)\right)\right| T_i(\theta)\, T_{i+1}(\theta)\, dv \\
& + E\sup_\theta \int \left|\left(R_i(\theta, v)\right)\right| T_i(\theta)\, T_{i+1}^c(\theta)\, dv \\
\Rightarrow\ & E\sup_\theta \int \left|R_0(\theta, v)\, T_0(\theta)\right| dv \\
\leq\ & E\sup_\theta \int \left|R_{i^*}(\theta, v)\, T_{i^*}(\theta)\right| dv \\
& + \sum_{i=0}^{i^*-1}\left\{E\sup_\theta \int \left|R_{i+1}(\theta, v)\right| T_i^c(\theta)\, T_{i+1}(\theta)\, dv\right. \\
& + E\sup_\theta \int \left|R_i(\theta, v)\right| T_i(\theta)\, T_{i+1}^c(\theta)\, dv \\
& \left. + E\sup_\theta \int \left|\left(R_i(\theta, v) - R_{i+1}(\theta, v)\right)\right| T_i(\theta)\, T_{i+1}(\theta)\, dv\right\}. \qquad \text{(C.33)}
\end{aligned}$$

We need to show that each of the terms above is bounded. First, we show that summation is bounded. Lemmas C.14 and C.15 (below) imply that there exists a constant K such that each of the terms in the summation is no greater than $K\sqrt{\log N(\delta_i)}\delta_i$. Therefore, equation

(C.33) implies

$$E\sup_{\theta}\int\left|R_i\left(\theta,v\right)T_i\left(\theta\right)\right|dv\leq E\sup_{\theta}\int\left|R_{i^*}\left(\theta,v\right)T_{i^*}\left(\theta\right)\right|dv+K\sum_{j=i}^{i^*-1}\sqrt{\log N\left(\delta_j\right)}\delta_j. \quad \text{(C.34)}$$

We now show that as $i^*\to\infty$,

$$E\sup_{\theta}\int\left|R_{i^*}\left(\theta,v\right)T_{i^*}\left(\theta\right)\right|dv\to 0.$$

For any i, we have that,

$$\begin{aligned}
& E\sup_{\theta}\int\left|R_i\left(\theta,v\right)T_i\left(\theta\right)\right|dv \\
= \;& E\sup_{\theta}\int\left|\sqrt{N}\left(\mu_{X_N}-\mu_X\right)\Psi\left(X\right)\left[f_\varepsilon\left(v-h\left(X;\theta\right)\right)-f_\varepsilon\left(v-h\left(X;B\left(\theta;i\right)\right)\right)\right]\right|dv \\
\leq \;& \sqrt{N}E\sup_{\theta}\int\left(\mu_{X_N}\left\|\Psi\right\|_\infty\left|f_\varepsilon\left(v-h\left(X;\theta\right)\right)-f_\varepsilon\left(v-h\left(X;B\left(\theta;i\right)\right)\right)\right|\right. \\
& \left.+\mu_X\left\|\Psi\right\|_\infty\left|f_\varepsilon\left(v-h\left(X;\theta\right)\right)-f_\varepsilon\left(v-h\left(X;B\left(\theta;i\right)\right)\right)\right|\right)dv \\
\leq \;& \sqrt{N}\left\|\Psi\right\|_\infty E\sup_{\theta}\frac{1}{N}\sum_{j=1}^{N}\int\left(\left|f_\varepsilon\left(v-h\left(X_j;\theta\right)\right)-f_\varepsilon\left(v-h\left(X_j;B\left(\theta;i\right)\right)\right)\right|\right. \\
& \left.+\mu_X\left\|\Psi\right\|_\infty\left|f_\varepsilon\left(v-h\left(X;\theta\right)\right)-f_\varepsilon\left(v-h\left(X;B\left(\theta;i\right)\right)\right)\right|\right)dv \\
\leq \;& \sqrt{N}\left\|\Psi\right\|_\infty E\sup_{\theta}\frac{1}{N}\sum_{j=1}^{N}C_\varepsilon\left(\left|h\left(X_j;\theta\right)-h\left(X_j;B\left(\theta;i\right)\right)\right|+\mu_X\left|h\left(X;\theta\right)-h\left(X;B\left(\theta;i\right)\right)\right|\right) \\
\leq \;& \sqrt{N}\left\|\Psi\right\|_\infty E\left(\frac{1}{N}\sum_{j=1}^{N}C_\varepsilon\left(2\delta_i\left|h_{LC}\left(X_j\right)\right|+\mu_X 2\delta_i\left|h_{LC}\left(X\right)\right|\right)\right) \\
= \;& 4\delta_i\left\|\Psi\right\|_\infty C_\varepsilon\sqrt{N}\mu_X h_{LC}\left(X\right). \qquad \text{(C.35)}
\end{aligned}$$

Hence, equation (C.34) implies that for any $i^*>i$,

$$E\sup_{\theta}\int\left|R_i\left(\theta,v\right)T_i\left(\theta\right)\right|dv\leq 4\delta_{i^*}\left\|\Psi\right\|_\infty C_\varepsilon\sqrt{N}\mu_X h_{LC}\left(X\right)+K\sum_{j=i}^{\infty}\delta_j\sqrt{\log N\left(\delta_j\right)}.$$

Therefore, for a universal constant K',

$$\begin{aligned}
E\sup_{\theta}\int\left|R_i\left(\theta,v\right)T_i\left(\theta\right)\right|dv &\leq K'\int_0^{\delta_i}\sqrt{\log N\left(\delta\right)}d\delta \\
\text{and } E\sup_{\theta}\int\left|R_0\left(\theta,v\right)T_0\left(\theta\right)\right|dv &\leq K'\int_0^{\infty}\sqrt{\log N\left(\delta\right)}d\delta<\infty.
\end{aligned}$$

Note that

$$\begin{aligned}
& E\sup_{\theta}\int\left|R_i\left(\theta,v\right)T_i^c\left(\theta\right)\right|dv \\
= \; & E\sup_{\theta}\int_{-\infty}^{\infty}\left|\sqrt{N}\left(\mu_{X_N}-\mu_X\right)\left[f_\varepsilon\left(v-h\left(X;\theta\right)\right)-f_\varepsilon\left(v-h\left(X;B\left(\theta;i\right)\right)\right)\right]\right| \\
& \left\{X:h_{LC}\left(X\right)>\frac{V_i}{2\delta_i}\right\}dv \\
& \leq 2\delta_i E\sup_{\theta}\sqrt{N}\left(\mu_{X_N}+\mu_X\right)h_{LC}\left(X\right)\left\{X:h_{LC}\left(X\right)>\frac{\sqrt{N}}{2\sqrt{\log N\left(\delta_i\right)}}\right\}\int_{-\infty}^{\infty}\left|f_\varepsilon'\left(v\right)\right|dv \\
\leq \; & 4\delta_i\sqrt{N}\int_{-\infty}^{\infty}\left|f_\varepsilon'\left(v\right)\right|dv\mu_X h_{LC}\left(X\right)^4\left[\frac{2\sqrt{\log N\left(\delta_i\right)}}{\sqrt{N}}\right]^3 \\
= \; & 32\frac{1}{N}\int_{-\infty}^{\infty}\left|f_\varepsilon'\left(v\right)\right|dv\mu_X h_{LC}\left(X\right)^4\delta_i\left(\log N\left(\delta_i\right)\right)^{\frac{3}{2}}
\end{aligned}$$

Since $N\left(\delta_i\right)$ is not greater than some polynomial in $\frac{1}{\delta_i}$, $\sup_i \delta_i\left(\log N\left(\delta_i\right)\right)^{\frac{3}{2}}<\infty$, we have that

$$\begin{aligned}
& \sup_N E\sup_{\theta}\int\left|R_0\left(\theta,v\right)\right|dv \\
\leq \; & \sup_N E\sup_{\theta}\int\left|R_0\left(\theta,v\right)T_0\left(\theta\right)\right|dv+\sup_N E\sup_{\theta}\int\left|R_0\left(\theta,v\right)T_0^c\left(\theta\right)\right|dv \\
\leq \; & K'\int_0^{\infty}\sqrt{\log N\left(\delta\right)}d\delta+32\frac{1}{N}\int_{-\infty}^{\infty}\left|f_\varepsilon'\left(v\right)\right|dv\mu_X h_{LC}\left(X\right)^4\delta_0\left(\log N\left(\delta_0\right)\right)^{\frac{3}{2}} \\
< \; & \infty.
\end{aligned}$$

This completes the proof for Part (i). Similarly, for any sequence of $i_N\to\infty$, as $N\to\infty$,

$$\begin{aligned}
& E\sup_{\theta}\int\left|R_{i_N}\left(\theta,v\right)\right|dv \\
\leq \; & E\sup_{\theta}\int\left|R_{i_N}\left(\theta,v\right)T_{i_N}\left(\theta\right)\right|dv+E\sup_{\theta}\int\left|R_{i_N}\left(\theta,v\right)T_{i_N}^c\left(\theta\right)\right|dv\to 0
\end{aligned}$$

■

We are now ready to show the main result:

Proposition C.3 *If the following assumptions are satisfied*

(i) Γ_X *and* Γ_Z *are respectively* μ_X*- and* μ_Z*- Donsker*

(ii) f_ε *and* f_η *are bounded away from zero on any compact interval of* $\mathbb{R}$*, and have continuous*

first derivatives

(iii) there exist constants C_1, $C_2 > 0$ *such that*

$$\max\left\{f_\varepsilon(v),\ f_\eta(v),\ \left|f'_\varepsilon(v)\right|,\ \left|f'_\eta(v)\right|, \sup_{\theta\in\Theta} P\left(\left|h\left(x;\theta\right)\right| > v\right), \sup_{\theta\in\Theta} P\left(\left|g\left(z;\theta\right)\right| > v\right)\right\}$$
$$\leq C_1 \exp\left(-C_2\left|v\right|\right)$$

(iv) $\int h_{LC}(X)^4\, d\mu_X$, $\int g_{LC}(Z)^4\, d\mu_Z$, *and* $\left\|\nabla\tilde{\psi}_q\right\|_\infty$ *are finite*

(v) $\Psi(x_1, x_2, z) = \sum_{k=1}^K a_k \Psi_1^k(x_1)\Psi_2^k(x_2)\Psi_z(z)$ *with* $\left\|\Psi_\cdot^k\right\|_\infty < \infty$ *for some constants* $a_1, \ldots, a_K$

then for any sequence of positive δ_N *and* r_N *decreasing to 0*

$$\sqrt{N}E \sup_{\|\theta-\theta_0\|\leq r_N} \left|\left(\psi\left[\mu_X, \mu_Z\right](\theta) - \psi\left[\mu_{X_N}, \mu_{Z_N}\right](\theta)\right) - \left(\psi^{\delta_N}\left[\mu_X, \mu_Z\right](\theta) - \psi^{\delta_N}\left[\mu_{X_N}, \mu_{Z_N}\right](\theta)\right)\right|$$

converges to 0 *as* $N \to \infty$.

Proof. The proof proceeds by first manipulating this expression into a sum of similar terms which can all be handed by Lemma C.13. To ease notation, define

$$\begin{aligned}
\phi_\eta(q, z;\theta) &= f_\eta\left(F^{-1}_{U;\theta}(q) - g\left(z;\theta\right)\right) \\
\phi_{\eta,N}(q, z;\theta) &= f_\eta\left(F^{-1}_{N,U;\theta}(q) - g\left(z;\theta\right)\right) \\
\phi_\varepsilon(q, x;\theta) &= f_\varepsilon\left(F^{-1}_{V;\theta}(q) - h\left(x;\theta\right)\right) \\
\phi_{\varepsilon,N}(q, x;\theta) &= f_\varepsilon\left(F^{-1}_{N,V;\theta}(q) - h\left(x;\theta\right)\right).
\end{aligned}$$

First, note that

$$
\begin{aligned}
& \left(\psi\left[\mu_X, \mu_Z\right](\theta) - \psi\left[\mu_{X_N}, \mu_{Z_N}\right](\theta)\right) - \left(\psi^{\delta}\left[\mu_X, \mu_Z\right](\theta) - \psi^{\delta}\left[\mu_{X_N}, \mu_{Z_N}\right](\theta)\right) \\
= & \left[\int_0^1 - \int_{\delta}^{1-\delta}\right] \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\
& - \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}} dq \\
= & \left(\left[\int_0^1 - \int_{\delta}^{1-\delta}\right] \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} \right. \\
& \left. - \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \right) \\
& + \left(\left[\int_0^1 - \int_{\delta}^{1-\delta}\right] \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} \right. \\
& \left. - \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}} dq \right) \\
= & A_1 + A_2
\end{aligned}
$$

First, we bound the absolute value of A_1. Since $q_{N,V;\theta}(q) = F_{V;\theta}\left(F^{-1}_{N,V;\theta}(q)\right)$ and $\phi_{\varepsilon}\left(q_{N,V;\theta}(q), x; \theta\right) = \phi_{\varepsilon,N}(q, x; \theta)$, we have that

$$
\begin{aligned}
A_1 = & \left[\int_0^1 - \int_{\delta}^{1-\delta}\right] \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q, x_1; \theta)\, \phi_{\varepsilon}(q, x_2; \theta)\, \phi_{\eta}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \\
& - \left[\int_0^1 - \int_{\delta}^{1-\delta}\right] \frac{\int \Psi(x_1, x_2, z)\, \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon,N}(q, x_1; \theta)\, \phi_{\varepsilon,N}(q, x_2; \theta)\, \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq
\end{aligned}
$$

By a first order Taylor expansion,

$$
\begin{aligned}
\sqrt{N} E \sup_{\theta} |A_1| & \leq \sqrt{N} \left[\int_0^1 - \int_{\delta}^{1-\delta}\right] \left\|\nabla \tilde{\psi}_q\right\|_{\infty} \left(2E \sup_{\theta} |q_{N,V;\theta}(q) - q| + E \sup_{\theta} |q_{N,U;\theta}(q) - q|\right) dq \\
& \leq 4\delta \left\|\nabla \tilde{\psi}_q\right\|_{\infty} \left(E \sup_{\theta} |q_{N,V;\theta}(q) - q| + E \sup_{\theta} |q_{N,U;\theta}(q) - q|\right).
\end{aligned}
$$

Since $\left\|\nabla \tilde{\psi}_q\right\|_{\infty} < \infty$, we only need to show that $\sqrt{N} E \sup_{q,\theta} |q_{N,V;\theta}(q) - q|$ and

$\sqrt{N}E\sup_{q,\theta}|q_{N,U;\theta}(q)-q|$ are finite. Note that

$$\begin{aligned} q_{N,V;\theta}(q)-q &= F_{V;\theta}\left(F^{-1}_{N,V;\theta}(q)\right)-F_{N,V;\theta}\left(F^{-1}_{N,V;\theta}(q)\right) \\ &= (\mu_X-\mu_{X_N})\left(F_\varepsilon\left(F^{-1}_{N,V;\theta}(q)-h(X;\theta)\right)\right) \\ \Rightarrow \sqrt{N}E\sup_{q,\theta}|q_{N,V;\theta}(q)-q| &\leq \sqrt{N}E\sup_{v,\theta}|(\mu_X-\mu_{X_N})(F_\varepsilon(v-h(X;\theta)))| \\ &< \infty \end{aligned}$$

since $F_\varepsilon(v-h(X;\theta))$ indexed by v and θ is μ_X-Donsker. An identical argument implies that $\sqrt{N}E\sup_{q,\theta}|q_{N,U;\theta}(q)-q|$ is finite.

To bound the absolute value of A_2, let

$$\rho_{\eta,N;\theta}(v,z)=f_\eta\left(F^{-1}_{N,U;\theta}(F_{N,V;\theta}(v))-g(z;\theta)\right)$$

$$\begin{aligned} &A_2 = \\ &\left(\left[\int_0^1-\int_\delta^{1-\delta}\right]\frac{\int\Psi(x_1,x_2,z)\,\phi_{\varepsilon,N}(q,x_1;\theta)\,\phi_{\varepsilon,N}(q,x_2;\theta)\,\phi_{\eta,N}(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_{\varepsilon,N}(q,x_1;\theta)\,\phi_{\varepsilon,N}(q,x_2;\theta)\,\phi_{\eta,N}(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}\right. \\ &\left.-\frac{\int\Psi(x_1,x_2,z)\,\phi_{\varepsilon,N}(q,x_1;\theta)\,\phi_{\varepsilon,N}(q,x_2;\theta)\,\phi_{\eta,N}(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}{\int\phi_{\varepsilon,N}(q,x_1;\theta)\,\phi_{\varepsilon,N}(q,x_2;\theta)\,\phi_{\eta,N}(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}dq\right)= \\ &\left(\int_{-\infty}^{\infty}-\int_{F^{-1}_{N,V;\theta}(\delta)}^{F^{-1}_{N,V;\theta}(1-\delta)}\right)\int f_\varepsilon(v-h(x_1;\theta))\,(d\mu_{X_N}-d\mu_X)\,x \\ &\frac{\int\Psi(x_1,x_2,z)\,f_\varepsilon(v-h(x_1;\theta))\,f_\varepsilon(v-h(x_2;\theta))\,\rho_{\eta,N;\theta}(v,z)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int f_\varepsilon(v-h(x_1;\theta))\,f_\varepsilon(v-h(x_2;\theta))\,\rho_{\eta,N;\theta}(v,z)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dv \\ &+\left(\int_{-\infty}^{\infty}-\int_{F^{-1}_{N,V;\theta}(\delta)}^{F^{-1}_{N,V;\theta}(1-\delta)}\right)\frac{\int f_\varepsilon(v-h(x_1;\theta))\,d\mu_X}{\int f_\varepsilon(v-h(x_1;\theta))\,d\mu_X}x \\ &\frac{\int\Psi(x_1,x_2,z)\,f_\varepsilon(v-h(x_1;\theta))\,f_\varepsilon(v-h(x_2;\theta))\,\rho_{\eta,N;\theta}(v,z)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int f_\varepsilon(v-h(x_1;\theta))\,\rho_{\eta,N;\theta}(v,z)\,d\mu_{X_1}d\mu_Z}dv \\ &-\frac{\int\Psi(x_1,x_2,z)\,f_\varepsilon(v-u_\theta(x_1))\,f_\varepsilon(v-h(x_2;\theta))\,\rho_{\eta,N;\theta}(v,z)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}{\int f_\varepsilon(v-h(x_1;\theta))\,\rho_{\eta,N;\theta}(v,z)\,d\mu_{X_{N,1}}d\mu_{Z_N}}dv= \\ &T_1+T_2-T_3 \end{aligned}$$

where the equality follows from the change of variable $v = F_{N,V;\theta}^{-1}(q)$.

Note that

$$\begin{aligned}
\sqrt{N}|T_1| &\leq \sqrt{N}\|\Psi\|_\infty \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)}\right) \left|\int f_\varepsilon(v - h(x_1;\theta))(d\mu_{X_N} - d\mu_X)\right| dv \\
&\leq \sqrt{N}\|\Psi\|_\infty \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)}\right) \left|\int f_\varepsilon(v - h(x_1;\theta_0))(d\mu_{X_N} - d\mu_X)\right| dv \\
&\quad + \sqrt{N}\|\Psi\|_\infty \int_{-\infty}^{\infty} \left|\int [f_\varepsilon(v - h(x_1;\theta)) - f_\varepsilon(v - h(x_1;\theta_0))](d\mu_{X_N} - d\mu_X)\right| dv.
\end{aligned}$$

Hence, $E\sqrt{N}\sup_{\|\theta-\theta_0\|\leq r_N}(|T_1|)|_{\delta=\delta_N} \to 0$ for any sequence of positive δ_N and r_N decreasing to 0 by Lemmas C.13 and C.18.

Now we bound $T_2 - T_3$ by splitting it into three terms, and bounding them,

$$\begin{aligned}
T_2 - T_3 &= \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)}\right) \\
&\frac{\int \Psi(x_1,x_2,z) f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int f_\varepsilon(v - h(x_1;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_1} d\mu_Z} \\
&- \frac{\int \Psi(x_1,x_2,z) f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{X_2} d\mu_{Z_N}}{\int f_\varepsilon(v - h(x_1;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{Z_N}} \\
&- \frac{\int \Psi(x_1,x_2,z) f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\left(\mu_{X_{N,2}} - \mu_{X_2}\right) d\mu_{Z_N}}{\int f_\varepsilon(v - h(x_1;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{Z_N}} \\
&= R_1 + R_2 - R_3,
\end{aligned}$$

where

$$\begin{aligned}
&R_3 = \\
&\sum a_k \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)}\right) \frac{\int \Psi_1^k(x_1)\Psi_z^k(z) f_\varepsilon(v - h(x_1;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{Z_N}}{\int f_\varepsilon(v - h(x_1;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{Z_N}} \times \\
&\int \Psi_2^k(x_2) f_\varepsilon(v - h(x_2;\theta)) d\left(\mu_{X_{N,2}} - \mu_{X_2}\right) d\mu_{X_2} dv \\
\Rightarrow\ &\sqrt{N}|R_3| \leq \sum_{k=1}^{K} a_k \left\|\Psi_1^k\right\|_\infty \left\|\Psi_z^k\right\|_\infty \sqrt{N} \\
&\left[\begin{array}{l}
\left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)}\right) \left|\int \Psi_2^k(x) f_\varepsilon(v - h(x;\theta_0)) \left(d\mu_{X_{N,2}} - d\mu_{X_2}\right)\right| dv \\
+ \int_{-\infty}^{\infty} \left|\int \Psi_2^k(x) [f_\varepsilon(v - h(x;\theta)) - f_\varepsilon(v - h(x;\theta_0))] \left(d\mu_{X_{N,2}} - d\mu_{X_2}\right)\right| dv
\end{array}\right].
\end{aligned}$$

Hence, $E\sqrt{N}\sup_{\|\theta-\theta_0\|\leq r_N}(|R_3|)|_{\delta=\delta_N} \to 0$ for any sequence of positive δ_N and r_N decreasing

to 0 by Lemmas C.13 and C.18.

We will now break $R_1 + R_2$ into three terms

$$
\begin{aligned}
& \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \\
& \frac{\int \Psi(x_1, x_2, z) f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int f_\varepsilon(v - h(x_1;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_1} d\mu_Z} \\
& - \frac{\int \Psi(x_1, x_2, z) f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{X_2} d\mu_{Z_N}}{\int f_\varepsilon(v - h(x_1;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{Z_N}} dv \\
= & \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \int f_\varepsilon(v - h(x;\theta)) d\mu_X \\
& \frac{\int \Psi(x_1, x_2, z) f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dv \\
& - \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \int f_\varepsilon(v - h(x;\theta)) d\mu_X \\
& \frac{\int \Psi(x_1, x_2, z) f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{X_2} d\mu_{Z_N}}{\int f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{X_2} d\mu_{Z_N}} dv \\
= & \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \int f_\varepsilon(v - h(x;\theta)) d\mu_X \\
& \frac{\int \Psi(x_1, x_2, z) f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dv \\
& - \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \int f_\varepsilon(v - h(x;\theta)) d\mu_{X_N} \\
& \frac{\int \Psi(x_1, x_2, z) f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{X_2} d\mu_{Z_N}}{\int f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{X_2} d\mu_{Z_N}} dv \\
& + \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \int f_\varepsilon(v - h(x;\theta)) (d\mu_{X_N} - d\mu_X) \\
& \frac{\int \Psi(x_1, x_2, z) f_\varepsilon(u - u_\theta(x_1)) f_\varepsilon(u - u_\theta(x_2)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{X_2} d\mu_{Z_N}}{\int f_\varepsilon(v - h(x_1;\theta)) f_\varepsilon(v - h(x_2;\theta)) \rho_{\eta,N;\theta}(v,z) d\mu_{X_{N,1}} d\mu_{X_2} d\mu_{Z_N}} dv \\
= & M_1 - M_2 + M_3,
\end{aligned}
$$

where

$$
\begin{aligned}
\sqrt{N} |M_3| \leq & \sqrt{N} \|\Psi\|_\infty \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \left| \int f_\varepsilon(v - h(x;\theta_0)) (d\mu_{X_N} - d\mu_X) \right| dv \\
& + \sqrt{N} \|\Psi\|_\infty \int_{-\infty}^{\infty} \left| \int [f_\varepsilon(v - h(x;\theta)) - f_\varepsilon(v - h(x;\theta_0))] (d\mu_{X_N} - d\mu_X) \right| dv,
\end{aligned}
$$

so $E \sup_{\|\theta-\theta_0\| \leq r_N} \sqrt{N} \left(|M_3| \right) |_{\delta=\delta_N} \rightarrow 0$ for our sequences r_N, δ_N by the same argument applied to T_1.

We rewrite $M_1 - M_2$ as

$$
\begin{aligned}
&= \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \int f_\varepsilon \left(v - h \left(x; \theta \right) \right) d\mu_X \\
&\left(\frac{\int \Psi \left(x_1, x_2, z \right) f_\varepsilon \left(v - h \left(x_1; \theta \right) \right) f_\varepsilon \left(v - h \left(x_2; \theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int f_\varepsilon \left(v - h \left(x_1; \theta \right) \right) f_\varepsilon \left(v - u_\theta \left(x_2 \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \right. \\
&\left. - \frac{\int \Psi \left(x_1, x_2, z \right) f_\varepsilon \left(v - h \left(x_1; \theta \right) \right) f_\varepsilon \left(v - h \left(x_2; \theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_{N,1}} d\mu_{X_2} d\mu_{Z_N}}{\int f_\varepsilon \left(v - h \left(x_1; \theta \right) \right) f_\varepsilon \left(v - h \left(x_2; \theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_1} d\mu_{X_2} d\mu_{Z_N}} \right) dv \\
&= \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \\
&\left(\frac{\int \Psi \left(x_1, x_2, z \right) f_\varepsilon \left(v - h \left(x_1; \theta \right) \right) f_\varepsilon \left(v - h \left(x_2; \theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int f_\varepsilon \left(v - h \left(x_2; \theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_2} d\mu_Z} \right. \\
&\left. - \frac{\int \Psi \left(x_1, x_2, z \right) f_\varepsilon \left(u - u_\theta \left(x_1 \right) \right) f_\varepsilon \left(u - u_\theta \left(x_2 \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_1} d\mu_{X_2} d\mu_{Z_N}}{\int f_\varepsilon \left(u - u_\theta \left(x_2 \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_2} d\mu_{Z_N}} \right) dv \\
&+ \int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \\
&\frac{\int \Psi \left(x_1, x_2, z \right) f_\varepsilon \left(v - h \left(x_1; \theta \right) \right) f_\varepsilon \left(v - h \left(x_2; \theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) \left(d\mu_{X_{N,1}} - d\mu_{X_1} \right) d\mu_{X_2} d\mu_{Z_N}}{\int f_\varepsilon \left(v - h \left(x_2; \theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_2} d\mu_{Z_N}} \\
&dv \\
&= N_1 + N_2,
\end{aligned}
$$

where

$$
\begin{aligned}
&\sqrt{N} \left| N_2 \right| \leq \sqrt{N} \sum_{k=1}^{K} a_k \left\| \Psi_2^k \right\|_\infty \left\| \Psi_z^k \right\|_\infty \\
&\left[\begin{array}{c} \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \left| \int \Psi_1^k \left(x \right) f_\varepsilon \left(v - h \left(x; \theta_0 \right) \right) \left(d\mu_{X_{N,1}} - d\mu_{X_1} \right) \right| dv \\ \int_{-\infty}^{\infty} \left| \int \Psi_1^k \left(x \right) \left[f_\varepsilon \left(v - h \left(x; \theta \right) \right) - f_\varepsilon \left(v - h \left(x; \theta_0 \right) \right) \right] \left(d\mu_{X_{N,1}} - d\mu_{X_1} \right) \right| dv \end{array} \right]
\end{aligned}
$$

so $E \sup_{\|\theta-\theta_0\| \leq r_N} \sqrt{N} \left(|N_2| \right) |_{\delta=\delta_N} \rightarrow 0$ by the same argument bounding T_1. We now split N_1 into three pieces

$$
\begin{aligned}
N_1 &= \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \int f_\varepsilon \left(v - h\left(x;\theta \right) \right) d\mu_X \\
&\frac{\int \Psi \left(x_1, x_2, z \right) f_\varepsilon \left(v - h\left(x_1;\theta \right) \right) f_\varepsilon \left(v - h\left(x_2;\theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int f_\varepsilon \left(v - h\left(x_1;\theta \right) \right) f_\varepsilon \left(v - h\left(x_2;\theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dv \\
&- \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \int f_\varepsilon \left(v - h\left(x;\theta \right) \right) d\mu_X \\
&\frac{\int \Psi \left(x_1, x_2, z \right) f_\varepsilon \left(v - h\left(x_1;\theta \right) \right) f_\varepsilon \left(v - h\left(x_2;\theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_1} d\mu_{X_2} d\mu_{Z_N}}{\int f_\varepsilon \left(v - h\left(x_1;\theta \right) \right) f_\varepsilon \left(v - h\left(x_2;\theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_1} d\mu_{X_2} d\mu_{Z_N}} dv \\
&+ \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \\
&\frac{\int \Psi \left(x_1, x_2, z \right) f_\varepsilon \left(v - h\left(x_1;\theta \right) \right) f_\varepsilon \left(v - h\left(x_2;\theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) \left(d\mu_{X_1} - d\mu_{X_{N,1}} \right) d\mu_{X_2} d\mu_{Z_N}}{\int f_\varepsilon \left(v - h\left(x_2;\theta \right) \right) \rho_{\eta,N;\theta} \left(v, z \right) d\mu_{X_2} d\mu_{Z_N}} dv \\
&= O_1 + O_2 + O_3
\end{aligned}
$$

where

$$
\begin{aligned}
&\sqrt{N} \left| O_3 \right| \leq \sqrt{N} \sum_{k=1}^{K} a_k \left\| \Psi_2^k \right\|_\infty \left\| \Psi_z^k \right\|_\infty \\
&\left[\begin{array}{c} \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \left| \int \Psi_1 \left(x \right) f_\varepsilon \left(v - h\left(x;\theta_0 \right) \right) \left(d\mu_{X_1} - d\mu_{X_{N,1}} \right) \right| dv \\ + \int_{-\infty}^{\infty} \left| \int \Psi_1 \left(x \right) \left(f_\varepsilon \left(v - h\left(x;\theta \right) \right) - f_\varepsilon \left(v - h\left(x;\theta_0 \right) \right) \right) \left(d\mu_{X_1} - d\mu_{X_{N,1}} \right) \right| dv \end{array} \right],
\end{aligned}
$$

$E \sup_{\left\| \theta - \theta_0 \right\| \leq r_N} \sqrt{N} \left(\left| O_3 \right| \right) |_{\delta = \delta_N} \to 0$ by the same argument bounding T_1.

Now we rewrite $O_1 + O_2$ by substituting $\rho_{\eta,N;\theta} \left(v, z \right) = f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1} \left(v \right) \right) - g\left(z;\theta \right) \right)$.Let

$$
m(v,z) = \int \Psi \left(x_1, x_2, z \right) f_\varepsilon \left(v - h\left(x_1;\theta \right) \right) f_\varepsilon \left(v - h\left(x_2;\theta \right) \right) d\mu_{X_1} d\mu_{X_2},
$$

and

$$
n(v) = n(v,z) = \int f_\varepsilon \left(v - h\left(x_1;\theta \right) \right) f_\varepsilon \left(v - h\left(x_2;\theta \right) \right) d\mu_{X_1} d\mu_{X_2}.
$$

$$
\begin{aligned}
&O_1 + O_2 = \\
&\left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \\
&\left(\frac{\int m(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_Z}{\int n(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_Z} \right. \\
&\left. - \frac{\int m(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_{Z_N}}{\int n(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_{Z_N}} \right) \int f_\varepsilon (v - h(x;\theta)) \, d\mu_{X_N} dv \\
&+ \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \left(\frac{\int m(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_Z}{\int n(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_Z} \right. \\
&\left. - \frac{\int m(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_{Z_N}}{\int n(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_{Z_N}} \right) \int f_\varepsilon (v - h(x;\theta)) \left(d\mu_X - d\mu_{X_N} \right) dv \\
&= P_1 + P_2,
\end{aligned}
$$

where

$$
\sqrt{N} |P_2| \leq \sqrt{N} 2 \|\Psi\|_\infty \left[\begin{array}{c} \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)} \right) \left| \int f_\varepsilon (v - h(x;\theta_0)) \left(d\mu_X - d\mu_{X_N} \right) \right| dv \\ + \int_{-\infty}^{\infty} \left| \int \left[f_\varepsilon (v - h(x;\theta)) - f_\varepsilon (v - h(x;\theta_0)) \right] \left(d\mu_X - d\mu_{X_N} \right) \right| dv \end{array} \right],
$$

so $E \sup_{\|\theta - \theta_0\| \leq r_N} \sqrt{N} \left(|P_2| \right)|_{\delta = \delta_N} \to 0$ by the same argument bounding T_1.

By change of variables,

$$
\begin{aligned}
& q = F_{N,V;\theta}^{-1}(v) \\
\Rightarrow \quad & dq = \int f_\varepsilon (v - h(x;\theta)) \, d\mu_{X_N} dv \text{ and } F_{N,V;\theta}^{-1}(q) = v
\end{aligned}
$$

followed by a change of variables

$$
q = \int F_\eta (u - g(z;\theta)) \, d\mu_{Z_N}
$$

we rewrite P_1 as

$$
\begin{aligned}
& \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta)}^{F_{N,V;\theta}^{-1}(1-\delta)}\right) \left(\frac{\int m(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_Z}{\int n(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_Z} \right. \\
& \left. - \frac{\int m(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_{Z_N}}{\int n(v,z) f_\eta \left(F_{N,U;\theta}^{-1} \left(F_{N,V;\theta}^{-1}(v) \right) - g(z;\theta) \right) d\mu_{Z_N}} \right) \int f_\varepsilon (v - h(x;\theta)) \, d\mu_{X_N} dv \\
= & \left(\int_0^1 - \int_\delta^{1-\delta}\right) \left(\frac{\int \Psi(x_1,x_2,z) \, \phi_{\varepsilon,N}(q,x_1;\theta) \, \phi_{\varepsilon,N}(q,x_2;\theta) \, \phi_{\eta,N}(q,x_1;\theta) \, d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon,N}(q,x_1;\theta) \, \phi_{\varepsilon,N}(q,x_2;\theta) \, \phi_{\eta,N}(q,x_1;\theta) \, d\mu_{X_1} d\mu_{X_2} d\mu_Z} \right. \\
& \left. - \frac{\int \Psi(x_1,x_2,z) \, \phi_{\varepsilon,N}(q,x_1;\theta) \, \phi_{\varepsilon,N}(q,x_2;\theta) \, \phi_{\eta,N}(q,x_1;\theta) \, d\mu_{X_1} d\mu_{X_2} d\mu_{Z_N}}{\int \phi_{\varepsilon,N}(q,x_1;\theta) \, \phi_{\varepsilon,N}(q,x_2;\theta) \, \phi_{\eta,N}(q,x_1;\theta) \, d\mu_{X_1} d\mu_{X_2} d\mu_{Z_N}} \right) dq \\
= & \left(\int_{-\infty}^{\infty} - \int_{F_{N,U;\theta}^{-1}(\delta)}^{F_{N,U;\theta}^{-1}(1-\delta)}\right) \left(\frac{\int m(F_{N,V;\theta}^{-1} \left(F_{N,U;\theta}^{-1}(u) \right), z) f_\eta (u - g(z;\theta)) \, d\mu_Z}{\int n(F_{N,V;\theta}^{-1} \left(F_{N,U;\theta}^{-1}(u) \right), z) f_\eta (u - g(z;\theta)) \, d\mu_Z} \right. \\
& \left. - \frac{\int m(F_{N,V;\theta}^{-1} \left(F_{N,U;\theta}^{-1}(u) \right), z) f_\eta (u - g(z;\theta)) \, d\mu_{Z_N}}{\int n(F_{N,V;\theta}^{-1} \left(F_{N,U;\theta}^{-1}(u) \right), z) f_\eta (u - g(z;\theta)) \, d\mu_{Z_N}} \right) \\
& \int f_\eta (u - g(z;\theta)) \, d\mu_{Z_N} du \\
= & \left(\int_{-\infty}^{\infty} - \int_{F_{N,U;\theta}^{-1}(\delta)}^{F_{N,U;\theta}^{-1}(1-\delta)}\right) \frac{\int m(F_{N,V;\theta}^{-1} \left(F_{N,U;\theta}^{-1}(u) \right), z) f_\eta (u - g(z;\theta)) \, d\mu_{X_1} d\mu_{X_2} d\mu_Z}{n(v)} du \\
& + \left(\int_{-\infty}^{\infty} - \int_{F_{N,U;\theta}^{-1}(\delta)}^{F_{N,U;\theta}^{-1}(1-\delta)}\right) \frac{\int m(F_{N,V;\theta}^{-1} \left(F_{N,U;\theta}^{-1}(u) \right), z) f_\eta (u - g(z;\theta)) \, d\mu_Z}{\int n(F_{N,V;\theta}^{-1} \left(F_{N,U;\theta}^{-1}(u) \right), z) f_\eta (u - g(z;\theta)) \, d\mu_Z} \\
& \times \int f_\eta (u - g(z;\theta)) \left(d\mu_{Z_N} - d\mu_Z \right) du \\
& - \left(\int_{-\infty}^{\infty} - \int_{F_{N,U;\theta}^{-1}(\delta)}^{F_{N,U;\theta}^{-1}(1-\delta)}\right) \frac{\int m(F_{N,V;\theta}^{-1} \left(F_{N,U;\theta}^{-1}(u) \right), z) f_\eta (u - g(z;\theta)) \, d\mu_{Z_N}}{n(v)} du \\
= & \; Q_1 + Q_2 + Q_3
\end{aligned}
$$

where $\sqrt{N}\,|Q_2|$ is not greater than

$$
\sqrt{N}\,\|\Psi\|_\infty \begin{bmatrix} \left(\int_{-\infty}^{\infty} - \int_{F_{N,U;\theta}^{-1}(\delta)}^{F_{N,U;\theta}^{-1}(1-\delta)}\right) \left| \int f_\eta (u - g(z;\theta_0)) \left(d\mu_{Z_N} - d\mu_Z \right) \right| du \\ \int_{-\infty}^{\infty} \left| \int \left[f_\eta (u - g(z;\theta)) - f_\eta (u - g(z;\theta_0)) \right] \left(d\mu_{Z_N} - d\mu_Z \right) \right| du \end{bmatrix}, \text{ and so}
$$

$$
E \sup_{\|\theta - \theta_0\| \le r_N} \sqrt{N} \left(|Q_2| \right)|_{\delta = \delta_N} \to 0
$$

by the same argument bounding T_1.. Finally, to bound $Q_1 + Q_3$, note that

$$
\begin{aligned}
& \sqrt{N}\left|Q_1+Q_3\right| \\
\leq\ & \sqrt{N}\sum a_k \left\|\Psi_1\right\|_\infty \left\|\Psi_2\right\|_\infty \\
& \left[\begin{array}{l}
\left(\int_{-\infty}^{\infty} - \int_{F^{-1}_{N,U;\theta}(\delta)}^{F^{-1}_{N,U;\theta}(1-\delta)}\right) \int_{-\infty}^{\infty} \left|\int \Psi_z(z) f_\eta(u-g(z;\theta_0))\left(d\mu_Z - d\mu_{Z_N}\right)\right| du \\
+\int_{-\infty}^{\infty} \left|\int \Psi_z(z)\left[f_\eta(u-g(z;\theta)) - f_\eta(u-g(z;\theta_0))\right]\left(d\mu_Z - d\mu_{Z_N}\right)\right| du
\end{array}\right],
\end{aligned}
$$

and so $E\sup_{\|\theta-\theta_0\|\leq r_N}\sqrt{N}\left(\left|Q_1+Q_3\right|\right)|_{\delta=\delta_N}\to 0$ by the same argument bounding T_1.By the triangle inequality, the expression

$$
\begin{aligned}
& \sqrt{N}\left(\int_0^1 - \int_\delta^{1-\delta}\right) \\
& \frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta)\,\phi_\varepsilon(q,x_2;\theta)\,\phi_\eta(q,z;\theta)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z} \\
& -\frac{\int \Psi(x_1,x_2,z)\,\phi_{\varepsilon,N}(q,x_1;\theta)\,\phi_{\varepsilon,N}(q,x_2;\theta)\,\phi_{\eta,N}(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}}{\int \phi_{\varepsilon,N}(q,x_1;\theta)\,\phi_{\varepsilon,N}(q,x_2;\theta)\,\phi_{\eta,N}(q,z;\theta)\,d\mu_{X_{N,1}}d\mu_{X_{N,2}}d\mu_{Z_N}} dq
\end{aligned}
$$

has abolute value not greater than

$$
\begin{aligned}
& \sqrt{N}\left|A_1\right| + \sqrt{N}\left|T_1\right| + \sqrt{N}\left|R_3\right| + \sqrt{N}\left|M_3\right| \\
& + \sqrt{N}\left|N_2\right| + \sqrt{N}\left|O_3\right| + \sqrt{N}\left|P_2\right| + \sqrt{N}\left|Q_1+Q_3\right| + \sqrt{N}\left|Q_2\right|
\end{aligned}
$$

so $\sqrt{N}E\sup_{\|\theta-\theta_0\|\leq r_N}$
$\left|\left(\psi\left[\mu_X,\mu_Z\right](\theta) - \psi\left[\mu_{X_N},\mu_{Z_N}\right](\theta)\right) - \left(\psi^{\delta_N}\left[\mu_X,\mu_Z\right](\theta) - \psi^{\delta_N}\left[\mu_{X_N},\mu_{Z_N}\right](\theta)\right)\right| = o(1)$
as desired. ■

Preliminaries for Proposition C.3

Lemma C.14 *If $C_\varepsilon = \int_{-\infty}^{\infty}\left|f_\varepsilon'(v)\right| dv$, $\mu_X h_{LC}(X)^2$ and $\Psi(X)$ are bounded, then*

$$
E\sup_\theta \int \left|R_{i+1}(\theta,v)\right| T_i^c(\theta)\, T_{i+1}(\theta)\, dv \leq \sqrt{N}\left\|\Psi\right\|_\infty 6\delta_{i+1}\frac{\delta_i}{V_i} C_\varepsilon dv \mu_X h_{LC}(X)^2
$$

and

$$
E\sup_\theta \int \left|\left(R_i(\theta,v)\right) T_i(\theta)\, T_{i+1}^c(\theta)\right| dv \leq 6\left\|\Psi\right\|_\infty \delta_i \frac{\delta_{i+1}}{V_{i+1}} C_\varepsilon \mu_X h_{LC}(X)^2 .
$$

Proof. We first show that

$$E\sup_{\theta}\int\left|R_{i+1}\left(\theta,v\right)\right|T_{i}^{c}\left(\theta\right)T_{i+1}\left(\theta\right)dv\leq\sqrt{N}6\left\|\Psi\right\|_{\infty}\delta_{i+1}\frac{\delta_{i}}{V_{i}}C_{\varepsilon}dv\mu_{X}h_{LC}\left(X\right)^{2}.$$

Note that

$$\begin{aligned}
& E\sup_{\theta}\int\left|R_{i+1}\left(\theta,v\right)\right|T_{i}^{c}\left(\theta\right)T_{i+1}\left(\theta\right)dv\\
\leq\ & E\sup_{\theta}\int\left|R_{i+1}\left(\theta,v\right)\right|T_{i}^{c}\left(\theta\right)dv\\
\leq\ & E\sup_{\theta}\frac{1}{\sqrt{N}}\int\sum_{i=1}^{n}\\
& \left|\Psi\left(X\right)\left\{f_{\varepsilon}\left(v-h\left(X_{j};\theta\right)\right)-f_{\varepsilon}\left(v-h\left(X_{j};B\left(\theta;i+1\right)\right)\right)\right\}\right|\left\{h_{LC}\left(X_{j}\right)>\frac{V_{i}}{2\delta_{i}}\right\}dv\\
& +\sqrt{N}\mu_{X}\int\\
& \left(\left|\Psi\left(X\right)f_{\varepsilon}\left(v-h\left(X_{j};\theta\right)\right)-f_{\varepsilon}\left(v-h\left(X_{j};B\left(\theta;i+1\right)\right)\right)\right|\left\{h_{LC}\left(X\right)>\frac{V_{i}}{2\delta_{i}}\right\}\right)dv
\end{aligned}$$

where the last inequality is a consequence of the triangle inequality. The second term is not greater than

$$\begin{aligned}
& \sqrt{N}\left\|\Psi\right\|_{\infty}\delta_{i+1}\int_{-\infty}^{\infty}\left|f_{\varepsilon}'\left(v\right)\right|dv\mu_{X}h_{LC}\left(X\right)\left\{h_{LC}\left(X\right)>\frac{V_{i}}{2\delta_{i}}\right\}\\
\leq\ & 2\left\|\Psi\right\|_{\infty}\sqrt{N}\frac{\delta_{i}\delta_{i+1}}{V_{i}}C_{\varepsilon}\mu_{X}h_{LC}\left(X\right)^{2}
\end{aligned}$$

and the first term is not greater than

$$\begin{aligned}
& E\sup_{\theta}\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\left\{h_{LC}\left(X_{j}\right)>\frac{V_{i}}{2\delta_{i}}\right\}\\
& \int\left\|\Psi\right\|_{\infty}\left|\left\{f_{\varepsilon}\left(v-h\left(X_{j};\theta\right)\right)-f_{\varepsilon}\left(v-h\left(X_{j};B\left(\theta;i+1\right)\right)\right)\right\}\right|dv\\
\leq\ & \left\|\Psi\right\|_{\infty}E\sup_{\theta}\frac{1}{\sqrt{N}}C_{\varepsilon}\sum_{i=1}^{n}\left\{h_{LC}\left(X_{j}\right)>\frac{V_{i}}{2\delta_{i}}\right\}\left|h\left(X_{j};\theta\right)-h\left(X_{j};B\left(\theta;i+1\right)\right)\right|\\
\leq\ & \left\|\Psi\right\|_{\infty}E\sup_{\theta}2\delta_{i+1}\sqrt{N}C_{\varepsilon}\mu_{X_{N}}\left|h_{LC}\left(X\right)\left\{h_{LC}\left(X\right)>\frac{V_{i}}{2\delta_{i}}\right\}\right|\\
\leq\ & 4\left\|\Psi\right\|_{\infty}\delta_{i}\frac{\delta_{i+1}}{V_{i}}\sqrt{N}C_{\varepsilon}\mu_{X}\left|h_{LC}\left(X\right)^{2}\right|.
\end{aligned}$$

By an identical argument,

$$\begin{aligned} & E\sup_{\theta}\int\left|\left(R_i\left(\theta,v\right)\right)T_i\left(\theta\right)T_{i+1}^{c}\left(\theta\right)\right|dv \\ \leq\ & 6\left\|\Psi\right\|_{\infty}\delta_i\frac{\delta_{i+1}}{V_{i+1}}C_{\varepsilon}\mu_X h_{LC}\left(X\right)^2. \end{aligned}$$

■

Lemma C.15 *Let $\mathcal{E}\left(x\right)=2\frac{\exp(x)-1-x}{x^2}$, and let $N\left(\delta_{i+1}\right)$ be the δ_{i+1} covering number of Θ in the Euclidean metric. If Assumption C.3 is satisfied, then*

$$\begin{aligned} & E\sup_{\theta}\int\left|R_i\left(\theta,v\right)-R_{i+1}\left(\theta,v\right)\right|T_i\left(\theta\right)T_{i+1}\left(\theta\right)dv \\ \leq\ & \delta_i\sqrt{\log\left(2N\left(\delta_{i+1}\right)\right)}\left(1+12C_{\varepsilon}^2\left\|\Psi\right\|_{\infty}^2\mu_X\left(h_{LC}\left(X\right)^2\right)\right. \\ & \left.+18C_{\varepsilon}^4\left\|\Psi\right\|_{\infty}^4\mu_X\left(h_{LC}\left(X\right)^2\right)\mathcal{E}\left(6\right)\right)+2\left\|\Psi\right\|_{\infty}K\delta_i. \end{aligned}$$

for some constant K. Hence, if $N\left(\delta_{i+1}\right)$ is finite, there is a $K_1<\infty$ such that

$$E\sup_{\theta}\int\left|R_i\left(\theta,v\right)-R_{i+1}\left(\theta,v\right)\right|T_i\left(\theta\right)T_{i+1}\left(\theta\right)dv<K_1\delta_i\sqrt{\log\left(N\left(\delta_i\right)\right)}.$$

Proof. To simplify notation, let $\Delta_i^f\left(X;\theta,v\right)=f_{\varepsilon}\left(v-h\left(X;B\left(\theta;i+1\right)\right)\right)-f_{\varepsilon}\left(v-h\left(X;B\left(\theta;i\right)\right)\right)$ and $\Delta_i^h\left(X;\theta\right)=h\left(X;B\left(\theta;i+1\right)\right)-h\left(X;B\left(\theta;i\right)\right)$. By the triangle inequality,

$$\begin{aligned} & E\sup_{\theta}\int\left|\left(R_i\left(\theta,v\right)-R_{i+1}\left(\theta,v\right)\right)\right|T_i\left(\theta\right)T_{i+1}\left(\theta\right)dv \\ \leq\ & E\sup_{\theta}\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{h_{LC}\left(X\right)\leq\frac{V_i}{2\delta_i}\right\}dv \\ \leq\ & E\sup_{\theta}\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv \\ \leq\ & E\sup_{\theta}\left[\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right. \\ & \left.-E\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right] \\ & +\sup_{\theta}E\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv \quad \text{(C.36)} \end{aligned}$$

We now bound the two terms individually.

The first term in equation (C.36) is bounded by using the bound on its moment generating function for a fixed θ (derived in Lemma C.16) and the concentration inequality of Theorem 2 in Boucheron *et al.* (2003).

By Jensen's inequality, note that for any λ_i,

$$\begin{aligned}
& \exp\left(\lambda_i\left(E\sup_{\theta}\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right.\right. \\
& \left.\left.-E\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right)\right) \\
\leq\ & E\exp\left(\lambda_i\left(\sup_{\theta}\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right.\right. \\
& \left.\left.-E\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right)\right).
\end{aligned}$$

Note that $B\left(\theta;i+1\right)$ takes on at most $N\left(\delta_{i+1}\right)$. Since the expectation of a maximum of finitely many nonnegative random variables is less than the sum of their expectations, the expression above is no greater than

$$\begin{aligned}
& \sum_{\theta\in ImB(\theta;i+1)}E\exp\left(\lambda_i\left|\int\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right.\right. \\
& \left.\left.-E\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right|\right) \\
\leq\ & \sum_{\theta\in ImB(\theta;i+1)}E\exp\left(\lambda_i\left(\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right.\right. \\
& \left.\left.-E\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right)\right) \\
& +E\exp\left(-\lambda_i\left(\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right.\right. \\
& \left.\left.-E\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right)\right)
\end{aligned}$$

Lemma C.16 implies this is not greater than

$$\begin{aligned}
& 2N\left(\delta_{i+1}\right)\max_{\theta\in\Theta}\exp \\
& \left(C_{\varepsilon}^{2}\left\|\Psi\right\|_{\infty}^{2}\lambda_{i}^{2}12\mu_{X}\left(\Delta_{i}^{h}\left(X;\theta\right)\right)^{2}+\frac{18}{n}C_{\varepsilon}^{4}\left\|\Psi\right\|_{\infty}^{4}\lambda_{i}^{4}V_{i}^{2}E\left(\left|\Delta_{i}^{h}\left(X_{j};\theta\right)\right|^{2}\right)\mathcal{E}\left(\frac{6\lambda_{i}^{2}}{n}V_{i}^{2}\right)\right) \\
\leq\ & 2N\left(\delta_{i+1}\right)\max_{\theta\in\Theta}\exp \\
& \left(C_{\varepsilon}^{2}\left\|\Psi\right\|_{\infty}^{2}\lambda_{i}^{2}\delta_{i}^{2}12\mu_{X}\left(h_{LC}\left(X\right)^{2}\right)+\frac{18}{n}C_{\varepsilon}^{4}\left\|\Psi\right\|_{\infty}^{4}\lambda_{i}^{4}\delta_{i}^{2}V_{i}^{2}\mu_{X}\left(h_{LC}\left(X\right)^{2}\right)\mathcal{E}\left(\frac{6\lambda_{i}^{2}}{n}V_{i}^{2}\right)\right)
\end{aligned}$$

It follows that

$$\begin{aligned}
& E\sup_{\theta}\int\left|\sqrt{N}\left(\left(\mu_{X}-\mu_{X_{N}}\right)\Psi\left(X\right)\Delta_{i}^{f}\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_{i}^{h}\left(X;\theta\right)\right|\leq V_{i}\right\}dv \\
& -E\int\left|\sqrt{N}\left(\left(\mu_{X}-\mu_{X_{N}}\right)\Psi\left(X\right)\Delta_{i}^{f}\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_{i}^{h}\left(X;\theta\right)\right|\leq V_{i}\right\}dv \\
\leq\ & \frac{\log\left(2N\left(\delta_{i+1}\right)\right)}{\lambda_{i}}+\frac{1}{\lambda_{i}}* \\
& \left(C_{\varepsilon}^{2}\left\|\Psi\right\|_{\infty}^{2}\delta_{i}^{2}\lambda_{i}^{2}12\mu_{X}\left(h_{LC}\left(X\right)\right)^{2}+\frac{18}{n}C_{\varepsilon}^{4}\left\|\Psi\right\|_{\infty}^{4}\lambda_{i}^{4}\delta_{i}^{2}V_{i}^{2}\mu_{X}\left(h_{LC}\left(X\right)^{2}\right)\mathcal{E}\left(\frac{6\lambda_{i}^{2}}{n}V_{i}^{2}\right)\right)
\end{aligned}$$

Recall that $V_i=\frac{\sqrt{N}}{\lambda_i}$ and choose $\lambda_i=\frac{\sqrt{N}}{V_i}=\frac{\sqrt{\log(2N(\delta_{i+1}))}}{\delta_i}$ which yields the upper bound

$$\delta_{i}\sqrt{\log\left(2N\left(\delta_{i+1}\right)\right)}\left(1+12C_{\varepsilon}^{2}\left\|\Psi\right\|_{\infty}^{2}\mu_{X}\left(h_{LC}\left(X\right)^{2}\right)+18C_{\varepsilon}^{4}\left\|\Psi\right\|_{\infty}^{4}\mu_{X}\left(h_{LC}\left(X\right)^{2}\right)\mathcal{E}\left(6\right)\right)$$

for the first term in equation (C.36).

We bound the second term in equation (C.36) using Lemma C.17. Note that

$$\begin{aligned}
& \sup_{\theta}E\int\left|\sqrt{N}\left(\left(\mu_{X}-\mu_{X_{N}}\right)\Psi\left(X\right)\Delta_{i}^{f}\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_{i}^{h}\left(X;\theta\right)\right|\leq V_{i}\right\}dv \\
\leq\ & \sup_{\theta}\int E\left|\sqrt{N}\left(\left(\mu_{X}-\mu_{X_{N}}\right)\Psi\left(X\right)\Delta_{i}^{f}\left(X;\theta,v\right)\right)\right|dv
\end{aligned}$$

By Jensen's inequality this is not greater than

$$\begin{aligned}
& \sup_{\theta} \int \sqrt{E\left(\sqrt{N}\left(\left(\mu_X - \mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right)^2}dv \\
= & \sup_{\theta} \int \sqrt{\mu_X \Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)^2}dv \\
\leq & \left\|\Psi\right\|_\infty \sup_{\theta} \int \sqrt{\mu_X\left(f_\varepsilon\left(v - h\left(X;B\left(\theta;i+1\right)\right)\right) - f_\varepsilon\left(v - h\left(X;B\left(\theta;i\right)\right)\right)\right)^2}dv \\
\leq & \left\|\Psi\right\|_\infty K \sup_{\theta_i \in B(\theta;i)} \left\|\theta_{i+1} - \theta_i\right\| \\
\leq & 2\left\|\Psi\right\|_\infty \delta_i K
\end{aligned}$$

for some constant $K \in (0,\infty)$. The second to last inequality follows from Lemma C.17, and the last inequality follows from the definitions of $B\left(\theta,i\right)$ and δ_i. ■

Lemma C.16 *For each $\theta \in \Theta$, and any $\lambda_i > 0$,*

$$\begin{aligned}
& E\exp\left(\pm\lambda_i\left(\int\left|\sqrt{N}\left(\left(\mu_X - \mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right| \leq V_i\right\}dv\right.\right. \\
& -E\int\left|\sqrt{N}\left(\left(\mu_X - \mu_{X_N}\right)\Psi\left(X\right)\left\{h\left(X;B\left(\theta;i\right)\right) \leq v\right\} - \left\{h\left(X;B\left(\theta;i+1\right)\right) \leq v\right\}\right)\right| \\
& \left.\left.\left\{\left|\Delta_i^h\left(X;\theta\right)\right| \leq V_i\right\}dv\right)\right) \\
\leq & \exp\left(\lambda_i^2 C_\varepsilon^2\left\|\Psi\right\|_\infty^2 12\mu_X\Delta_i^h\left(X;\theta\right)^2 + \frac{18}{n}C_\varepsilon^4\left\|\Psi\right\|_\infty^4\lambda_i^4 V_i^2 E\left(\left|\Delta_i^h\left(X;\theta\right)\right|^2\right)\mathcal{E}\left(\frac{6\lambda_i^2}{n}V_i^2\right)\right)
\end{aligned}$$

where $\Delta_i^f\left(X;\theta,v\right) = f_\varepsilon\left(v - h\left(X;B\left(\theta;i+1\right)\right)\right) - f_\varepsilon\left(v - h\left(X;B\left(\theta;i\right)\right)\right)$ *and* $\Delta_i^h\left(X;\theta\right) = h\left(X;B\left(\theta;i+1\right)\right) - h\left(X;B\left(\theta;i\right)\right)$.

Proof. Let $\left(X_{(1)},X_{(2)},...,X_{(n)}\right)$ be an independently drawn copy of $\left(X_1,X_2,...X_n\right)$, and let $\mu_X^{n,(j)}$ be the empirical measure induced by replacing X_j by $X_{(j)}$. Let

$$Z = \int\left|\sqrt{N}\left(\left(\mu_X - \mu_{X_N}\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right| \leq V_i\right\}dv$$

and

$$Z^{(j)} = \int\left|\sqrt{N}\left(\left(\mu_X - \mu_X^{n,(j)}\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right| \leq V_i\right\}dv.$$

By Theorem 2 in Boucheron *et al.* (2003), for any $0 < \theta < \frac{1}{|\lambda_i|}$

$$\log E\exp\left(\pm\lambda_i\left(Z-E\left[Z\right]\right)\right)\leq\frac{\lambda_i\theta}{1-\lambda_i\theta}\log E\exp\left(\frac{\lambda_i}{\theta}E\left[\sum_{j=1}^{n}\left(Z-Z^{(j)}\right)^2|\mu_{X_N}\right]\right)$$

so it is enough to bound the moment generating function of $E\left(\sum_{j=1}^{n}\left(Z-Z^{(j)}\right)^2|\mu_{X_N}\right)$ to prove the lemma. Note that

$$\begin{aligned}
&\left|\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right.\\
&\left.-\int\left|\sqrt{N}\left(\left(\mu_X-\mu_X^{n,(j)}\right)\Psi\left(X\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right|\\
\leq\;&\frac{1}{\sqrt{N}}\left|\int\left|\Psi\left(X_j\right)\left(\Delta_i^f\left(X_j;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X_j;\theta\right)\right|\leq V_i\right\}dv\right|\\
&+\frac{1}{\sqrt{N}}\left|\int\left|\Psi\left(X_{(j)}\right)\left(\Delta_i^f\left(X_{(j)};\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X_{(j)};\theta\right)\right|\leq V_i\right\}dv\right|\\
\leq\;&\frac{1}{\sqrt{N}}\left\|\Psi\right\|_\infty\int_{-\infty}^{\infty}\left|f_\varepsilon'\left(v\right)\right|dv\min\left(\left|\Delta_i^h\left(X_j;\theta\right)\right|,V_i\right)+\\
&\frac{1}{\sqrt{N}}\left\|\Psi\right\|_\infty\int_{-\infty}^{\infty}\left|f_\varepsilon'\left(v\right)\right|dv\left|\Delta_i^h\left(X_{(j)};\theta\right)\right|.
\end{aligned}$$

Since $\left(a+b\right)^2\leq 3a^2+3b^2$, it follows that

$$\sum_{j=1}^{n}\left(Z-Z^{(j)}\right)^2\leq\frac{3}{n}C_\varepsilon^2\left\|\Psi\right\|_\infty^2\sum_{j=1}^{n}\min\left(\left|\Delta_i^h\left(X_j;\theta\right)\right|^2,V_i^2\right)+\left(\Delta_i^h\left(X_{(j)};\theta\right)\right)^2,$$

and this upper bound has conditional expectation given μ_{X_N} of

$$\begin{aligned}
&3C_\varepsilon^2\left\|\Psi\right\|_\infty^2\mu_X\left(\Delta_i^h\left(X;\theta\right)\right)^2+3C_\varepsilon^2\left\|\Psi\right\|_\infty^2\mu_{X_N}\min\left(\left|\Delta_i^h\left(X;\theta\right)\right|^2,V_i^2\right)\\
\leq\;&6C_\varepsilon^2\left\|\Psi\right\|_\infty^2\mu_X\left(\Delta_i^h\left(X;\theta\right)\right)^2+3C_\varepsilon^2\left\|\Psi\right\|_\infty^2\left(\mu_{X_N}-\mu_X\right)\min\left(\left|\Delta_i^h\left(X;\theta\right)\right|^2,V_i^2\right)
\end{aligned}$$

Hence, the moment generating function of this conditional expectation is not greater than

$$\begin{aligned}
&\exp\left(\lambda_iC_\varepsilon^2\left\|\Psi\right\|_\infty^2 6\mu_X\left(\Delta_i^h\left(X;\theta\right)\right)^2\right)\\
&\times E\exp\left(\lambda_iC_\varepsilon^2\left\|\Psi\right\|_\infty^2 3\left(\mu_{X_N}-\mu_X\right)\min\left(\left(\Delta_i^h\left(X;\theta\right)\right)^2,V_i^2\right)\right).
\end{aligned}$$

Since μ_{X_N} is a sum of i.i.d. random variables,

$$\begin{aligned} & E\exp\left(\lambda_i C_\varepsilon^2 \|\Psi\|_\infty^2 3\left(\mu_{X_N}-\mu_X\right)\min\left(\left(\Delta_i^h\left(X;\theta\right)\right)^2, V_i^2\right)\right) \\ = & \prod_{j=1}^n E\exp\left(C_\varepsilon^2\|\Psi\|_\infty^2\frac{3\lambda_i}{n}\left(\min\left(\left|\Delta_i^h\left(X_j;\theta\right)\right|^2, V_i^2\right) - \mu_X \min\left(\left|\Delta_i^h\left(X;\theta\right)\right|^2, V_i^2\right)\right)\right). \end{aligned}$$

To bound this note that

$$\exp\left(x\right) = 1 + x + \frac{1}{2}x^2\mathcal{E}\left(x\right)$$

where $\mathcal{E}\left(x\right) = 2\frac{\exp(x)-1-x}{x^2}$ is strictly increasing. This implies that if V is a mean zero random variable bounded by a constant K,

$$E\exp\left(\lambda V\right) \leq 1 + \frac{1}{2}\lambda^2\mathcal{E}\left(\lambda K\right)E\left(V^2\right) \leq \exp\left(\frac{1}{2}\lambda^2\mathcal{E}\left(\lambda K\right)E\left(V^2\right)\right).$$

Hence,

$$\begin{aligned} & E\exp\left(C_\varepsilon^2\|\Psi\|_\infty^2\frac{3\lambda_i}{n}\left(\min\left(\left|\Delta_i^h\left(X_j;\theta\right)\right|^2, V_i^2\right)\right.\right. \\ & \left.\left. -\mu_X\min\left(\left|\Delta_i^h\left(X;\theta\right)\right|^2, V_i^2\right)\right)\right) \\ \leq & \exp\left(\frac{9}{2}C_\varepsilon^4\|\Psi\|_\infty^4\frac{\lambda_i^2}{n^2}Var\left(\min\left(\left|\Delta_i^h\left(X_j;\theta\right)\right|^2, V_i^2\right)^2\right)\mathcal{E}\left(C_\varepsilon^2\|\Psi\|_\infty^2\frac{3\lambda_i}{n}V_i^2\right)\right) \\ \leq & \exp\left(\frac{9}{2}C_\varepsilon^4\|\Psi\|_\infty^4\frac{\lambda_i^2}{n^2}V_i^2E\left(\Delta_i^h\left(X_j;\theta\right)^2\right)\mathcal{E}\left(C_\varepsilon^2\|\Psi\|_\infty^2\frac{3\lambda_i}{n}V_i^2\right)\right), \end{aligned}$$

which implies that

$$\begin{aligned} & E\exp\left(\lambda_i E\left[\sum_{j=1}^n\left(Z-Z^{(j)}\right)^2|\mu_{X_N}\right]\right) \\ \leq & \exp\left(\lambda_i C_\varepsilon^2\|\Psi\|_\infty^2 6\mu_X\left(\Delta_i^h\left(X;\theta\right)^2\right) + \right. \\ & \left.\frac{9}{2}C_\varepsilon^4\|\Psi\|_\infty^4\frac{\lambda_i^2}{n}V_i^2E\left(\Delta_i^h\left(X_j;\theta\right)\right)^2\mathcal{E}\left(C_\varepsilon^2\|\Psi\|_\infty^2\frac{3\lambda_i}{n}V_i^2\right)\right) \end{aligned}$$

By Theorem 2 of Boucheron *et al.* (2003), this implies for all $\gamma_i > 0$ and $\lambda_i \in \left(0, \frac{1}{\gamma_i}\right)$

$$
\begin{aligned}
& \log E \exp\left(\pm\lambda_i\left(\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right.\right. \\
& \left.\left.-E\int\left|\sqrt{N}\left(\left(\mu_X-\mu_{X_N}\right)\Delta_i^f\left(X;\theta,v\right)\right)\right|\left\{\left|\Delta_i^h\left(X;\theta\right)\right|\leq V_i\right\}dv\right)\right) \\
\leq\ & \frac{\lambda_i\gamma_i}{1-\lambda_i\gamma_i}\log E\exp\left(\frac{\lambda_i}{\gamma_i}E\left[\sum_{j=1}^{n}\left(Z-Z^{(j)}\right)^2|\mu_{X_N}\right]\right) \\
\leq\ & \frac{\lambda_i\gamma_i}{1-\lambda_i\gamma_i}\frac{\lambda_i}{\gamma_i}C_\varepsilon^2\left\|\Psi\right\|_\infty^2 6\mu_X\left(\left|\Delta_i^h\left(X_j;\theta\right)\right|^2\right) \\
& +\frac{\lambda_i\gamma_i}{1-\lambda_i\gamma_i}\frac{9}{2n}C_\varepsilon^4\left\|\Psi\right\|_\infty^4\left(\frac{\lambda_i}{\gamma_i}\right)^2V_i^2E\left(\left|\Delta_i^h\left(X_j;\theta\right)\right|^2\right)\mathcal{E}\left(\frac{3\lambda_i}{n\gamma_i}V_i^2\right)
\end{aligned}
$$

If we pick γ_i so that $\lambda_i\gamma_i=\frac{1}{2}$ we get the upper bound

$$
\lambda_i^2C_\varepsilon^2\left\|\Psi\right\|_\infty^2 12\mu_X\left(\Delta_i^h\left(X;\theta\right)^2\right)+\frac{18}{n}C_\varepsilon^4\lambda_i^4\left\|\Psi\right\|_\infty^4V_i^2E\left(\left|\Delta_i^h\left(X_j;\theta\right)\right|^2\right)\mathcal{E}\left(\frac{6\lambda_i^2}{n}V_i^2\right)
$$

as desired. ■

Lemma C.17 *Suppose that*

(i) for some constants C_1, $C_2>0$, *we have that* $\max\left\{\left|f_\varepsilon'\left(v\right)\right|,\sup_{\theta\in\Theta}P\left(\left|h\left(x;\theta\right)\right|>v\right)\right\}\leq C_1\exp\left(-C_2\left|v\right|\right)$

(ii) $\int h_{LC}\left(X\right)^4d\mu_X$ *is finite*

then there exists a constant such that

$$
\left|\int\sqrt{\mu_X\left(f_\varepsilon\left(v-h\left(X;\theta_1\right)\right)-f_\varepsilon\left(v-h\left(X;\theta_2\right)\right)\right)^2}dv\right|\leq K\left\|\theta_1-\theta_2\right\|.
$$

Proof. It is enough to show that the following term

$$
\sup_{\theta\in\Theta}\int\sqrt{\mu_X\left(\nabla_\theta f_\varepsilon\left(v-h\left(X;\theta\right)\right)\right)^2}dv\leq\sup_{\theta\in\Theta}\int_{-\infty}^{\infty}\left(\int f_\varepsilon'\left(v-h\left(x;\theta\right)\right)^2h_{LC}^2\left(x\right)d\mu_X\right)^{\frac{1}{2}}dv
$$

is finite. By the Cauchy-Schwarz inequality,

$$\begin{aligned}
& \int \left(\int f_\varepsilon' \left(v - h\left(x;\theta \right) \right)^2 h_{LC}^2 \left(x \right) d\mu_X \right)^{\frac{1}{2}} dv \\
\leq & \int \left(\int f_\varepsilon' \left(v - h\left(x;\theta \right) \right)^4 d\mu_X \int h_{LC}^4 \left(x \right) d\mu_X \right)^{\frac{1}{4}} dv \\
= & \left(\int h_{LC}^4 \left(x \right) d\mu_X \right)^{\frac{1}{4}} \int \left(\int f_\varepsilon' \left(v - h\left(x;\theta \right) \right)^4 d\mu_X \right)^{\frac{1}{4}} dv.
\end{aligned}$$

The first term is bounded by assumption. The second term is finite if, for all $\theta \in \Theta$, the integrand

$$\int f_\varepsilon' \left(v - h\left(x;\theta \right) \right)^4 d\mu_X \leq K_1 \exp\left(-K_2 \left| v \right| \right)$$

for some constants K_1 and K_2. Note that

$$\begin{aligned}
& \int f_\varepsilon' \left(v - h\left(x;\theta \right) \right)^4 d\mu_X \\
= & \int \left\{ \left| h\left(x;\theta \right) \right| \geq \frac{v}{2} \right\} f_\varepsilon' \left(v - h\left(x;\theta \right) \right)^4 d\mu_X + \int \left\{ \left| h\left(x;\theta \right) \right| < \frac{v}{2} \right\} f_\varepsilon' \left(v - h\left(x;\theta \right) \right)^4 d\mu_X \\
\leq & C_1 \left\| f_\varepsilon' \right\|_\infty^4 \exp\left(-C_2 \left| \frac{v}{2} \right| \right) + \int \left\{ \left| h\left(x;\theta \right) \right| < \frac{v}{2} \right\} f_\varepsilon' \left(v - h\left(x;\theta \right) \right)^4 d\mu_X \\
\leq & C_1 \left\| f_\varepsilon' \right\|_\infty^2 \exp\left(-C_2 \left| \frac{v}{2} \right| \right) + \int C_1 \exp\left(-4C_2 \left| \frac{v}{2} \right| \right) d\mu_X \\
= & K_1 \exp\left(-K_2 \left| v \right| \right)
\end{aligned}$$

since $\left\| f_\varepsilon' \right\|_\infty < C_1$ by our bound. ■

Lemma C.18 *If the Assumptions in Proposition C.3 are satisfied, then for any sequence of positive numbers δ_N and r_N decreasing to 0, as $N \to \infty$,*

$$\sqrt{N} E \sup_{\left\| \theta - \theta_0 \right\| \leq r_N} \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta_N)}^{F_{N,V;\theta}^{-1}(1-\delta_N)} \right) \left| \int \Psi\left(x \right) f_\varepsilon \left(v - h\left(x;\theta_0 \right) \right) \left(d\mu_{X_N} - d\mu_X \right) \right| dv \to 0.$$

Proof. We bound this term as follows:

$$\begin{aligned}
& \sqrt{N}E \sup_{\|\theta-\theta_0\|\leq r_N} \left(\int_{-\infty}^{\infty} - \int_{F^{-1}_{N,V;\theta}(\delta_N)}^{F^{-1}_{N,V;\theta}(1-\delta_N)} \right) \left| \int \Psi(x) f_\varepsilon (v - h(x;\theta_0)) (d\mu_{X_N} - d\mu_X) \right| dv \\
\leq \; & \sqrt{N}E \left(\int_{-\infty}^{\infty} - \int_{V_1}^{V_2} \right) \left| \int \Psi(x) f_\varepsilon (v - h(x;\theta_0)) (d\mu_{X_N} - d\mu_X) \right| dv \\
& + \sqrt{N}E \sup_{\|\theta-\theta_0\|\leq r_N} \left[\left\{ F^{-1}_{N,V;\theta}(\delta_N) \geq V_1 \right\} + \left\{ F^{-1}_{N,V;\theta}(1-\delta_N) \leq V_2 \right\} \right] * \\
& \int_{-\infty}^{\infty} \left| \int \Psi(x) f_\varepsilon (v - h(x;\theta_0)) (d\mu_{X_N} - d\mu_X) \right| dv \\
\leq \; & \left(\int_{-\infty}^{\infty} - \int_{V_1}^{V_2} \right) \left(\int \Psi(x)^2 f_\varepsilon^2 (v - h(x;\theta_0)) d\mu_X \right)^{\frac{1}{2}} dv \\
& + \left[E \sup_{\|\theta-\theta_0\|\leq r_N} \left[\left\{ F^{-1}_{N,V;\theta}(\delta_N) \geq V_1 \right\} + \left\{ F^{-1}_{N,V;\theta}(1-\delta_N) \leq V_2 \right\} \right] \right]^{\frac{1}{2}} * \\
& \int_{-\infty}^{\infty} \left(\int \Psi(x)^2 f_\varepsilon^2 (v - h(x;\theta_0)) d\mu_X \right)^{\frac{1}{2}} dv \\
\\
\leq \; & \|\Psi\|_\infty \left(\int_{-\infty}^{\infty} - \int_{V_1}^{V_2} \right) \left(\int f_\varepsilon^2 (v - h(x;\theta_0)) d\mu_X \right)^{\frac{1}{2}} dv \\
& + \|\Psi\|_\infty \left[E \sup_{\|\theta-\theta_0\|\leq r_N} \left[\left\{ F^{-1}_{N,V;\theta}(\delta_N) \geq V_1 \right\} + \left\{ F^{-1}_{N,V;\theta}(1-\delta_N) \leq V_2 \right\} \right] \right]^{\frac{1}{2}} * \\
& \int_{-\infty}^{\infty} \left(\int f_\varepsilon^2 (v - h(x;\theta_0)) d\mu_X \right)^{\frac{1}{2}} dv
\end{aligned}$$

We now show that $E \sup_{\|\theta-\theta_0\|\leq r_N} \left[\left\{ F^{-1}_{N,V;\theta}(\delta_N) \geq V_1 \right\} \right]$ converges to zero. Note that for any $\epsilon > 0$

$$\begin{aligned}
& E \sup_{\|\theta-\theta_0\|\leq r_N} \left\{ F^{-1}_{N,V;\theta}(\delta) \geq V_1 \right\} \\
\leq \; & \left\{ F^{-1}_{V;\theta_0}(\delta) \geq V_1 - 2\epsilon \right\} + E \sup_{\|\theta-\theta_0\|\leq r_N} \left\{ \left| F^{-1}_{V;\theta}(\delta) - F^{-1}_{N,V;\theta}(\delta) \right| \geq \epsilon \right\} \\
& + \sup_{\|\theta-\theta_0\|\leq r_N} \left\{ \left| F^{-1}_{V;\theta_0}(\delta) - F^{-1}_{V;\theta}(\delta) \right| \geq \epsilon \right\}
\end{aligned}$$

We first bound these terms for a fixed δ. The first term equals 0 for $\delta < F_{V;\theta_0}(V_1 - 2\epsilon)$. By

definition,

$$\begin{aligned}
&\delta = \mu_X\left(F_\varepsilon\left(F_{V;\theta}^{-1}(\delta) - h(x;\theta)\right)\right) \\
&= \mu_{X_N}\left(F_\varepsilon\left(F_{N,V;\theta}^{-1}(\delta) - h(x;\theta)\right)\right) \\
&\Rightarrow \mu_X\left(F_\varepsilon\left(F_{V;\theta}^{-1}(\delta) - h(x;\theta)\right)\right) - \mu_X\left(F_\varepsilon\left(F_{N,V;\theta}^{-1}(\delta) - h(x;\theta)\right)\right) \\
&= \left(\mu_{X_N} - \mu_X\right)\left(F_\varepsilon\left(F_{N,V;\theta}^{-1}(\delta) - h(x;\theta)\right)\right)
\end{aligned}$$

Note that $\sqrt{N}E\sup_\theta\left(\mu_{X_N} - \mu_X\right)\left(F_\varepsilon\left(F_{N,V;\theta}^{-1}(\delta) - h(x;\theta)\right)\right) \leq \sqrt{N}E\sup_{\theta,v}\left(\mu_{X_N} - \mu_X\right)\left(F_\varepsilon\left(v - h(x;\theta)\right)\right) < \infty$. Thus, $E\sup_\theta\left|\mu_X\left(F_\varepsilon\left(F_{V;\theta}^{-1}(\delta) - h(x;\theta)\right)\right) - \mu_X\left(F_\varepsilon\left(F_{N,V;\theta}^{-1}(\delta) - h(x;\theta)\right)\right)\right| = O\left(1/\sqrt{N}\right)$. Lemma C.11 implies that $\frac{d}{dv}\mu_X\left(F_\varepsilon\left(v - h(x;\theta)\right)\right) = \int f_\varepsilon\left(v - h(x;\theta)\right)d\mu_X$ is bounded away from 0 over all θ and all v in a compact intervals. Therefore, we have that $E\sup_\theta\left|F_{V;\theta}^{-1}(\delta) - F_{N,V;\theta}^{-1}(\delta)\right| \to 0$. Finally, for any $\delta > 0$ and $q \in (\delta, 1-\delta)$

$$\nabla_\theta F_{V;\theta}^{-1}(q) = \frac{\int \nabla_\theta h(X;\theta) f_\varepsilon\left(F_{V;\theta}^{-1}(q) - h(X;\theta)\right)d\mu_X}{\int f_\varepsilon\left(F_{V;\theta}^{-1}(q) - h(X;\theta)\right)d\mu_X}$$

is bounded over all $\theta \in \Theta$ since $\nabla_\theta h(X;\theta) \leq h_{LC}(X)$ and $\int h_{LC}(X)^2 d\mu_X < \infty$ and $\int f_\varepsilon\left(F_{V;\theta}^{-1}(q) - h(X;\theta)\right)d\mu_X$ is bounded away from zero. Hence, for r_N sufficiently small, the third term is

$$\sup_{\|\theta-\theta_0\|\leq r_N}\left\{\left|F_{V;\theta_0}^{-1}(\delta) - F_{V;\theta}^{-1}(\delta)\right| \geq \epsilon\right\} = 0.$$

Therefore, there exists a sequence of $\tilde{\delta}_N$ decreasing to 0, such that the following expression converges to 0

$$\sup_{\delta\in\left(\tilde{\delta}_N, 1-\tilde{\delta}_N\right)}\left[E\sup_{\|\theta-\theta_0\|\leq r_N}\left\{\left|F_{V;\theta}^{-1}(\delta) - F_{N,V;\theta}^{-1}(\delta)\right| \geq \epsilon\right\} + \left\{\left|F_{V;\theta_0}^{-1}(\delta) - F_{V;\theta}^{-1}(\delta)\right| \geq \epsilon\right\}\right].$$

Since $\left\{F_{V;\theta_0}^{-1}\left(\tilde{\delta}_N\right) \geq V_1 - 2\epsilon\right\} \to 0$, we have that

$$E \sup_{\|\theta-\theta_0\|\leq r_N} \left\{F_{N,V;\theta}^{-1}\left(\delta_N\right) \geq V_1\right\} \leq E \sup_{\|\theta-\theta_0\|\leq r_N} \left\{F_{N,V;\theta}^{-1}\left(\max\left(\delta_N, \tilde{\delta}_N\right)\right) \geq V_1\right\} \to 0.$$

Similar arguments show that $E \sup_{\|\theta-\theta_0\|\leq r_N} \left\{F_{N,V;\theta}^{-1}\left(1-\delta_N\right) \leq V_2\right\} \to 0$. It follows that there exist a sequence of $V_{1,N} \to -\infty$, $V_{2,N} \to \infty$ such that

$$\left[E \sup_{\|\theta-\theta_0\|\leq r_N} \left[\left\{F_{N,V;\theta}^{-1}\left(\delta_N\right) \geq V_{1,N}\right\} + \left\{F_{N,V;\theta}^{-1}\left(1-\delta_N\right) \leq V_{2,N}\right\}\right]\right]^{\frac{1}{2}} \to 0.$$

Therefore,

$$\sqrt{N}E \sup_{\|\theta-\theta_0\|\leq r_N} \left(\int_{-\infty}^{\infty} - \int_{F_{N,V;\theta}^{-1}(\delta_N)}^{F_{N,V;\theta}^{-1}(1-\delta_N)}\right) \left|\int \Psi\left(x\right) f_\varepsilon\left(v - h\left(x;\theta_0\right)\right)\left(d\mu_{X_N} - d\mu_X\right)\right| dv \to 0.$$

■

C.3.5 Primitives for Assumption 3.6(ii) c.

Proposition C.4 *If Assumption C.3 is satisfied,then, for any sequence of positive numbers b_N decreasing to 0, and for any $\delta > 0$,*

$$\sup_{\|\theta-\theta_0\|\leq b_N} \left|\nabla_{(G_X,G_Z)}\psi^\delta\left(\theta\right) - \nabla_{(G_X,G_Z)}\psi^\delta\left(\theta_0\right)\right| = o_p\left(1\right).$$

Proof. For a fixed $\delta > 0$, consider the Gaussian process $\nabla_{(G_X,G_Z)}\psi^\delta\left(\theta\right)$, indexed by Θ. The expression for this term (given in Appendix C.2.2) is a sum and product of finitely many terms of the form of the following two expressions

$$\int_\delta^{1-\delta} \frac{\int \Psi\left(x_1,x_2,z\right)\phi_\varepsilon\left(q,x_1,\theta_1\right)\phi_\varepsilon\left(q,x_2,\theta_1\right)\phi_\eta\left(q,z,\theta_1\right)dG_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon\left(q,x_1,\theta_1\right)\phi_\varepsilon\left(q,x_2,\theta_1\right)\phi_\eta\left(q,z,\theta_1\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z} dq,$$

$$\int_\delta^{1-\delta} G_V^q\left(\theta_1\right)$$

$$\frac{\int \Psi\left(x_1,x_2,z\right) f_\varepsilon'\left(F_{V;\theta_1}^{-1}\left(q\right) - h\left(x_1;\theta_1\right)\right)\phi_\varepsilon\left(q,x_2;\theta_1\right)\phi_\eta\left(q,z;\theta_1\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon\left(q,x_1;\theta_1\right)\phi_\varepsilon\left(q,x_2;\theta_1\right)\phi_\eta\left(q,z;\theta_1\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z} dq$$

and analogous terms with G_Z and G_U^q instead of G_{X_1} and G_V^q. We will show that for any

sequence of positive numbers b_N decreasing to 0,

$$\sup_{\|\theta-\theta_0\|\leq b_N} \left| \nabla_{(G_X,G_Z)} \psi^\delta(\theta) - \nabla_{(G_X,G_Z)} \psi^\delta(\theta_0) \right| = o_p(1)$$

by individually analyzing these terms.

First consider the Gaussian process $\tilde{G}(\theta)$ indexed by Θ, given by

$$\tilde{G}(\theta) = \int_\delta^{1-\delta} \frac{\int \Psi(x_1, x_2, z) \phi_\varepsilon(q, x_1, \theta_1) \phi_\varepsilon(q, x_2, \theta_1) \phi_\eta(q, z, \theta_1) dG_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q, x_1, \theta_1) \phi_\varepsilon(q, x_2, \theta_1) \phi_\eta(q, z, \theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq.$$

We show that for any sequence of positive numbers b_N decreasing to 0, we have that

$$\sup_{\|\theta-\theta_0\|\leq b_N} \left| \tilde{G}(\theta) - \tilde{G}(\theta_0) \right| = o_p(1).$$

To do so, it is enough to show that $\tilde{G}$ has almost surely uniformly continuous sample paths in θ. By Dudley's Theorem (e.g. Theorem 2.6.1 of Dudley (2014)), $\tilde{G}(\theta)$ has almost surely uniformly continuous sample paths if $\int_0^\infty \sqrt{\log N_{\tilde{G}}(\epsilon)} d\epsilon$ is finite, where $N_{\tilde{G}}(\epsilon)$ is the $\epsilon - L_2$ covering number for $\tilde{G}$. Note that if $N_{\tilde{G}}(\epsilon) \leq C_0 \epsilon^d$ for some constant C_0 and natural number d, this integral is finite. A sufficient condition is that $\left(E\left(\tilde{G}(\theta_1) - \tilde{G}(\theta_2) \right)^2 \right)^{\frac{1}{2}} < K \|\theta_1 - \theta_2\|$ since Θ is finite dimensional.

Hence, we must bound

$$\left[E\left(\begin{array}{c}\int_{\delta}^{1-\delta}\frac{\int\Psi(x_1,x_2,z)\phi_\varepsilon(q,x_1,\theta_1)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1,\theta_1)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)d\mu_{X_1}d\mu_{X_2}d\mu_Z}dqdG_{X_1}\\ -\int_{\delta}^{1-\delta}\frac{\int\Psi(x_1,x_2,z)\phi_\varepsilon(q,x_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_2)d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_2)d\mu_{X_1}d\mu_{X_2}d\mu_Z}dqdG_{X_1}\end{array}\right)^2\right]^{\frac{1}{2}}$$

$$= Var\left(\begin{array}{c}\int_{\delta}^{1-\delta}\frac{\int\Psi(X_1,x_2,z)\phi_\varepsilon(q,X_1,\theta_1)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1,\theta_1)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq\\ -\int_{\delta}^{1-\delta}\frac{\int\Psi(X_1,x_2,z)\phi_\varepsilon(q,X_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_2)d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_2)d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq\end{array}\right)^{\frac{1}{2}}$$

$$\leq \left(E\left(\int_{\delta_1}^{1-\delta_1}\left(\begin{array}{c}\frac{\int\Psi(X_1,x_2,z)\phi_\varepsilon(q,X_1,\theta_1)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)}{\int\phi_\varepsilon(q,x_1,\theta_1)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)d\mu_{X_1}d\mu_{X_2}d\mu_Z}\\ -\frac{\int\Psi(X_1,x_2,z)\phi_\varepsilon(q,X_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_2)d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_2)d\mu_{X_1}d\mu_{X_2}d\mu_Z}\end{array}\right)dqd\mu_{X_2}d\mu_Z\right)^2\right)^{\frac{1}{2}}$$

$$\leq \left(E\left(\int_0^1\left(\begin{array}{c}\frac{\int\Psi(X_1,x_2,z)\phi_\varepsilon(q,X_1,\theta_1)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)}{\int\phi_\varepsilon(q,x_1,\theta_1)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)d\mu_{X_1}d\mu_{X_2}d\mu_Z}\\ -\frac{\int\Psi(X_1,x_2,z)\phi_\varepsilon(q,X_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)d\mu_{X_1}d\mu_{X_2}d\mu_Z}\end{array}\right)dqd\mu_{X_2}d\mu_Z\right)^2\right)^{\frac{1}{2}}$$

$$+\left(E\left(\int_0^1\left(\begin{array}{c}\frac{\int\Psi(X_1,x_2,z)\phi_\varepsilon(q,X_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_1)\phi_\eta(q,z,\theta_1)d\mu_{X_1}d\mu_{X_2}d\mu_Z}\\ -\frac{\int\Psi(X_1,x_2,z)\phi_\varepsilon(q,X_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_1)d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_1)d\mu_{X_1}d\mu_{X_2}d\mu_Z}\end{array}\right)dqd\mu_{X_2}d\mu_Z\right)^2\right)^{\frac{1}{2}}$$

$$+\left(E\left(\int_0^1\left(\begin{array}{c}\frac{\int\Psi(X_1,x_2,z)\phi_\varepsilon(q,X_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_1)d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_1)d\mu_{X_1}d\mu_{X_2}d\mu_Z}\\ -\frac{\int\Psi(X_1,x_2,z)\phi_\varepsilon(q,X_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_2)d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1,\theta_2)\phi_\varepsilon(q,x_2,\theta_2)\phi_\eta(q,z,\theta_2)d\mu_{X_1}d\mu_{X_2}d\mu_Z}\end{array}\right)dqd\mu_{X_2}d\mu_Z\right)^2\right)^{\frac{1}{2}}.$$

By a change of variables, $v = F_{V;\theta_1}^{-1}(q)$, the first of these 3 terms is not greater than

$$\leq \|\Psi\|_\infty\left(E\left(\int_{-\infty}^{\infty}\left|\left(f_\varepsilon\left(v-h\left(X_1;\theta_1\right)\right)-f_\varepsilon\left(v-h\left(X_1;\theta_2\right)\right)\right)\right|dv\right)^2\right)^{\frac{1}{2}}$$

$$= \|\Psi\|_\infty\left(E\left(h\left(X_1;\theta_1\right)-h\left(X_1;\theta_2\right)\right)^2\right)^{\frac{1}{2}}\int_{-\infty}^{\infty}\left|f_\varepsilon'\left(v\right)\right|dv$$

$$\leq \|\theta_1-\theta_2\|\left(\int h_{LC}\left(X\right)^2d\mu_X\right)^{\frac{1}{2}}\|\Psi\|_\infty\int_{-\infty}^{\infty}\left|f_\varepsilon'\left(v\right)\right|dv < K\|\theta_1-\theta_2\|$$

for a finite constant K. The next two terms are handled similarly. Hence, $\tilde{G}(\theta)$ has almost surely uniformly continuous sample paths.

By a similar argument, a bound on

$$\left[E \left[\begin{array}{c} \int_{\delta}^{1-\delta} G_V^q(\theta_1) \frac{\int \Psi(x_1,x_2,z) f_\varepsilon' \left(F_{V;\theta_1}^{-1}(q) - h(x_1;\theta_1) \right) \phi_\varepsilon(q,x_2;\theta_1) \phi_\eta(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta_1) \phi_\varepsilon(q,x_2;\theta_1) \phi_\eta(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \\ - \int_{\delta}^{1-\delta} G_V^q(\theta_2) \frac{\int \Psi(x_1,x_2,z) f_\varepsilon' \left(F_{V;\theta_2}^{-1}(q) - h(x_1;\theta_2) \right) \phi_\varepsilon(q,x_2;\theta_2) \phi_\eta(q,z;\theta_2) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta_2) \phi_\varepsilon(q,x_2;\theta_2) \phi_\eta(q,z;\theta_2) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \end{array} \right]^2 \right]^{\frac{1}{2}}$$

implies that

$$\int_{\delta}^{1-\delta} G_V^q(\theta_1) \frac{\int \Psi(x_1,x_2,z) f_\varepsilon' \left(F_{V;\theta_1}^{-1}(q) - h(x_1;\theta_1) \right) \phi_\varepsilon(q,x_2;\theta_1) \phi_\eta(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta_1) \phi_\varepsilon(q,x_2;\theta_1) \phi_\eta(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq$$

has almost surely uniformly continuous sample paths. Note that

$$
\begin{aligned}
& \left[E \left[\begin{array}{c} \int_{\delta}^{1-\delta} G_V^q(\theta_1) \frac{\int \Psi(x_1,x_2,z) f_{\varepsilon}' \left(F_{V;\theta_1}^{-1}(q) - h(x_1;\theta_1) \right) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q,x_1;\theta_1) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \\ - \int_{\delta}^{1-\delta} G_V^q(\theta_2) \frac{\int \Psi(x_1,x_2,z) f_{\varepsilon}' \left(F_{V;\theta_2}^{-1}(q) - h(x_1;\theta_2) \right) \phi_{\varepsilon}(q,x_2;\theta_2) \phi_{\eta}(q,z;\theta_2) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q,x_1;\theta_2) \phi_{\varepsilon}(q,x_2;\theta_2) \phi_{\eta}(q,z;\theta_2) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \end{array} \right]^2 \right]^{\frac{1}{2}} \\
= & \left[E \left[\begin{array}{c} \int_{\delta}^{1-\delta} \frac{1}{f_{V;\theta_1}\left(F_{V;\theta_1}^{-1}(q) \right)} \int G_X \left(1 \left\{ h(x;\theta_1) + \varepsilon \leq F_{V;\theta_1}^{-1}(q) \right\} \right) dF_{\varepsilon} * \\ \frac{\int \Psi(x_1,x_2,z) f_{\varepsilon}' \left(F_{V;\theta_1}^{-1}(q) - h(x_1;\theta_1) \right) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q,x_1;\theta_1) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \\ - \int_{\delta}^{1-\delta} \frac{1}{f_{V;\theta_2}\left(F_{V;\theta_2}^{-1}(q) \right)} \int G_X \left(1 \left\{ h(x;\theta_2) + \varepsilon \leq F_{V;\theta_2}^{-1}(q) \right\} \right) dF_{\varepsilon} * \\ \frac{\int \Psi(x_1,x_2,z) f_{\varepsilon}' \left(F_{V;\theta_2}^{-1}(q) - h(x_1;\theta_2) \right) \phi_{\varepsilon}(q,x_2;\theta_2) \phi_{\eta}(q,z;\theta_2) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q,x_1;\theta_2) \phi_{\varepsilon}(q,x_2;\theta_2) \phi_{\eta}(q,z;\theta_2) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \end{array} \right]^2 \right]^{\frac{1}{2}} \\
\leq & \left[E \left[\begin{array}{c} \int_{\delta}^{1-\delta} (\frac{1}{f_{V;\theta_1}\left(F_{V;\theta_1}^{-1}(q) \right)} \int G_X \left(1 \left\{ h(x;\theta_1) + \varepsilon \leq F_{V;\theta_1}^{-1}(q) \right\} \right) dF_{\varepsilon} - \\ \frac{1}{f_{V;\theta_2}\left(F_{V;\theta_2}^{-1}(q) \right)} \int G_X \left(1 \left\{ h(x;\theta_2) + \varepsilon \leq F_{V;\theta_2}^{-1}(q) \right\} \right) dF_{\varepsilon}) \times \\ \frac{\int \Psi(x_1,x_2,z) f_{\varepsilon}' \left(F_{V;\theta_1}^{-1}(q) - h(x_1;\theta_1) \right) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q,x_1;\theta_1) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \end{array} \right]^2 \right]^{\frac{1}{2}} \\
& + \left[E \left[\begin{array}{c} \int_{\delta}^{1-\delta} \frac{1}{f_{V;\theta_2}\left(F_{V;\theta_2}^{-1}(q) \right)} \int G_X \left(1 \left\{ h(x;\theta_2) + \varepsilon \leq F_{V;\theta_2}^{-1}(q) \right\} \right) dF_{\varepsilon} \times \\ \left(\begin{array}{c} \frac{\int \Psi(x_1,x_2,z) f_{\varepsilon}' \left(F_{V;\theta_2}^{-1}(q) - h(x_1;\theta_2) \right) \phi_{\varepsilon}(q,x_2;\theta_2) \phi_{\eta}(q,z;\theta_2) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q,x_1;\theta_2) \phi_{\varepsilon}(q,x_2;\theta_2) \phi_{\eta}(q,z;\theta_2) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\ - \frac{\int \Psi(x_1,x_2,z) f_{\varepsilon}' \left(F_{V;\theta_1}^{-1}(q) - h(x_1;\theta_1) \right) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q,x_1;\theta_1) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \end{array} \right) dq \end{array} \right]^2 \right]^{\frac{1}{2}} \\
= & \left[E \left[\begin{array}{c} \int_{\delta}^{1-\delta} (\frac{1}{f_{V;\theta_1}\left(F_{V;\theta_1}^{-1}(q) \right)} \int 1 \left\{ h(X;\theta_1) + \varepsilon \leq F_{V;\theta_1}^{-1}(q) \right\} dF_{\varepsilon} - \\ \frac{1}{f_{V;\theta_2}\left(F_{V;\theta_2}^{-1}(q) \right)} \int 1 \left\{ h(X;\theta_2) + \varepsilon \leq F_{V;\theta_2}^{-1}(q) \right\} dF_{\varepsilon}) \times \\ \frac{\int \Psi(x_1,x_2,z) f_{\varepsilon}' \left(F_{V;\theta_1}^{-1}(q) - h(x_1;\theta_1) \right) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q,x_1;\theta_1) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq \end{array} \right]^2 \right]^{\frac{1}{2}} \\
& + \left[E \left[\begin{array}{c} \int_{\delta}^{1-\delta} \frac{1}{f_{V;\theta_2}\left(F_{V;\theta_2}^{-1}(q) \right)} \int 1 \left\{ h(X;\theta_2) + \varepsilon \leq F_{V;\theta_2}^{-1}(q) \right\} dF_{\varepsilon} \times \\ \left(\begin{array}{c} \frac{\int \Psi(x_1,x_2,z) f_{\varepsilon}' \left(F_{V;\theta_2}^{-1}(q) - h(x_1;\theta_2) \right) \phi_{\varepsilon}(q,x_2;\theta_2) \phi_{\eta}(q,z;\theta_2) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q,x_1;\theta_2) \phi_{\varepsilon}(q,x_2;\theta_2) \phi_{\eta}(q,z;\theta_2) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \\ - \frac{\int \Psi(x_1,x_2,z) f_{\varepsilon}' \left(F_{V;\theta_1}^{-1}(q) - h(x_1;\theta_1) \right) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_{\varepsilon}(q,x_1;\theta_1) \phi_{\varepsilon}(q,x_2;\theta_1) \phi_{\eta}(q,z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \end{array} \right) dq \end{array} \right]^2 \right]^{\frac{1}{2}} \\
= & T_1 + T_2
\end{aligned}
$$

where the last equality follows from the definition of G_X's covariance kernel. To bound T_1, note that for any $\delta > 0$ and all $q \in (\delta, 1-\delta)$, we have that

$$\left| \frac{\int \Psi(x_1, x_2, z) f'_\varepsilon \left(F^{-1}_{V;\theta_1}(q) - h(x_1;\theta_1) \right) \phi_\varepsilon(q, x_2;\theta_1) \phi_\eta(q, z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q, x_1;\theta_1) \phi_\varepsilon(q, x_2;\theta_1) \phi_\eta(q, z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z} \right| < M < \infty$$

since $\inf_{\theta, q \in (\delta, 1-\delta)} \int \phi_\varepsilon(q, x_1;\theta_1) \phi_\varepsilon(q, x_2;\theta_1) \phi_\eta(q, z;\theta_1) d\mu_{X_1} d\mu_{X_2} d\mu_Z > 0$ (Lemma C.11) and the numerator is uniformly bounded. Hence, the T_1 no greater than

$$\begin{aligned}
& M \left[E \left[\int_\delta^{1-\delta} \left| \begin{array}{l} \frac{1}{f_{V;\theta_1}\left(F^{-1}_{V;\theta_1}(q)\right)} \int 1\left\{ h(X;\theta_1) + \varepsilon \le F^{-1}_{V;\theta_1}(q) \right\} dF_\varepsilon \\ -\frac{1}{f_{V;\theta_2}\left(F^{-1}_{V;\theta_2}(q)\right)} \int 1\left\{ h(X;\theta_2) + \varepsilon \le F^{-1}_{V;\theta_2}(q) \right\} dF_\varepsilon \end{array} \right| dq \right]^2 \right]^{\frac{1}{2}} \\
\le \; & M \left[E \left[\int_\delta^{1-\delta} \left| \begin{array}{l} \frac{1}{f_{V;\theta_1}\left(F^{-1}_{V;\theta_1}(q)\right)} \int 1\left\{ h(X;\theta_1) + \varepsilon \le F^{-1}_{V;\theta_1}(q) \right\} dF_\varepsilon \\ -\frac{1}{f_{V;\theta_1}\left(F^{-1}_{V;\theta_1}(q)\right)} \int 1\left\{ h(X;\theta_2) + \varepsilon \le F^{-1}_{V;\theta_2}(q) \right\} dF_\varepsilon \end{array} \right| dq \right]^2 \right]^{\frac{1}{2}} \\
& + M \left[E \left[\int_\delta^{1-\delta} \left| \begin{array}{l} \frac{1}{f_{V;\theta_1}\left(F^{-1}_{V;\theta_1}(q)\right)} \int 1\left\{ h(X;\theta_2) + \varepsilon \le F^{-1}_{V;\theta_2}(q) \right\} dF_\varepsilon \\ -\frac{1}{f_{V;\theta_2}\left(F^{-1}_{V;\theta_2}(q)\right)} \int 1\left\{ h(X;\theta_2) + \varepsilon \le F^{-1}_{V;\theta_2}(q) \right\} dF_\varepsilon \end{array} \right| dq \right]^2 \right]^{\frac{1}{2}} \\
\le \; & M \left[E \left[\int_\delta^{1-\delta} \left| \frac{1}{f_{V;\theta_1}\left(F^{-1}_{V;\theta_1}(q)\right)} \begin{pmatrix} F_\varepsilon\left(F^{-1}_{V;\theta_1}(q) - h(X;\theta_1)\right) \\ -F_\varepsilon\left(F^{-1}_{V;\theta_2}(q) + h(X;\theta_2)\right) \end{pmatrix} \right| dq \right]^2 \right]^{\frac{1}{2}} \\
& + M \left[E \left[\int_\delta^{1-\delta} \left| \begin{array}{l} \frac{1}{f_{V;\theta_1}\left(F^{-1}_{V;\theta_1}(q)\right)} \int 1\left\{ h(X;\theta_2) + \varepsilon \le F^{-1}_{V;\theta_2}(q) \right\} dF_\varepsilon \\ -\frac{1}{f_{V;\theta_2}\left(F^{-1}_{V;\theta_2}(q)\right)} \int 1\left\{ h(X;\theta_2) + \varepsilon \le F^{-1}_{V;\theta_2}(q) \right\} dF_\varepsilon \end{array} \right| dq \right]^2 \right]^{\frac{1}{2}} \\
= \; & R_1 + R_2
\end{aligned}$$

Note that

$$\nabla_\theta F^{-1}_{V;\theta}(q) = \frac{\int f_\varepsilon \left(F^{-1}_{V;\theta}(q) - h(x;\theta) \right) \nabla_\theta h(x;\theta) d\mu_X}{f_{V;\theta}\left(F^{-1}_{V;\theta}(q)\right)}.$$

and $\inf_{\theta,q\in(\delta,1-\delta)} f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right) > 0$ (Lemma C.11). Hence, R_1 is no greater than

$$
\begin{aligned}
& M\left(\frac{1}{\inf_{\theta,q\in(\delta,1-\delta)} f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)}\right) \times \\
& \left[E\left[\int_{\delta}^{1-\delta} \|f_{\varepsilon}\|_{\infty} \left|\frac{\int f_{\varepsilon}\left(F_{V;\theta}^{-1}(q) - h(x;\theta)\right) \nabla_{\theta} h(x;\theta)\, d\mu_X}{f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)} - \nabla_{\theta} h(X;\theta)\right|\right.\right. \\
& \left.\left. \|\theta_1 - \theta_2\|\, dq\right]^2\right]^{\frac{1}{2}} \\
\leq\ & M\frac{\|f_{\varepsilon}\|_{\infty}\|\theta_1 - \theta_2\|}{\inf_{\theta,q\in(\delta,1-\delta)} f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)} * \\
& \left(\left|\frac{\|f_{\varepsilon}\|_{\infty}}{\inf_{\theta,q\in(\delta,1-\delta)} f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)} \int h_{LC}(x)\, d\mu_X\right| + \left(\int h_{LC}(x)^2\, d\mu_X\right)^{\frac{1}{2}}\right) \\
<\ & K_1 M\|\theta_1 - \theta_2\|
\end{aligned}
$$

since $\frac{1}{\inf_{\theta,q\in(\delta,1-\delta)} f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)}$, $\|f_{\varepsilon}\|_{\infty}$ and $\left(\int h_{LC}(x)^2\, d\mu_X\right)^{\frac{1}{2}}$ are finite.

Similarly, to bound R_2, note that $\nabla_{\theta}\left(\frac{1}{f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)}\right)$ is given by

$$
\begin{aligned}
& -\frac{\int\left(\nabla_{\theta} F_{V;\theta}^{-1}(q) - \nabla_{\theta} h(x;\theta)\right) f_{\varepsilon}'\left(F_{V;\theta}^{-1}(q) - h(x;\theta)\right) d\mu_X}{\left(f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)\right)^2} \\
=\ & -\frac{1}{\left(f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)\right)^2} \int\left(\frac{\int f_{\varepsilon}\left(F_{V;\theta}^{-1}(q) - h(x;\theta)\right) \nabla_{\theta} h(x;\theta)\, d\mu_X}{f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)} - \nabla_{\theta} h(x;\theta)\right) \\
& f_{\varepsilon}'\left(F_{V;\theta}^{-1}(q) - h(x;\theta)\right) d\mu_X.
\end{aligned}
$$

Hence, $\sup_{\theta,q\in(\delta,1-\delta)}\left|\nabla_{\theta}\left(\frac{1}{f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)}\right)\right|$ is at most

$$
\begin{aligned}
& \left(\frac{1}{\inf_{\theta,q\in(\delta,1-\delta)} f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)}\right)^2 \|f_{\varepsilon}'\|_{\infty} * \\
& \left(\frac{1}{\inf_{\theta,q\in(\delta,1-\delta)} f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)} \|f_{\varepsilon}\|_{\infty} + 1\right) \int |h_{LC}(x)|\, d\mu_X < K_2 < \infty
\end{aligned}
$$

for each $q \in (\delta, 1-\delta)$. Therefore, the R_2 is at most $MK_2\|\theta_1 - \theta_2\|$. Similarly, since

$\left|\frac{1}{f_{V;\theta_2}\left(F^{-1}_{V;\theta_2}(q)\right)}\int 1\left\{h\left(X;\theta_2\right)+\varepsilon\leq F^{-1}_{V;\theta_2}\left(q\right)\right\}dF_\varepsilon\right|$ is bounded, a uniform bound on the derivative of

$$\frac{\int\Psi\left(x_1,x_2,z\right)f'_\varepsilon\left(F^{-1}_{V;\theta_1}\left(q\right)-h\left(x_1;\theta_1\right)\right)\phi_\varepsilon\left(q,x_2;\theta_1\right)\phi_\eta\left(q,z;\theta_1\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon\left(q,x_1;\theta_1\right)\phi_\varepsilon\left(q,x_2;\theta_1\right)\phi_\eta\left(q,z;\theta_1\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}$$

with respect to θ_1 implies that $T_2\leq K_3\left\|\theta_1-\theta_2\right\|$ for some constant K_3. This follows from identical arguments as the ones above.

Hence,

$$\begin{aligned}&\left[E\left[\begin{array}{c}\int_\delta^{1-\delta}G_V^q\left(\theta_1\right)\frac{\int\Psi(x_1,x_2,z)f'_\varepsilon\left(F^{-1}_{V;\theta_1}(q)-h(x_1;\theta_1)\right)\phi_\varepsilon(q,x_2;\theta_1)\phi_\eta(q,z;\theta_1)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta_1)\phi_\varepsilon(q,x_2;\theta_1)\phi_\eta(q,z;\theta_1)d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq\\-\int_\delta^{1-\delta}G_V^q\left(\theta_2\right)\frac{\int\Psi(x_1,x_2,z)f'_\varepsilon\left(F^{-1}_{V;\theta_2}(q)-h(x_1;\theta_2)\right)\phi_\varepsilon(q,x_2;\theta_2)\phi_\eta(q,z;\theta_2)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int\phi_\varepsilon(q,x_1;\theta_2)\phi_\varepsilon(q,x_2;\theta_2)\phi_\eta(q,z;\theta_2)d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq\end{array}\right]^2\right]^{\frac{1}{2}}\\&\leq T_1+T_2\leq K\left\|\theta_1-\theta_2\right\|\end{aligned}$$

for some constant $K\in(0,\infty)$.

The proof for the remaining terms in $\nabla_{(G_X,G_Z)}\psi^\delta\left(\theta\right)$ is analogous. Therefore,

$$\left(E\left(\nabla_{(G_X,G_Z)}\psi^\delta\left(\theta_1\right)-\nabla_{(G_X,G_Z)}\psi^\delta\left(\theta_2\right)\right)^2\right)^{\frac{1}{2}}<\tilde{K}\left\|\theta_1-\theta_2\right\|$$

for some constant $\tilde{K}$, implying that the $\epsilon-L^2$ covering numbers are bounded above by a polynomial in $\frac{1}{\epsilon}$, completing the proof. ■

C.3.6 Primitives for Assumption 3.6(ii) d.

Proposition C.5 *If $\left\|\Psi\right\|_\infty<\infty$, $\left\|\nabla\tilde{\psi}_q\right\|_\infty^2<\infty$, $F_{U;\theta_0}$ and $F_{V;\theta_0}$ have full support on $\mathbb{R}$, and and $g\left(Z;\theta_0\right)$ and $h\left(X;\theta_0\right)$ have finite second moments, then*

$$\left|\nabla_{\tilde{G}}\psi^\delta\left[\mu_X,\mu_Z\right]\left(\theta_0\right)-\nabla_{\tilde{G}}\psi^0\left[\mu_X,\mu_Z\right]\left(\theta_0\right)\right|$$

converges in probability to 0 as $\delta\to 0$.

Proof. The expression for $\nabla_{\tilde{G}}\psi^\delta\left[\mu_X,\mu_Z\right]$ is given in equation (C.30). We show convergence

of each of the terms in $Lim_{G;\delta}(\theta_0)$ as $\delta \to 0$. First, we show that

$$\left(\int_0^1 - \int_\delta^{1-\delta}\right) \frac{\int \Psi(x_1, x_2, z)\, \phi_\varepsilon(q, x_1; \theta_0)\, \phi_\varepsilon(q, x_2; \theta_0)\, \phi_\eta(q, z; \theta_0)\, dG_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q, x_1; \theta_0)\, \phi_\varepsilon(q, x_2; \theta_0)\, \phi_\eta(q, z; \theta_0)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq$$

converges weakly as $\delta \to 0$.

This term has mean zero and variance not greater than

$$\begin{aligned}
& \|\Psi\|_\infty^2 \int \left[\left(\int_0^1 - \int_\delta^{1-\delta}\right) \frac{\int \phi_\varepsilon(q, X_1; \theta_0)\, \phi_\varepsilon(q, x_2; \theta_0)\, \phi_\eta(q, z; \theta_0)\, d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q, x_1; \theta_0)\, \phi_\varepsilon(q, x_2; \theta_0)\, \phi_\eta(q, z; \theta_0)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq\right]^2 d\mu_{X_1} \\
= & \|\Psi\|_\infty^2 \int \left[\left(\int_0^1 - \int_\delta^{1-\delta}\right) \frac{f_\varepsilon\left(F_{V;\theta_0}^{-1}(q) - h(X_1; \theta_0)\right)}{\int f_\varepsilon\left(F_{V;\theta_0}^{-1}(q) - h(X_1; \theta_0)\right) d\mu_{X_1}} dq\right]^2 d\mu_{X_1} \\
= & \|\Psi\|_\infty^2 \int \left[\left(\int_{-\infty}^{\infty} - \int_{F_{V;\theta_0}^{-1}(\delta)}^{F_{V;\theta_0}^{-1}(1-\delta)}\right) f_\varepsilon(v - h(X_1; \theta_0))\, dv\right]^2 d\mu_{X_1}.
\end{aligned}$$

where the last equality follows from a change of variables, $v = F_{V;\theta_0}^{-1}(q)$. Since $\int_{-\infty}^{\infty} f_\varepsilon(v - h(X_1; \theta_0))\, dv = 1$ for all X_1, and $F_{V;\theta_0}^{-1}(\delta) \to -\infty$ and $F_{V;\theta_0}^{-1}(1-\delta) \to \infty$ as $\delta \to 0$, the bound above converges to 0 as $\delta \to 0$ by the dominated convergence theorem. This proves that the term

$$\left(\int_0^1 - \int_\delta^{1-\delta}\right) \frac{\int \Psi(x_1, x_2, z)\, \phi_\varepsilon(q, x_1; \theta_0)\, \phi_\varepsilon(q, x_2; \theta_0)\, \phi_\eta(q, z; \theta_0)\, dG_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon(q, x_1; \theta_0)\, \phi_\varepsilon(q, x_2; \theta_0)\, \phi_\eta(q, z; \theta_0)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq$$

converges to 0 in probability as $\delta \to 0$.

Next, recall that

$$\frac{1}{f_{U;\theta}\left(F_{U,\theta}^{-1}(q)\right)} \int G_Z\left(1\left\{g(z; \theta) + \eta \leq F_{U;\theta}^{-1}(q)\right\}\right) dF_\eta = G_U^q(\theta).$$

Consider the terms that include $G_U^q(\theta_0)$ in the expression for $\nabla_{(G_X, G_Z)} \psi^\delta[\mu_X, \mu_Z](\theta_0)$. The

sum of these are given by

$$\int_{\delta}^{1-\delta} G_U^q(\theta_0) * \frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta_0)\,\phi_\varepsilon(q,x_2;\theta_0)\,f'_\eta\left(F^{-1}_{U;\theta}(q)-g(z;\theta_0)\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta_0)\,\phi_\varepsilon(q,x_2;\theta_0)\,\phi_\eta(q,z;\theta_0)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq$$
$$-\int_{\delta}^{1-\delta}\frac{\int \Psi(x_1,x_2,z)\,\phi_\varepsilon(q,x_1;\theta_0)\,\phi_\varepsilon(q,x_2;\theta_0)\,\phi_\eta(q,z;\theta_0)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta_0)\,\phi_\varepsilon(q,x_2;\theta_0)\,\phi_\eta(q,z;\theta_0)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}\times$$
$$G_U^q(\theta)\,\frac{\int \phi_\varepsilon(q,x_1;\theta_0)\,\phi_\varepsilon(q,x_2;\theta_0)\,f'_\eta\left(F^{-1}_{U;\theta}(q)-g(z;\theta_0)\right)d\mu_{X_1}d\mu_{X_2}d\mu_Z}{\int \phi_\varepsilon(q,x_1;\theta_0)\,\phi_\varepsilon(q,x_2;\theta_0)\,\phi_\eta(q,z;\theta_0)\,d\mu_{X_1}d\mu_{X_2}d\mu_Z}dq.$$

Note that this term is equal to

$$\begin{aligned}&\int_{\delta}^{1-\delta} G_U^q(\theta)\,\frac{\partial}{\partial q_3}\tilde{\psi}_q dq\\ =\;&\int_{\delta}^{1-\delta}\frac{1}{f_{U;\theta}\left(F^{-1}_{U,\theta}(q)\right)}\int G_Z\left(1\left\{g(z;\theta)+\eta\le F^{-1}_{U;\theta}(q)\right\}\right)dF_\eta\frac{\partial}{\partial q_3}\tilde{\psi}_q dq.\end{aligned}$$

Therefore,

$$\begin{aligned}&\nabla_{(G_X,G_Z)}\psi^\delta[\mu_X,\mu_Z](\theta_0)-\nabla_{(G_X,G_Z)}\psi^0[\mu_X,\mu_Z](\theta_0)\\ =\;&\left(\int_0^1-\int_\delta^{1-\delta}\right)\frac{1}{f_{U;\theta}\left(F^{-1}_{U,\theta}(q)\right)}\int G_Z\left(1\left\{g(z;\theta)+\eta\le F^{-1}_{U;\theta}(q)\right\}\right)dF_\eta\frac{\partial}{\partial q_3}\tilde{\psi}_q dq\\ =\;&\left(\int_{-\infty}^{\infty}-\int_{F^{-1}_{U;\theta_0}(\delta)}^{F^{-1}_{U;\theta_0}(1-\delta)}\right)\int G_Z\left(1\{g(z;\theta)+\eta\le u\}\right)dF_\eta\left.\frac{\partial}{\partial q_3}\tilde{\psi}_q\right|_{q_3=F_{U;\theta}(u)}du\end{aligned}$$

has mean zero and variance not greater than

$$\left\|\nabla\tilde{\psi}_q\right\|_\infty^2\int\left[\left(\int_{-\infty}^{\infty}-\int_{F^{-1}_{U;\theta_0}(\delta)}^{F^{-1}_{U;\theta_0}(1-\delta)}\right)\left[F_\eta(u-g(Z;\theta_0))-EF_\eta(u-g(Z;\theta_0))\right]du\right]^2 d\mu_Z.$$

By the Efron-Stein inequality, let $Z^{(i)}$ have the same distribution as Z, and note that

$$\begin{aligned}
& \int \left[\int_{-\infty}^{\infty} \left[F_\eta\left(u - g\left(Z;\theta_0\right)\right) - EF_\eta\left(u - g\left(Z;\theta_0\right)\right)\right] du\right]^2 d\mu_Z \\
\leq\ & \frac{1}{2}\int\int \left[\int_{-\infty}^{\infty} \left[F_\eta\left(u - g\left(Z;\theta_0\right)\right) - F_\eta\left(u - g\left(Z^{(i)};\theta_0\right)\right)\right] du\right]^2 d\mu_Z d\mu_{Z^{(i)}} \\
=\ & \frac{1}{2}\int\int \left[\int_{-\infty}^{\infty} \left[\int_{g(Z;\theta_0)}^{g\left(Z^{(i)};\theta_0\right)} f_\eta\left(u - g\right) dg\right] du\right]^2 d\mu_Z d\mu_{Z^{(i)}} \\
=\ & \frac{1}{2}\int\int \left[\int_{g(Z;\theta_0)}^{g\left(Z^{(i)};\theta_0\right)} \int_{-\infty}^{\infty} f_\eta\left(u - g\right) du dg\right]^2 d\mu_Z d\mu_{Z^{(i)}} \\
=\ & \frac{1}{2}\int\int \left[g\left(Z;\theta_0\right) - g\left(Z^{(i)};\theta_0\right)\right]^2 d\mu_Z d\mu_{Z^{(i)}} \\
=\ & Var\left(g\left(Z;\theta_0\right)\right) < \infty
\end{aligned}$$

where the second-last equality follows from the fact that $\int_{-\infty}^{\infty} f_\eta\left(u - g\right) du = 1$. Since $F_{V;\theta_0}^{-1}\left(\delta\right) \to -\infty$ and $F_{V;\theta_0}^{-1}\left(1-\delta\right) \to \infty$ as $\delta \to 0$,

$$\int \left[\int_{-\infty}^{\infty} \left[F_\eta\left(u - g\left(Z;\theta_0\right)\right) - EF_\eta\left(u - g\left(Z;\theta_0\right)\right)\right] du\right]^2 d\mu_Z$$

converges to 0 as $\delta \to 0$ by the dominated convergence theorem.

The other terms in the expression for $Lim_{G;\delta}\left(\theta_0\right)$ converge to 0 in probability by analogous arguments. ■

C.4 Parametric Bootstrap

Let $\left\{z_j\right\}_{j=1}^{J}$ be a sample of firm characteristics and $\left\{x_i\right\}_{i=1}^{N}$ denote a sample of worker characteristics. The parametric bootstrap for the estimate $\hat{\theta} = \arg\min_{\theta\in\Theta} \hat{Q}_N\left(\theta\right)$ is constructed by the following procedure for $b = \{1, \ldots, 500\}$

1. Sample J firms with replacement from the empirical sample $\left\{z_j\right\}_{j=1}^{J}$. Denote this sample with $\left\{z_j^b\right\}_{j=1}^{J}$.
2. Draw N^b workers with replacement from the empirical sample $\left\{x_i\right\}_{i=1}^{N}$, where $N^b = \sum c_j^b$ and c_j^b is capacity of the j-th sampled firm in the bootstrap sample.

3. Simulate the unobservables ε_j^b and η_i^b.

4. Compute the quantities v_i^b and u_j^b at $\hat{\theta}$ from equations (**??**) and (**??**).

5. Compute a pairwise stable match for the bootstrap sample.

6. Compute $\hat{\theta}_b = \arg\min_{\theta \in \Theta} \hat{Q}_N^b(\theta)$ using the bootstrap pairwise stable match and an independent set of simulations for $\hat{Q}_N^b(\theta)$.