ABSTRACT

Title of dissertation: ESSAYS ON NONPARAMETRIC ESTIMATION OF HETEROGENEOUS CAUSAL EFFECTS

Sungho Noh
Doctor of Philosophy, 2018

Dissertation directed by: Professor Guido Kuersteiner
Department of Economics

My dissertation studies semi- and non-parametric estimation strategies for the distribution of heterogeneous causal effects with applications to labor economics and macroeconomics.

In the first chapter, I propose a nonparametric strategy to identify the distribution of heterogeneous causal effects. A set of identifying restrictions proposed in this chapter differs from existing approaches in three ways. First, it extends the random coefficient model by allowing potentially non-linear interaction between distributional parameters and the set of covariates. Second, the treatment effect distribution identified in this chapter offers an alternative interpretation to that of the the rank invariance assumption. Third, the identified distribution lies within a sharp bound of distributions of the treatment effect. An estimator exploiting the identifying restriction is developed by extending the classical version of statistical deconvolution method to the Rubin causal framework. I show that the estimator is uniformly consistent for the distribution of causal effects.

In chapter two, I apply the nonparametric method developed in the previous chapter to the estimation of heterogeneous effects of displacement on earnings losses. Using the Current Population Survey (CPS) individual-level data from 1996 to 2016, I show that the decline in labor incomes of displaced workers is not only substantial in magnitude compared to their non-displaced counterparts, but also varies significantly within groups characterized by, for example, tenure and educational attainment. I find that displaced workers, on average, lose 19% of their potential earnings while the dispersion of losses among workers is wide. In addition, estimated quantile effects of displacement are more dispersed when the local unemployment rate is higher.

In the third chapter, co-authored with Guido Kuersteiner, we develop a new asymptotic theory for flexible semi-parametric estimators of dynamic causal effects in data with discrete policy interventions. Our framework extends existing theory of propensity score weighted estimators to weakly dependent processes. We show uniform consistency and asymptotic normality by applying a newly-developed asymptotic theory for the series estimator over a non-compact support. The estimator proposed in this chapter captures non-linear and asymmetric impulse response functions that are often difficult to be accommodated in parametric models.

ESSAYS ON NONPARAMETRIC ESTIMATION OF HETEROGENEOUS CAUSAL EFFECTS

by

Sungho Noh

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Guido Kuersteiner, Chair/Advisor
Professor John Chao
Professor Judith Hellerstein
Professor Haluk Unal
Professor Sergio Urzúa

Dedication

To my parents and my partner, Eunhee.

Acknowledgments

I would like to thank my dissertation committee members at the University of Maryland. I would like to express the deepest appreciation to my committee chair Professor Guido Kuersteiner. This research could not be completed without his guidance and continuous support. I am deeply indebted to Professor Sergio Urzúa who has provided extremely useful feedbacks on my work. I am thankful to Professor John Chao who provided invaluable comments on theoretical results and encouraged me to move further. I also thank Professor Judy Hellerstein for her constructive criticism and kind words. I thank Professor Haluk Unal for his service as a committee member and valuable comments during the defense.

I am grateful to all seminar participants at the University of Maryland and Southern Economic Association Annual Conference for their invaluable comments on the earlier version of my work. Especially, I want to thank Professor Katherine Abraham, Professor Borağan Aruoba, Professor Jessica Goldberg, Professor Leslie Turner, and Professor John Shea for their fruitful discussions.

I am also very grateful to my colleagues, Jake Blackwood, Hyung Suk Choi, Joonkyu Choi, Diyue Guo, Bryan Hardy, Xing Hong, Shen Hui, Karam Jo, Keonwoo Kim, Min Kim, Zi-Ying Mao, Jongho Park, Yongjoon Park, Yong Paek, Christian Sanchez, Riley Wilson, Youngjin Yun, and Ying Zheng. Intellectually challenging conversations with them have been a great help for me.

This journey would not have been possible without the support of my family. I am especially grateful to my parents, who supported my emotionally and financially.

I always knew that you believed in me. Thank you to my mother, Insook Son, and my father, Sanggyun Lho. My special thanks go to Eunhee Lee, for always standing by my side and pushing me to my fullest potential. Thank you for celebrating all my achievements and loving me even more through all my failures. I love you.

Table of Contents

List of Tables

List of Figures

Chapter 1: Nonparametric Identification and Estimation of Heterogeneous Causal Effects under Conditional Independence

1.1 Introduction

This chapter introduces a new nonparametric strategy to identify the distribution of heterogeneous causal effects. In past decades, a growing number of papers pointed out possible heterogeneity embedded within the causal effects even after controlling for individual characteristics.[1] With the presence of heterogeneity, interpretation and policy implications that arise from quantitative analyses of causal effects will be substantially different from a case with homogeneous effects. For example, consider a social experiment providing a job training program for displaced workers to support their performance in the job market. If the purpose of the program is to raise the average earnings of workers who participate the job training program, it is sufficient to estimate mean effects and check if the result is positive and statistically

[1]For example, Heckman and Robb (1985), Heckman, Smith, and Clements (1997), and Heckman, Urzúa, and Vytlacil (2006) discuss how the understanding of heterogeneity in causal effects can alter policy implications in the context of empirical policy evaluation with observational data.

significant. However, if the goal is to benefit the maximum number of workers, a better measurement would be the probability of having positive effects for those who have participated in the experiment. As illustrated in this example, understanding the distributional information of the treatment effect is more important, both theoretically and empirically.

The causal effect is formally defined via Rubin (1974)'s potential outcome framework. The outcome of interest, denoted by Y, is supposed to be a partial observation of a pair of potential outcomes. That is, $Y = DY(1) + (1 - D)Y(0)$ where $Y(1)$ is the potential outcome of treated, $Y(0)$ is the potential outcome of controlled, and D is the binary random variable indicating treatment status. As an example, suppose that we are interested in estimating the potential earnings losses of workers after being displaced from their previous jobs. If a worker has been displaced from her previous work, she is assigned to the treated group ($D = 1$). In this case, her observed wage is equal to the potential wage of treated ($Y = Y(1)$). Otherwise, if a worker has been continuously working for at least three years, then she is classified as control group ($D = 0$), and her wage is considered to be equal to the potential wage of the controlled group ($Y = Y(0)$). The causal effect is given by $\Delta = Y(1) - Y(0)$ which is the difference between potential outcomes. Specifically the example of displaced workers, the causal effect of displacement refers to the earnings differences of a worker who has been displaced from her previous match, measured by the changes relative to the counterfactual wage that she would have earned if she were able to continue working in her previous job.

Since the realized value of the causal effects may differ across individuals, a natural way to describe heterogeneity is through the distribution function of Δ. However, without a further identifying restriction imposed, the distribution of causal effects is only partially identified up to a range of distribution functions (Fan and Park, 2010). One caveat of the partial identification approach to the distribution of causal effects is that it is often difficult to propose concrete policy implications from a set of distributions of casual effects, as the resulting set of identified distribution may not exclude unrealistic estimates (Heckman et al., 1997). I consider, instead, the case where the gains from treatment are independent from the potential outcome absent of treatment, conditioning on a set of observable characteristics. More precisely, Δ is assumed to be a random variable that is orthogonal to $Y(0)$ conditional on a vector of covariates.

The conditionally independent gains assumption is comparable with other existing frameworks in two ways. First, it is a natural extension to the case when Δ is constant within a group characterized by X. This type of restriction is commonly used in regression analysis, as illustrated in Section 1.2. However, the constant within-group effect restricts the causal effect Δ to be fully deterministic with respect to the observable dimension. This implies that there is no dispersion within the group characterized by X.

Second, the restriction considered in this chapter has similar intuition to that of rank invariance which is commonly assumed in the context of quantile effects and quantile regression. The quantile effects function has been one of the most popular approaches to describing heterogeneity since the earlier work of Doksum (1974). Chernozhukov and Hansen (2005) study the role of the rank invariance condition for identifying the conditional quantile effect functions. In a later section, I show that both conditionally independent gains and the rank invariance assumption imply positive stochastic dependence between potential outcomes conditional on observable characteristics. The difference is that there is a positive chance of switching ranks between the two treatment statuses under the conditionally independent gains assumption while rank invariance does not allow for such a case. In this sense, the restriction I consider in this chapter has a more flexible interpretation in practice.

The nonparametric estimator that corresponds to the identifying assumption is developed by extending the theory of statistical deconvolution. The intuition is to exploit the natural linearity in the definition of the causal effect. Notice that the potential outcome of the treated can be written as the sum of the baseline outcome and the gains from treatment, such that $Y(1) = Y(0) + \Delta$. If Δ is assumed to be orthogonal to $Y(0)$ conditioning on a set of observable characteristics, it follows that the marginal distribution of $Y(1)$ is equivalent with the convolution of the marginal distributions of $Y(0)$ and Δ conditioning on covariates. Then the distribution of Δ is nonparametrically identified by the inverse Fourier transformation on the ratio of the characteristic functions of $Y(1)$ and $Y(0)$.

The nonparametric strategy presented in this chapter is an extension of the classical theory of statistical deconvolution. Since the early work of Carroll and Hall (1988) and Fan (1991b), asymptotic properties of statistical deconvolution estimators have been applied in various contexts. Fan and Truong (1993), Taupin (2001), Schennach (2004), and Carroll, Delaigle, and Hall (2009) study estimation strategies using deconvolution in regression models with measurement errors. On the other hand, Horowitz and Markatou (1996), Neumann (2007), and Arellano and Bonhomme (2012) apply the nonparametric deconvolution method to panel regression models to estimate the unknown distributions of random effects. In a methodological sense, the closest approach to the method proposed in this chapter is a recent paper by Gautier and Hoderlein (2011). They propose an estimation strategy that utilizes a deconvolution method in the context of selection on unobservables. However, the interpretation of the identified distribution of treatment effects is different. Gautier and Hoderlein (2011) identify the distribution of treatment effects conditional on the individuals who satisfy the monotonicity or, in other words, compliers. On the other hand, this chapter focuses on the distribution identified via comparison of individuals in treated and controlled groups, matched with observable characteristics.

A number of studies emphasize the presence of non-negligible dispersion in the causal effect across individuals and develop estimation strategies to capture heterogeneity. Under the rank invariance restriction, the quantile effect function is a useful object to describe the heterogeneity. Firpo (2007) presents an efficient semi-parametric estimation strategy for the quantile effect. More recently, Rothe (2010) and Chernozhukov, Fernández-Val, and Melly (2013) extend the theory to estimate the effect of policy on changes in counterfactual distributions.

Rank invariance can sometimes be overly restrictive as it fully identifies the counterfactual distribution conditional on the realized outcome. Alternatively, set identification has been considered as a more flexible approach. A major advantage of the set identification approach is that it does not impose a particular stochastic relationship between potential outcomes. Heckman et al. (1997) attempt to identify the range of distributions of the treatment effect by listing every possible match between quantiles of the marginal distributions of potential outcomes which are identified under selection on observables assumption. More recent papers by Chiburis (2010), Fan and Park (2010, 2014), and Fan, Sherman, and Shum (2014) construct tighter bounds for the set of identified treatment effect distributions. Kim (2014) considers improving the identification bound by imposing restrictions on the support of treatment effect distributions in an attempt to sharpen the range of identified distributions.[2] Compared to the partial identification approach, I propose a set of identifying restrictions under which the distribution of the treatment effect is point-identified nonparametrically given a pair of marginal distributions of potential outcomes.

The rest of this chapter consists of the following sections. In section 1.2, I introduce identifying restrictions to recover the distribution of heterogeneous treatment effects and discuss implications in relation to regression models and the propensity score matching estimator. In section 1.3, I establish a nonparametric estimator for the distribution of treatment effects and present useful results for later sections. Section 1.4 provides asymptotic properties of nonparametric estimators for the treatment effect distribution, and the quantile effect function. In section 1.5, I present a Monte Carlo experiment to evaluate finite sample properties of the nonparametric estimators of the mean effect and quantile effects.

[2] For a comprehensive review of the partial identification approach to distributional treatment effects, see Abbring and Heckman (2007).

Notation The Euclidean norm for matrix A is defined as $\|A\| = \sqrt{\operatorname{tr}(A'A)}$. For a square matrix A, $\lambda_{\min}(A)$ is the smallest eigenvalue of A and $\lambda_{\max}(A)$ is the largest eigenvalue of A. For a vector $x = (x_1, \cdots, x_d)' \in \mathbb{R}^d$, ℓ_α-norm is defined as $\|x\|_\alpha = (\sum_{j=1}^d x_j^\alpha)^{1/\alpha}$. For a generic multivariate function $g : \mathbb{R}^d \to \mathbb{R}$, $\mathcal{L}_\eta(\mu)$-norm is defined by $\|g\|_\eta = (\int \|g(x)\|^\eta d\mu(x))^{1/\eta}$ for μ being a probability measure defined over $\mathcal{X}$. In addition, partial derivatives of g are denoted by $\nabla^{\mathbf{s}} g(x) = \frac{\partial^{|\mathbf{s}|}}{\partial^{s_1} x_1 \cdots \partial^{s_d} x_d} g(x)$ where $\mathbf{s} = (s_1, \cdots, s_d)'$ is the index and $|\mathbf{s}| = s_1 + \cdots + s_d$. For a complex number $c \in \mathbb{C}$, the modulus is defined as $|c| = \sqrt{c\bar{c}} = \sqrt{\Re(c)^2 + \Im(c)^2}$ where $\bar{c}$ is the complex conjugate of c while $\Re(\cdot)$ and $\Im(\cdot)$ are real and imaginary parts, respectively.

1.2 Identification of Heterogeneous Effects

Heterogeneity in causal effects is described by the shape and size of the dispersion of Δ across individuals. In a formal notation, let $F_\Delta(\cdot)$ be the distribution function of Δ over the set of possible treatment effect values $\mathcal{T} \subseteq \mathbb{R}$. The possible set of treatment effects is simply defined as a Minkovski difference between supports of potential outcomes. That is, let $\mathcal{Y}_1, \mathcal{Y}_0 \subseteq \mathbb{R}$ be the support of $Y(1)$ and $Y(0)$, respectively. Then $\mathcal{T} = [\inf \mathcal{Y}_1 - \sup \mathcal{Y}_0, \sup \mathcal{Y}_1 - \inf \mathcal{Y}_0]$. In cases where, for example, $\mathcal{Y}_1$ is unbounded, then $\inf \mathcal{Y}_1 = -\infty$ and $\sup \mathcal{Y}_1 = \infty$.

The distribution of Δ can be further decomposed into two parts: heterogeneity within- and between-groups. Let X be a vector of covariates which is a random vector over the support $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ with marginal distribution μ. The heterogeneity embedded in Δ may be further specified with the conditional distribution which is denoted by $F_\Delta(\tau|x) = \text{Prob}(Y(1) - Y(0) \leq \tau|X = x)$ for $\tau \in \mathbb{R}$. Then it is natural to characterize the within-group heterogeneity by the size of dispersion in $F_\Delta(\tau|x)$ such as $F_\Delta(\tau'|x) - F_\Delta(\tau|x)$ for some $\tau, \tau' \in \mathbb{R}$. On the other hand, the between-group heterogeneity is represented by the changes in distribution $F_\Delta(\tau|x)$ such as $F_\Delta(\tau|x') - F_\Delta(\tau|x)$ for $x, x' \in \mathcal{X}$. In sum, identification of the conditional distribution of Δ is the key to uncovering heterogeneity in causal effects.

1.2.1 Selection on Observables

Identification of the causal effect is achieved with a set of restrictions that can eliminate the selection bias in a partially observed outcome. I consider a conventional framework where the treatment status is randomly determined within a group characterized by a set of observable variables. Precisely, the propensity score to treatment is denoted by $p(x) = \text{Prob}(D = 1|X = x)$ for $x \in \mathcal{X}$. The following statements collectively imply that, conditioning on X, D is randomly assigned and there are positive chances of having both $D = 1$ and $D = 0$.

Condition 1.1 (Selection on Observables). *For $j \in \{0, 1\}$, $Y(j) \perp D|X$ almost surely.*

Condition 1.2 (Overlapping). *There exist $\underline{p}, \overline{p} \in (0, 1)$ such that $\underline{p} \leq p(x) \leq \overline{p}$ almost surely.*

Earlier work of Rosenbaum and Rubin (1983) has shown that the two conditions are sufficient for identifying the conditional mean effect by balancing the samples using the inverse of the propensity scores as sampling weights. The idea is that both Conditions 1.1 and 1.2 collectively specify a stochastic model to compare the outcomes across different treatment statuses. In other words, a set of observed outcomes in the control group contains unbiased information on the counterfactual outcome of the treated group as long as the propensity score values of individuals in the two groups are the same (Rosenbaum and Rubin, 1983, Theorem 2). The argument relies on the correct specification of the propensity score function $p(x)$, which is usually unknown to the researcher. Later in section 1.3, I consider the series estimation that is robust to parametric specification of $p(x)$ for a fixed number of covariates X with possibly unbounded support.

Conditions 1.1 and 1.2 do not restrict underlying stochastic models to have a specific form as in linear regression models. However, for illustration, I discuss the role of selection on observables and overlapping conditions for identifying heterogeneous mean effects using a representative linear regression model as shown in the following example.

Example 1.1. Consider the following regression model:

$$Y = D\delta + X'\gamma + U \tag{1.1}$$

where Y is the outcome variable, D is the dummy variable indicating the treatment effect status, and X is a vector of covariates that are relevant for controlling possible endogeneity in selection into treatment. The error term denoted by U satisfies $E[U|X] = 0$ under the conditional independence assumption. The parameter of interest in this case is δ. If δ is a constant, it can be identified via OLS estimator and has a causal interpretation if Conditions 1.1 and 1.2 hold. However, the argument does not hold in general as δ may differ across individuals. Specifically, assume that δ is a random coefficient. It is natural to consider describing its heterogeneity as a function of observed individual characteristics. Let $E[\delta|X]$ be the portion of δ explained by X and denote $\varepsilon \equiv \delta - E[\delta|X]$ for the unexplained randomness. The variance decomposition formula yields the following result:

$$Var(\delta) = \underbrace{E[Var(\varepsilon|X)]}_{within-group\ heterogeneity} + \underbrace{Var(E[\delta|X])}_{between-group\ heterogeneity} \tag{1.2}$$

It can be shown that the between group heterogeneity is identified by running a regression model. Substituting $\delta = E[\delta|X] + \varepsilon$ to the equation (1.1),

$$Y = E[\delta|X]D + X'\gamma + \varepsilon D + U \equiv E[\delta|X]D + X'\gamma + V \tag{1.3}$$

where $V \equiv \varepsilon D + U$. The resulting equation is reasonably well approximated by a reduced form regression model that has multiple interaction terms between X and D. Notice that $E[V|X] = E[\varepsilon D|X] + E[U|X] = E[\varepsilon|D = 1, X]\text{Prob}(D = 1|X) = E[\varepsilon|X]\text{Prob}(D = 1|X) = 0$ as ε is independent of D under the conditional independence assumption. Therefore, the mean function $E[\delta|X]$ is identified by, for example, OLS coefficients associated with the interaction terms between X and D. Then what follows is that the between group heterogeneity is obtained by the variations in $E[\delta|X]$ across X.

1.2.2 Conditionally Independent Treatment Effects

As the parameter of interest is the unknown distribution of treatment effects, I introduce an additional restriction to recover the distribution nonparametrically. Notice that Conditions 1.1 and 1.2 are sufficient to identify the marginal distributions of the potential outcomes, $Y(1)$ and $Y(0)$, from observational data. Then it is easy to show that the characteristic functions of potential outcomes are separately identified as well under the assumption of selection on observables (Proposition 1.1). The characteristic functions of each potential outcome will be ingredients for nonparametrically identified distribution of causal effects via deconvolution formula.

The distribution of the treatment effects is identified only by imposing a stochastic relationship between potential outcomes. A straightforward method is to consider every possible relationship between potential outcomes and list a range of identified distributions of causal effects. However, as pointed out by Abbring and Heckman (2007), this method often results in uninformative bounds for the distribution of causal effects. The rank invariance assumption has been one of the most popular stochastic models to recover heterogeneity in causal effects. In spite of its benefits, however, rank invariance constrains the set of identified distributions by specifying the distribution of counterfactual outcomes as fully conditional on the observed outcome.[3]

I propose an alternative model by relaxing the strong dependence property between potential outcomes into a weaker version. Consider the case when X is the set of information which fully characterizes the heterogeneity in treatment effect

[3] An alternative is to parameterize the degree of dependence in terms of the correlation coefficient between ranks of the potential outcome distributions as in Zimmer and Trivedi (2006). However, a parametric approach not only restricts the set of identified distributions of the heterogeneous treatment effect but also has difficulties in justifying any particular parametric model for the dependence between potential outcomes unless it is based on a structural model with underlying behavioral assumptions.

Δ upto a random component that is independent with the counterfactual outcome $Y(0)$ conditional on X. A simple example is when Δ is additively separable into two parts. Suppose that $\Delta = \delta(X) + \varepsilon$ where $\delta(\cdot)$ is a non-stochastic function that governs the conditional mean of Δ and ε is a random error which is independent of $Y(0)$ conditional on X. The case may also be interpreted as a linear shift model where the distribution of ε governs the size of dispersion in Δ while X determines the conditional mean of Δ. Note that in this example, heterogeneity still exists, as Δ has non-degenerate distribution even after conditioning on X. Instead, the model restricts stochastic relationship between the gains from treatment and the baseline outcome to be fully specified for a given X.

The following is the statement of the restriction discussed above in more general form:

CONDITION 1.3 (Conditional Treatment Effect Independence). $Y(0) \perp \Delta | X$ *almost surely.*

The implication of Condition 1.3 can be illustrated as follows. Suppose that the effect of displacement on earnings is a random object that may have different realizations across workers. The effect in general can be correlated—without conditioning on worker's characteristics—with $Y(0)$ which is the counterfactual wage that a worker would have earned if she hasn't been displaced. Then what Condition 1.3 means is that within a particular group of workers characterized by X, the effect of displacement is independent of their rank within the counterfactual income distribution. For example, among the workers with lower levels of labor market experience ($X = low$), the distribution of potential losses of being displaced is equal regardless of their relative position in the income distribution (distribution of $Y(0)$ conditioning on X). However, workers with lower experience levels may suffer more on average compared to the group of highly trained workers ($X = high$) following the involuntary separation from their previous position.

Example 1.2. As stated in Example 1.1, an additional restriction is required to identify the within-group heterogeneity of the causal effect δ. Condition 1.3 works in this case as follows. First, note that the outcome Y for the treated group ($D = 1$) and controlled group ($D = 0$) are given as follows:

$$Y(D = 1) = \delta + X'\gamma + U \equiv E[\delta|X] + X'\gamma + \eta \tag{1.4}$$

$$Y(D = 0) = X'\gamma + U \tag{1.5}$$

where $\eta \equiv \varepsilon + U$. Note that from equation (1.5), U can be estimated by the residual of the regression of Y on X within a subsample of the controlled group ($D = 0$). Similarly, η can be approximated by the residuals from the regression of Y on the possibly non-linear function of X, conditional on the treated group ($D = 1$). In the context of the linear regression model (1.1), Condition 1.3 implies $\varepsilon \perp U|X$. As a result, there are three conditions that collectively identify the conditional variance of residual heterogeneity ε. First, $\eta = \varepsilon + U$ by construction. Second, both η and U are consistently estimated by the residuals from the regression equations (1.4) and (1.5) which leads to consistent estimates of the conditional variances of η and U. Third, ε and U are independent conditional on X. Then the conditional variance of ε is simply identified by $Var(\varepsilon|X) = Var(\eta|X) - Var(U|X)$ given that

$Cov(\varepsilon, U|X) = 0$.

Note that Condition 1.3 does not imply that the potential outcomes are independent from each other without the existence of any conditioning covariates. In fact, the potential outcomes are stochastically dependent through the definition which is $Y(1) = Y(0) + \Delta$. Here, $Y(1)$ can be interpreted as a linear function of $Y(0)$ shifted by the size of Δ. The definition itself does not give any intuition over the relationship between $Y(1)$ and $Y(0)$ as the shift Δ is realized differently across individuals and possibly correlated with the baseline outcome $Y(0)$ in any direction. Then by Condition 1.3, the stochastic relationship between $Y(0)$ and Δ is restricted within a group characterized by the set of observables. If there is a set of covariates that collectively contain rich enough information to control possible confounding factors which causes the selection bias in Δ, we may assume that the causal effect is *ex ante* homogeneous at least within a narrowly defined group of individuals.

Within a particular group of individuals indexed by $x \in \mathcal{X}$, the stochastic relationship between the two potential outcomes implied by Condition 1.3 becomes more clear. Precisely, Condition 1.3 implies positive stochastic dependence between potential outcomes in the following sense. Let $F_{1|0}$ be the conditional distribution of $Y(1)$ given $Y(0)$ for individuals of the same characteristics $X = x$. That is, $F_{1|0}(y_1|y_0, x) = \text{Prob}(Y(1) \leq y_1|Y(0) = y_0, X = x)$ for $y_1, y_0 \in \mathbb{R}$ and $x \in \mathcal{X}$. Conditioning on X, $Y(1)$ is a linear shift of $Y(0)$ by definition, and the size of the shift is orthogonal to the baseline outcome $Y(0)$. Therefore, the probability of having a larger draw of $Y(1)$ will increase with $Y(0)$. Such property is formally written as follows:

Proposition 1.1. *Suppose that Condition 1.3 is satisfied. Then $Y(1)$ is stochastically increasing (SI) in $Y(0)$ conditional on X–that is, $F_{1|0}(y_1|y_0, x)$ is weakly decreasing in y_0 for all $y_1 \in \mathbb{R}$ and $x \in \mathcal{X}$.*

The intuition of Proposition 1.1 is that it partially identifies the distributional characteristics of counterfactual outcome from the observed outcome of the individuals with same characteristic. For example, consider two workers in the controlled group who both share the same characteristics X. Suppose that each of the two workers has an observed wage of y_0 and y_0', respectively. Furthermore, consider the case $y_0 \leq y_0'$ without loss of generality. Then what Proposition 1.1 implies is that the post-treatment wage of the worker with observed wage y_0 is likely to be smaller than that of the worker with wage y_0'. The intuition is similar to that of rank invariance as in both cases, distribution of the post-treatment outcome conditional on a larger pre-treatment outcome stochastically dominates that of smaller pre-treatment outcome.

In the context of the empirical application with displaced workers, $Y(1)$ denotes the potential wages of displaced workers and $Y(0)$ is that of non-displaced workers. What follows from Proposition 1.1 is that if a worker is supposed to have a higher wage if she has not been displaced previously in her career (larger $Y(0)$), then she is "more likely" to have higher wage among the displaced workers. It is less restrictive than the case of the rank invariance under which the distribution of $Y(1)$ is sharply determined conditional on $Y(0)$. In the context of displaced workers, a worker who earned top 1% wage before separation must be at the top 1% after involuntary separation from her previous job based on the implication of the rank invariance.[4]

1.2.3 Nonparametric Identification of Treatment Effect Distribution

The identification of heterogeneous treatment effects in the case of conditionally independent gains is achieved by applying the theory of statistical deconvolution. Statistical deconvolution theory has been developed to identify and estimate the unknown distribution of a latent random variable when it is observed with classical measurement error with a known distribution (see, for example, Carroll and Hall, 1988; Fan, 1991b,a). Earlier work of Fan (1991b) derives a uniform rate of convergence for a nonparametric deconvolution estimator over a Sobolev class $\mathcal{F}_\alpha(\mathbb{R})$ of functions for $\alpha > 0$ which is defined as following:

$$\mathcal{F}_\alpha(\mathbb{R}) \equiv \left\{ f \in \mathcal{C}(\mathbb{R}) \ : \ \int_{\mathbb{R}} ||\varphi_f(\omega)\|^2 (1 + |\omega|^2)^\alpha d\omega < \infty \right\}$$

where $\varphi_f(\omega) \equiv \int_{\mathbb{R}} \exp(\iota\omega\tau) f(\tau) d\tau$ is the characteristic function of a density f. Later studies extend the theory to the case where measurement error distribution is unknown while researchers have proxy data to estimate it (Taupin, 2001; Carroll et al., 2009).

[4]In a formal notation, difference between Condition 1.3 and rank invariance is given as follows. Suppose that (Y_1, Y_0) and (Y_1', Y_0') are two random draws from the joint distribution of potential outcomes conditional on the same X. With Condition 1.3, there exists positive probability of "reversed rank" in the sense that $\text{Prob}((Y_1' - Y_1)(Y_0' - Y_0) < 0) > 0$. On the other hand, the rank invariance does not allow such possibility as $\text{Prob}((Y_1' - Y_1)(Y_0' - Y_0) > 0) = 1$.

The identification strategy proposed in this chapter is a modification of the recently developed statistical theory under which measurement error distribution is replaced by its empirical counterpart. The intuition is that I exploit the linearity in the definition of the treatment effect and interpret the distribution of observed outcome for controlled group, denoted by $Y(0)$, as proxy data for the measurement error. Then by applying Condition 1.3, it can be shown that the distribution of Δ is nonparametrically identified via deconvolution of the distribution functions of $Y(1)$ and $Y(0)$.

I first show that the characteristic functions of the distributions of potential outcomes are identified. Note that the Conditions 1.1 and 1.2 collectively imply the strong ignorability as in Rosenbaum and Rubin (1983). Their result implies that the marginal distributions of both $Y(1)$ and $Y(0)$ conditional on X are fully identified by the empirical distribution of the observed data conditional on the balancing score $p(X)$. I extend their result to identify the characteristic functions of $Y(1)$ and $Y(0)$.

Lemma 1.1. *Suppose that the Conditions 1.1 and 1.2 hold. Then,*

$$E[\exp(\iota\omega Y(1))|p(X)] = \frac{E[D\exp(\iota\omega Y)|p(X)]}{p(X)}$$
$$E[\exp(\iota\omega Y(0))|p(X)] = \frac{E[(1-D)\exp(\iota\omega Y)|p(X)]}{1-p(X)}$$

Lemma 1.1 states that the characteristic functions of $Y(1)$ and $Y(0)$ are identified as a product of observable quantities. The first part is the exponential transformation of outcome variables of either treated or controlled group. The second part is the inverse of the propensity score function which is bounded away from zero for all $X \in \mathcal{X}$ following the Condition 1.2. The result is similar to the identification strategy discussed in Gautier and Hoderlein (2011) in the sense that they also identify the characteristic functions of potential outcomes by using the inverse propensity score values as sampling weights. The difference is in that the propensity score values in Gautier and Hoderlein (2011) represent the probability of selection into treatment by unobservables whereas, in my framework, $p(X)$ is a sufficient statistic to control observable heterogeneity.

Given the characteristic functions of $Y(1)$ and $Y(0)$ in hand, Condition 1.3 is used to retrieve the characteristic function of the gains from treatment. Recall that $Y(1)$, the potential outcome of the treated, is the sum of the baseline outcome $Y(0)$ and the gains from treatment Δ. That is, $Y(1) = Y(0) + \Delta$. Condition 1.3 implies that $Y(0)$ and Δ are independent conditional on $p(X)$.

LEMMA 1.2. *Suppose that Condition 1.3 holds. Then,* $Y(0) \perp \Delta | p(X)$ *almost surely.*

The argument is interpreted as an extension of the result of Rosenbaum and Rubin (1983) as it also argues that the propensity score $p(x)$ is a sufficient statistic for the identification restriction (Condition 1.3). The result implies that the stochastic relationship between baseline outcomes and causal effects is fully captured by using $p(x)$ as a balancing score. In other words, among the individuals having the same propensity score value, differences in realized effects are not related to their potential outcomes. Therefore, we may aggregate the differences in realized outcomes of treated and controlled groups with the same propensity score weights to infer the distributional characteristics of causal effects. By indexing the individual observations with $p(x)$ instead of X, we may reduce the dimensionality dramatically and consequently, increase the credibility of the identified causal effect distribution.

As a result of Lemma 1.2, it can be shown that the the characteristic function of the causal effect Δ is identified by the ratio of conditional characteristic functions of the potential outcomes. Consider the characteristic function of the potential outcome of treated conditional on $p(X)$. Denote $\varphi_j(\omega|z)$ as the characteristic function of the potential outcome $Y(j)$ for $j \in \{0,1\}$ conditional on $p(X) = z$. That is, $\varphi_j(\omega|z) = E[\exp(\iota\omega Y(j))|p(X) = z]$. As suggested by Lemma 1.2, the conditional characteristic function of $Y(1)$ is written as a product of two characteristic functions, that of $Y(0)$ and Δ, conditional on $p(X)$. More precisely, $\varphi_1(\omega|z) = \varphi_0(\omega|z)\varphi_\Delta(\omega|z)$ where $\varphi_\Delta(\omega|z) = E[\exp(\iota\omega\Delta)|p(X) = z]$ denotes the characteristic function of Δ conditional on $p(X) = z$. If $\varphi_0(\omega|z)$ is assumed to be non-vanishing in the sense that $|\varphi_0(\omega|z)| > 0$ for almost all $\omega \in \mathbb{R}$, the conditional characteristic function of Δ is recovered by the ratio $\varphi_1(\omega|z)/\varphi_0(\omega|z)$. The formal statement of the identification result is as follows:

LEMMA 1.3. *Suppose that 1.1, 1.2, and 1.3 hold. If $\varphi_0(\omega|z) \geq C > 0$ for some constant $C < \infty$ almost everywhere in $\mathbb{R} \times [0,1]$, then $\varphi_\Delta(\omega|z)$ is identified.*

Moreover, the result further implies that the unconditional version of the characteristic function of causal effects could be identified as the following formula:

LEMMA 1.4. *Suppose that Conditions 1.1, 1.2, and 1.3, hold. Then $\varphi_\Delta(\omega)$ is identified by*

$$\varphi_\Delta(\omega) = E\Big[\frac{1-p(X)}{p(X)}\frac{E[D\exp(\iota\omega Y)|p(X)]}{E[(1-D)\exp(\iota\omega Y)|p(X)]}\Big] \tag{1.6}$$

Identification of conditional and unconditional characteristic functions are necessary conditions for the identification of the conditional and unconditional distribution of causal effects, respectively. It is well-known that the characteristic function uniquely defines its corresponding distribution function (Billingsley, 2008, p.365). However, for a more rigorous statement on the identifiability of the causal effect distribution, a formal statement for the class of "non-vanishing" characteristic functions is required. More specifically, the rate of decay of the characteristic functions $\varphi_1(\omega|z)$ and $\varphi_0(\omega|z)$ as $|\omega| \to \infty$ needs to be restricted. It is common to regulate the tail behavior of the characteristic function by uniform bounds that diminish faster than or equal to geometric rates. Such restrictions have been introduced initially by Carroll and Hall (1988) and Fan (1991a,b). Later they were widely adopted in the statistical deconvolution literature including Taupin (2001), Hall and Lahiri (2008), Johannes (2009), and Dattner, Goldenshluger, and Juditsky (2011), among others. Some exceptions are Bonhomme and Robin (2010) and Evdokimov (2010). I also use the regularization kernel as it can incorporate various types of underlying distributions of potential outcomes. The following is the formal statement of the "smoothness" restrictions over the tails of conditional characteristic functions of potential outcomes.

Condition 1.4. *For $j \in \{0, 1\}$, there exist continuous functions $\underline{\Upsilon}_j$, $\overline{\Upsilon}_j : \mathbb{R}^+ \to \mathbb{R}^+$ such that for an arbitrary constant $B < \infty$,*

$$\sup_{|\omega| \leq B, z \in [\underline{p}, \overline{p}]} |\varphi_j(\omega|z)| \leq \overline{\Upsilon}_j(B) \tag{1.7}$$

$$\inf_{|\omega|\leq B, z\in[\underline{p},\overline{p}]} |\varphi_j(\omega|z)| \geq \underline{\Upsilon}_j(B) \tag{1.8}$$

The role of Condition 1.4 is to restrict the tail behavior of the characteristic functions. The inequality (1.7) sets an upper bound that restricts the rate of decay for the conditional characteristic functions at their tails as we are particularly interested in the case when $B \to \infty$. The existing literature on statistical deconvolution considers a particular type of upper bound that is the type of case given in Condition 1.4. Specifically, if $\overline{\Upsilon}_j(B) = \overline{A}_j|B|^{c_j} \exp(-C_j|B|^{\gamma_j})$ for some $\overline{A}_j, C_j, c_j, \gamma_j > 0$, then the characteristic function of $Y(j)$ is said to be "super-smooth." On the other hand, if $\overline{\Upsilon}_j(B) = \overline{A}_j|B|^{-\gamma_j}$ for some $\overline{A}_j, \gamma_j > 0$, then it is said to be an "ordinary-smooth" case. Note that in both super- and ordinary-smooth cases, the characteristic functions $\varphi_j(\omega|z)$ for $j = 0, 1$ tend to zero as $|\omega| \to \infty$. The rate of decay is mostly governed by the parameter γ_j in both cases. For example, if $Y(1)$ is distributed in normal conditional on $p(X)$, it can be shown that $\varphi_1(\omega|z)$ is super-smooth with $\gamma_j = 2$.[5]

The inequality (1.8) defines a lower bound of the conditional characteristic functions of potential outcomes. The intuition behind this constraint is to make the rate of decay for the characteristic function of $Y(0)$ to be slow enough so that the denominator of the ratio $\varphi_1(\omega|z)/\varphi_0(\omega|z)$ does not converge to zero faster than its numerator. It is a necessary condition to achieve identification and smooth

[5]Most of the widely-used parametric distributions fall into the category of either super-smooth or ordinary-smooth distributions. For example, normal, mixture normal, and Cauchy distributions have super-smooth characteristic functions while gamma and exponential distributions have ordinary-smooth characteristic functions.

asymptotic behavior which will be discussed in later sections.

REMARK. Grafakos (2008) establishes a result stating the relationship between the smoothness of the underlying distribution and the rate of decay of its characteristic function. Formally, let f be an arbitrary density function and φ_f be its characteristic function. Suppose that for some integer $\gamma > 0$, $\partial^s f$ exists and are integrable for all $s \leq \gamma$. Then $\varphi_f(\omega)$ is bounded by $|\omega|^{-\gamma} \sup_{|s|\leq\gamma} \|\partial^s f\|$ (Grafakos, 2008, p. 180). Therefore, we may conclude that any random variable of which underlying distribution function is continuously differentiable is at least ordinary smooth. On the other hand, Theorem 3.2.2. in Grafakos (2008) states that given any rate of decay, there exists an integrable function of which the characteristic function has a slower rate of decay than what is assumed. This result suggest that the inequality (1.8) which constrains the lower bound of the rate of decay is not overly restrictive in practice.

From the uniqueness of the inverse Fourier transformation, the density function of Δ, denoted by f_Δ, is obtained by the following formula:

$$f_\Delta(\tau|z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\iota\tau\omega) \frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)} d\omega. \tag{1.9}$$

The expression in (1.9) is well-defined only if the ratio of characteristic functions is integrable over the real line. The integrability is satisfied by regulating the rate of decay of the ratio of characteristic functions in the tails.

I present the formal statement of the identification result of the distribution of treatment effects in two versions. The result for the conditional distribution of Δ

is given below. The proof is in Appendix A.2.

Proposition 1.2. *Suppose that Conditions 1.1–1.4 hold. If* $\int \underline{\Upsilon}_0(x)^{-2}dx < \infty$*, then* $f_\Delta \in \mathcal{F}_\alpha(\mathbb{R})$ *is identified by* (1.9).

The following example shows how the distribution of treatment effects can be identified via Proposition 1.2 in a simple model where potential outcomes are normally distributed.

Example 1.3. Consider a simple two-factor model where the potential outcomes are given by $Y(1) = \Delta + \varepsilon$ and $Y(0) = \varepsilon$ with both Δ and ε having normal distribution. Specifically, assume that $\Delta \sim N(\bar{\Delta}, \sigma_\Delta^2)$ and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ while $\sigma_\Delta, \sigma_\varepsilon < \infty$ denote the finite variances of Δ and ε, respectively. It is easy to verify that the marginal distributions of potential outcomes are given by $Y(1) = \Delta + \varepsilon \sim N(\bar{\Delta}, \sigma_\Delta^2 + \sigma_\varepsilon^2)$ and $Y(0) = \varepsilon \sim N(0, \sigma_\varepsilon^2)$ with the assumption of $\Delta \perp \varepsilon$, which is implied by the conditionally independent treatment effect (Condition 1.3). Let $\varphi_j(\omega)$ be the characteristic function of $Y(j)$ for $j \in \{0, 1\}$. It is well-known that the characteristic function of normal distributions is given by $\varphi_1(\omega) = \exp\big(\iota\omega\bar{\Delta} - \frac{\sigma_\Delta^2 + \sigma_\varepsilon^2}{2}\omega^2\big)$ and $\varphi_0(\omega) = \exp\big(-\frac{\sigma_\varepsilon^2}{2}\omega^2\big)$. Then $\varphi_0(\omega)$ satisfies the condition (ii) of Proposition 1.2 with $g(\omega) = \exp(-C\omega^2)$ for $C \geq \sigma_\varepsilon^2/2$. Taking the ratio between $\varphi_1(\omega)$ and $\varphi_0(\omega)$, the characteristic function of the causal effect is identified by $\varphi_1(\omega)/\varphi_0(\omega) = \exp\big(\iota\omega\bar{\Delta} - \frac{\sigma_\Delta^2}{2}\omega^2\big)$. As it is bounded above by 1, globally convex, and tends to zero as $|\omega| \to \infty$, there exists an absolutely continuous density defined by the inverse Fourier transformation of $\varphi_1(\omega)/\varphi_0(\omega)$. By uniqueness of the inverse

Fourier transformation, the density of Δ is equivalent to that of normal with mean $\bar{\Delta}$ and variance σ_{Δ}^2.

Using the result of Proposition 1.3 as the main ingredient, we may recover other kinds of parameters that are useful to describe the shape and size of heterogeneity of the causal effect Δ. First, the cumulative distribution function of the causal effect is given by the lower integral of the density f_{Δ} over $(-\infty, \tau]$, such as

$$F_{\Delta}(\tau|z) = \int_{-\infty}^{\tau} f_{\Delta}(t|z)dt \tag{1.10}$$

Earlier result of Gil-Pelaez (1951) shows that there exists an equivalent closed-form expression for the distribution of Δ which is explicitly given as a function of the underlying characteristic functions $\varphi_1(\omega|z)$ and $\varphi_0(\omega|z)$. The formula is given as follows:

$$F_{\Delta}(\tau|z) = \frac{1}{2} - \frac{1}{\pi}\int_{0}^{\infty} \omega^{-1}\Im\left(\exp(-\iota\omega\tau)\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\right) d\omega \tag{1.11}$$

I use the expression (1.11) instead of (1.10) as it only has to integrate once over a possibly unbounded support. This expression has advantage especially while constructing a nonparametric estimator as the integral operation should be approximated by numerical integration. The following proposition argues that the distribution of causal effects is identified nonparametrically by equation (1.11). It is a direct consequence of Proposition 1.3 that establishes the identification of the conditional characteristic function of the causal effect Δ.

Proposition 1.3. *Suppose that Conditions 1.1–1.4 hold. If* $\int \underline{\Upsilon}_0(x)^{-2}dx < \infty$, *then the distribution function of the causal effect* Δ *is identified by* (1.11).

The conditional quantile effect function of Δ is identified by the left inverse of the distribution function F_Δ. That is,

$$Q_\Delta(u|z) = \inf_{\tau \in \mathcal{T}} \{\tau : F_\Delta(\tau|z) \geq u\} \tag{1.12}$$

where $\mathcal{T}$ is the support of the heterogeneous causal effect. Notice that, by construction, (1.12) is weakly monotonic in $u \in [0, 1]$. Such property makes the expressions (1.12) to be easily interpretable as the quantiles of the causal effect Δ. For example, if $Q_\Delta(0.5) = 0$, then the treatment is expected to have zero or negative impact with a probability of 50%.

The following statement formally argues that the quantile effect function is identified under the same set of restrictions imposed in the previous results on the density and distribution functions of causal effects.

Corollary 1.1. *Suppose that Conditions 1.1, 1.2, 1.3, and 1.4 hold. Then the quantile effect function of* Δ *is identified by* (1.12).

Remark. The rank invariance assumption is an alternative approach to identify quantile effects in a nonparametric way. Earlier work of Doksum (1974) defines the quantile effect function as the difference between quantile functions of potential outcomes. Specifically,

$$Q_\Delta^{RI}(u|z) = F_1^{-}(u|z) - F_0^{-}(u|z) \tag{1.13}$$

where $F_j^{-}(u|z) = \inf\{y : F_j(y|z) \geq u\}$ is the quantile function of the potential outcome where $j \in \{0, 1\}$. While both F_1^{-} and F_0^{-} are identified only by imposing Conditions 1.1 and 1.2, Q_Δ^{RI} is in general not equivalent with the inverse of the conditional distribution of the treatment effect F_Δ which is given as the expression (1.12). In addition to Conditions 1.1 and 1.2, suppose that the rank of $Y(1)$ is equal to $Y(0)$ conditional on $p(X)$. Chernozhukov and Hansen (2005) show that under this assumption, the quantile treatment effects function (1.13) has a causal interpretation and therefore, expression (1.13) represents the quantiles of the distribution of Δ.

1.2.4 Comparison with other Identification Schemes

Full identification of the heterogeneous treatment effect cannot be achieved without specifying the dependence structure between potential outcomes $Y(1)$ and $Y(0)$. A convenient way to describe stochastic dependence between two marginal distributions is to use bivariate copula. A well-known result of Sklar (1959) shows that for any bivariate distribution, denoted by $F(y_1, y_0)$, there exists a non-decreasing function $C : [0,1]^2 \to [0,1]$ such that $F(y_1, y_0) = C\big(F_1(y_1), F_0(y_0)\big)$ where $F_j(\cdot)$, for $j \in \{0,1\}$, denotes the marginal distribution of $Y(j)$. On the other hand, any real-valued function over $[0,1]^2$ which generates a bivariate CDF by joining two marginal distributions is called (bivariate) copula.

Any bivariate copula is known to be bounded by two prototypical functions referred as Frechét-Hoeffding bounds. More precisely, for any bivariate copula $C(u, v)$, we have $C^L(u,v) \le C(u,v) \le C^U(u,v)$ where $C^L(u,v) = \max\{u+v-1, 0\}$ and $C^U(u,v) = \min\{u,v\}$.[6] The result is used to construct unrestricted bounds on the distributional causal effect. The following result from Makarov (1982) shows that any distribution of Δ identified with a choice of bivariate copula should be bounded

[6] See Nelsen (2007), Theorem 2.2.3.

by the distributions implied by Frechét-Hoeffding bounds.

Lemma 1.5 (Makarov, 1982). *Suppose that Conditions 1.1 and 1.2 hold. Then for any* $z \in [\underline{p}, \overline{p}]$, $F_\Delta(\tau|z)$ *is bounded by*

$$\max\Big\{\sup_{y\in\mathbb{R}}\big\{F_1(y|z) - F_0(y-\tau|z)\big\}, 0\Big\} \le F_\Delta(\tau|z) \le 1 + \min\Big\{\inf_{y\in\mathbb{R}}\big\{F_1(y|z) - F_0(y-\tau|z)\big\}, 0\Big\}$$

Instead of partial identification, the distribution of causal effects is point-identified by choosing a specific copula. The rank invariance condition, considered by Chernozhukov and Hansen (2005) and Firpo (2007) for instance, is one of the possible choices. It turns out that the bivariate copula implied by rank invariance is equal to the Frechét-Hoeffding upper bound (see proof of Proposition 1.4 in Appendix A.2 for details). Hence, given that the marginal distributions of $Y(1)$ and $Y(0)$ are identified, the distribution of causal effects is recovered as follows:

Proposition 1.4. *Suppose that Assumptions 1.1 and 1.2 hold. If* $(Y(1), Y(0))$ *are rank invariant conditional on* X*, distribution of* Δ *is identified by*

$$F_\Delta^{RI}(\tau|z) = 1 + \min\Big\{\inf_{y\in\mathbb{R}}\big\{F_1(y|z) - F_0(y-\tau|z)\big\}, 0\Big\},$$

for all $z \in [\underline{p}, \overline{p}]$.

Notice that the distribution of causal effect under rank invariance condition is equal to the upper bound in Lemma 1.5. Therefore, the distributional treatment effect identified under the conditional independence first-order stochastically dominates the one recovered under rank invariance case.

Corollary 1.2. *Suppose that Assumptions 1.1, 1.2, and 1.3 hold. Then,* $F_\Delta(\tau|z) \leq F_\Delta^{RI}(\tau|z)$ *for all* $\tau \in \mathbb{R}$ *and* $z \in [\underline{p}, \overline{p}]$.

The proof of Corollary 1.2 follows from Lemma 1.5 given that, by Proposition 1.4, F_Δ^{RI} is equal to the CDF implied by the Frechét-Hoeffding upper bound. The result is intuitive as can be seen in Figure A.1, which depicts the level curve of bivariate CDFs of the pair of potential outcomes, under two different assumptions–rank invariance and conditional independence. For illustrative purposes, I consider a simple case where both $Y(1)$ and $Y(0)$ are uniformly distributed over the $[0, 1]$ interval. This implies that the treatment effect $\Delta = Y(1) - Y(0)$ lies in between $[-1, 1]$. The blue shaded area in the left panel shows the area that is equivalent to $F_\Delta^{RI}(\tau)$ for some $\tau \in [-1, 1]$. On the other hand, the area of $F_\Delta(\tau)$ identified under conditional independence is shown as the red shaded region in the right panel. It is easily observed that the distribution function of Δ that is identified under rank invariance dominates that of conditional independence.

Example 1.4. Recall the example with Gaussian distributions for the potential outcomes. In case of $\sigma_\delta = 0$, the distributional causal effect is degenerate at $\delta = \bar{\delta}$. Proposition 1.2 can incorporate with this case as well. If $\delta = \bar{\delta}$ with probability one, both $Y(1)$ and $Y(0)$ follow normal distributions with the same variance σ_ε^2 but with different mean. That is, $E[Y(1)] = \bar{\delta}$ while $E[Y(0)] = 0$. This case fits into the set of restrictions in Proposition 1.2 as shown previously. The ratio between characteristic functions is given by $\varphi_1(\omega)/\varphi_0(\omega) = \exp(\iota\omega\bar{\delta})$, which implies that the density of δ is degenerated at $\delta = \bar{\delta}$. On the other hand, notice that as $Y(1)$ is simply a constant shift of $Y(0)$ at amount of $\bar{\delta}$, $(Y(1), Y(0))$ is now rank invariant. The quantile functions of each outcomes are given by $F_1^-(u) = \bar{\delta} + \sigma_\varepsilon \Phi^-(u)$ and $F_0^-(u) = \sigma_\varepsilon \Phi^-(u)$ where $u \in [0,1]$ and $\Phi^-(\cdot)$ is the inverse of standard normal CDF. By taking the difference, we get the quantile function of Δ as $Q_\Delta(u) \equiv F_1^-(u) - F_0^-(u) = \bar{\delta}$ for all $u \in [0,1]$, which is equivalent to the result obtained by Proposition 1.4.

1.3 Nonparametric Estimation

In this section, I describe how to construct the nonparametric estimator of the distribution of heterogeneous causal effects. I present theoretical results that collectively form the basis for the asymptotic theory which will be discussed in the following section. Nonparametric estimators of the conditional density function f_Δ and distribution function F_Δ are constructed by replacing the characteristic functions in (1.9) and (1.11) with their sample counterparts, respectively. However, a naive substitution of the characteristic functions with their empirical estimates will in general cause ill-posed inverse problem as noted by, for example, Carroll and Hall (1988), Fan (1991b), Taupin (2001), and many others. To alleviate this problem, I consider kernel-based regularization. This method sets a bandwidth and introduces a weighting kernel to regularize possibly irregular behavior of the empirical approximations of conditional characteristic functions at its center ($|\omega| \to 0$) and its tails ($|\omega| \to \infty$). Past studies including, but not limited to, Carroll and Hall (1988), Fan (1991b,a), Taupin (2001), Hall and Lahiri (2008), Carroll et al. (2009), and Dattner et al. (2011) show that regularization over the spectral domain guarantees convergence of the deconvolution estimator to its targeting distribution. In this chapter, I extend the previous results to attain uniform convergence of the deconvolution estimator for all conditional distributions over the covariate domain.[7]

The following is the set of assumptions that will be imposed throughout the discussion of the empirical implementation and asymptotic properties of nonparametric estimators.

Condition 1.5. *Let* $\{(Y_i, D_i, X_i)\}_{i=1}^n$ *be a set of independent samples identically distributed as* (Y, D, X).

Condition 1.6. *Suppose that for* $j \in \{0, 1\}$, *there exists* $\delta > 2$ *such that* $E[|Y(j)|^\delta | X = x] < \infty$ *almost surely on* $x \in \mathcal{X}$.

1.3.1 Estimation of the Propensity Score Function

The first stage is to find a consistent estimator of the propensity score function $p(x)$. The propensity score function $p(x)$ is an unknown function that is possibly nonlinear in x. Thus I consider a series approximation method discussed by Hirano, Imbens, and Ridder (2003) and compute the empirical counterpart of $p(x)$. Precisely, let $\Psi^\kappa(x) = (\psi_1(x), \psi_2(x), \cdots, \psi_\kappa(x))'$ be a vector of basis functions satisfying $(E[\|\Psi^\kappa(X)\|^\eta])^{1/\eta} = O(z_\kappa)$ for some $\eta \geq 1$ with a positive sequence z_κ. Without loss of generality, the vector of basis functions is normalized to satisfy

[7] Some studies on nonparametric estimation for panel regression models discuss technical conditions to achieve uniform convergence of the deconvolution estimator for all conditional distributions. See Horowitz and Markatou (1996), Neumann (2007), Bonhomme and Robin (2010), Evdokimov (2010), and Canay (2011).

$E[\|\Psi^\kappa(X)\Psi^\kappa(X)'\|] = I_\kappa$ (see Appendix A.1 for details). Denote the logistic cdf by $L(u) = \exp(u)/(1+\exp(u))$. The sieve approximation of $p(x)$ is defined by $\widehat{p}_n(x) = L(\Psi^\kappa(x)'\widehat{\gamma}_\kappa)$ where

$$\widehat{\gamma}_\kappa = \arg\max_{\gamma \in \mathbb{R}^\kappa} \sum_{i=1}^{n} \Big(D_i \ln L(\Psi^\kappa(X_i)'\gamma) + (1-D_i)\ln(1 - L(\Psi^\kappa(X_i)'\gamma)) \Big)$$

Previous studies have addressed that there are at least two advantages of using the series approximation for unknown propensity score functions. One is that if $p(x)$ is a smooth function of a continuously distributed covariates X, the series estimator gives a robust and consistent approximation of the true propensity score function (Newey, 1997). Moreover, Hirano et al. (2003) point out that using a consistent estimator of $p(x)$ actually improves finite sample precision of propensity score weighting estimators and achieve the semi-parametric efficiency bound as shown by Hahn (1998).

One limitation in existing theories of series estimation of propensity score functions is that the support $\mathcal{X}$ of the vector of covariates is restricted to a closed and bounded set in $\mathbb{R}^{d_x}$. In Appendix A.1, I discuss how to extend the previous results to be applied for the case when X has possibly non-compact support. It is done by adopting some new asymptotic theories recently developed by Belloni, Chernozhukov, Chetverikov, and Kato (2015) and Hansen (2015).

The following result shows that the series estimator of the propensity score function is consistent for the unknown function $p(x)$. The size of the approximation error, which is measured by a weighted norm, is matched with that of Hirano et al. (2003). The error depends geometrically on the length of basis functions κ_n which increases with respect to the sample size n. However, by controlling the rate of growth for κ_n properly, we may eliminate the approximation error in large samples. The proof of Lemma 1.6 as well as a detailed discussion regarding the series estimation of propensity score functions over a non-compact support can be found in Appendix A.1.

LEMMA 1.6. *Suppose that the Conditions 1.1, 1.2 and that of Lemma A.1.2 hold for* $\eta \geq 1$. *In addition, suppose that (i) there exists a positive sequence* z_κ *such that* $(E[\|\Psi^\kappa(X)\|^\eta])^{1/\eta} = O(z_\kappa)$, *(ii)* $E[\|\Psi^\kappa(X)\Psi^\kappa(X)'\|] = I_\kappa$, *and (iii) for* $\kappa = \kappa_n$, $\kappa_n \to \infty$ *and* $n^{-1/2}\kappa_n^{1/2(s/d_x+1)} = o(1)$ *as* $n \to \infty$. *Then,*

$$\|\widehat{p}_n(X) - p(X)\|_\eta = O_p\left(\sqrt{\frac{\kappa_n}{n}}z_\kappa\right)$$

The result has practical advantages as it allows the model to incorporate the case when the true propensity score is a smooth function over non-compact support. For example, suppose that the past wage profile of a worker is included in a vector of control variables X. It is usually the case that a series of past wages are considered to be a good predictor of the worker's productivity and therefore, the current wage level. Given that the cross-sectional distribution of earnings is usually right-skewed and has substantial mass on the upper tail of the distribution, it is arbitrary to constrain the support $\mathcal{X}$ with any finite value, unlike the lower bound which can be naturally set to be zero. In this case, the result of Lemma 1.6 is especially useful as it accounts possibility that $p(x)$ can be defined over unbounded support.

1.3.2 Statistical Deconvolution with Inverse Propensity Score Weighting

Having a consistent estimator of the propensity score function, the next step is to find sample counterparts of the conditional characteristic functions. In this chapter, I refer the estimator of the conditional characteristic functions as the empirical characteristic functions (hereby ECF).

I consider a non-linear projection of the exponential transformation of outcome variables over the interval $[0,1]$ to approximate the ECFs. From the result of Theorem 1.1, it is known that the characteristic functions of potential outcomes are identified as functions conditional on $p(X)$. Then for all z, both $\varphi_1(\omega|z)$ and $\varphi_0(\omega|z)$ are bounded functions of $\omega \in \mathbb{R}$ over a bounded support $[\underline{p}, \overline{p}]$. Therefore, with some additional restrictions, they are well approximated by a series of nonlinear functions indexed by $z \in [0,1]$.

The nonlinear approximation of conditional characteristic functions is done by projecting the exponential transformation of outcome variables onto a polynomial series approximation of propensity score function. Let $P^r(z)$ be a vector of B-spline basis functions of arbitrary order $r \geq 2$. Specifically, let $\{b_1, b_2, \cdots, b_{r-2}\}$ be equally-spaced nodes in $[0,1]$. The B-spline series is defined as $P^r(z) = (1, z, \max\{z - b_1, 0\}, \max\{z - b_2, 0\}, \cdots, \max\{z - b_{r-2}, 0\})'$ over $z \in [0,1]$. Then the projections of $\varphi_1(\omega|z)$ and $\varphi_0(\omega|z)$ are given as follows:

$$\widehat{\varphi}_{1,n}(\omega|z) = P^r(z)'\Big(\sum_{i=1}^n P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)'\Big)^{-1}\Big(\sum_{i=1}^n P^r(\widehat{Z}_i)D_i\exp(\iota\omega Y_i)\Big) \tag{1.14}$$

$$\widehat{\varphi}_{0,n}(\omega|z) = P^r(z)'\Big(\sum_{i=1}^n P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)'\Big)^{-1}\Big(\sum_{i=1}^n P^r(\widehat{Z}_i)(1-D_i)\exp(\iota\omega Y_i)\Big) \tag{1.15}$$

where $\widehat{Z}_i = \widehat{p}_n(X_i)$ and $\widehat{p}_n(\cdot)$ is the consistent estimator of $p(\cdot)$ as defined in the previous section.

The estimators (1.14) and (1.15) approximate conditional characteristic functions reasonably well only if both $\varphi_1(\omega|z)$ and $\varphi_0(\omega|z)$ are continuous and smooth over a compact set $[\underline{p}, \overline{p}]$ for all $\omega \in \mathbb{R}$. The following condition restricts the true characteristic function to be reasonably smooth at its tails.

Condition 1.7. *Let $B > 0$ be an arbitrary constant. For $j \in \{0, 1\}$, there exists α_j and $\beta^{re}_{j,r}(\omega)$ and $\beta^{im}_{j,r}(\omega)$ such that $\sup_{|\omega| \leq B} |\Re(\varphi_j(\omega|z)) - P^r(z)'\beta^{re}_{j,r}(\omega)| = O(r^{-\alpha_j} B)$ and $\sup_{|\omega| \leq B} |\Im(\varphi_j(\omega|z)) - P^r(z)'\beta^{im}_{j,r}(\omega)| = O(r^{-\alpha_j} B)$.*

Condition 1.7 ensures that the polynomial approximation is reasonably close to the actual conditional characteristic functions. The parameter α_j governs smoothness of $\varphi_j(\omega|z)$ uniformly for both real and imaginary parts. For example, consider a single spectrum ω. Then α_j is related to the smallest number of continuous derivatives of both real and imaginary parts in $\varphi_j(\omega|z)$ with respect to z. See the following example.

Example 1.5. Suppose that the distribution of $Y(j)$ conditional on $p(X) = z$ is normal with mean $\mu_j(z)$ and variance $\sigma^2 < \infty$. Then the characteristic function of $Y(j)$ is given by $\varphi_j(\omega|z) = \exp(\iota\omega\mu_j(z) - \frac{\sigma^2}{2}\omega^2) = \exp(-\frac{\sigma^2}{2}\omega^2)\exp(\iota\omega\mu_j(z))$. By the Euler expansion, $\exp(\iota\omega\mu_j(z)) = \cos(\omega\mu_j(z)) + \iota\sin(\omega\mu_j(z)) \approx (1 - \frac{1}{2!}\omega^2\mu_j(z)^2 + \frac{1}{4!}\omega^4\mu_j(z)^4 - \cdots) + \iota(\omega\mu_j(z) - \frac{1}{3!}\omega^3\mu_j(z)^3 + \frac{1}{5!}\omega^5\mu_j(z)^5 - \cdots)$ of which both real and imaginary parts are continuously differentiable functions with respect to $\mu_j(z)$.

Therefore, $\varphi_j(\omega|z)$ is well approximated by a polynomial series of z if and only if $\mu_j(z)$ is continuously differentiable up to α_j-th order.

The following result is the key ingredient for constructing a consistent estimator of the density of Δ. It formally states that the ECFs (1.14) and (1.15) are uniformly consistent for the actual characteristic functions of the potential outcomes.

THEOREM 1.1. *Suppose that Condition 1.7 and that of Lemma 1.6 hold. In addition, assume that $r = r_n$ satisfies (i) $z_\kappa r_n^{1/2} \kappa_n^{-s/2d_x} = o(1)$ and (ii) $B_n / r_n^{\alpha_j} = o(1)$. Then, for $j \in \{0, 1\}$,*

$$\sup_{|\omega| \leq B_n, z \in [\underline{p}, \overline{p}]} |\widehat{\varphi}_{j,n}(\omega|z) - \varphi_j(\omega|z)| = O_p\left(\sqrt{\frac{\kappa_n r_n^3}{n}} z_\kappa\right) + O_p(r_n^{-\alpha_j} B_n)$$

as $n \to \infty$.

The result shows that the empirical estimators of the characteristic functions given by (1.14) and (1.15) are uniformly consistent with their population counterparts. The rate of convergence consists of two terms as shown in Theorem 1.1. The first term is related to the approximation error caused by the series estimation of the propensity score function in the first stage. It is a function of three parameters: bandwidth κ_n, size of the basis function z_κ, and the order of spline basis r_n. The second term is due to the smoothness assumption (Condition 1.7) imposed on the conditional characteristic functions. This term is bounded with a proper choice of bandwidth B_n and corresponding order of splines r_n.

Given an empirical approximation of the ratio of ECFs, a naive way to construct an estimator of f_Δ is to replace $\varphi_1(\omega|z)/\varphi_0(\omega|z)$ in (1.9) with its sample counterpart. However, simply replacing the population characteristic functions with their empirical counterparts does not generally work because of the ill-posed inverse problem which is common in nonparametric estimators. To alleviate this problem, I regularize the ratio of ECFs using a kernel function over the spectral domain $\omega \in \mathbb{R}$ following the approach discussed by Carroll and Hall (1988) and Fan (1991a). The main idea is to control the rate of decay of the ratio of characteristic functions at their tails using the spectral cut-off.

Given the ECFs (1.14), (1.15) and the weighting kernel $\varphi_\xi(\cdot)$, the nonparametric estimator of the conditional density of heterogeneous causal effect is constructed by replacing the terms in (1.9) with their empirical counterparts. The formula is as follows:

$$\widehat{f}_{\Delta,n}(\tau|z) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp(-\iota\omega\tau)\varphi_\xi(h_n\omega)\frac{\widehat{\varphi}_{1,n}(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)}d\omega. \tag{1.16}$$

Expression (1.16) has the following implication. Suppose that the targeting quantity is the causal effect with a random noise instead of Δ itself. Specifically, consider shifting Δ with the size of ξ, which is a random variable independent with both $Y(1)$ and $Y(0)$. In addition, ξ is penalized by h_n such as $h_n\xi$. The target is then $\Delta_\xi \equiv Y(1) - Y(0) + h_n\xi$ which is essentially a linear shift of Δ by the amount of $h_n\xi$. The characteristic function of the "shifted" causal effect is equivalent to that of the original one multiplied by $\varphi_\xi(h_n\omega)$. The estimator (1.16) is then interpreted as an empirical counterpart of the inverse Fourier transformation of the conditional density of Δ_ξ. Intuitively, the density function of Δ_ξ will converge to that of the actual causal effect Δ as h_n tends to zero.

The ratio between empirical characteristic functions is penalized with a weighting function over the spectral domain. The weighting function consists of two components. One is the characteristic function of an auxiliary random variable, denoted by $\varphi_\xi(\cdot)$. This function serves as a kernel to constraint the region where the ratio of ECFs is integrated. Another element is the bandwidth which is denoted by h_n. This sequence of non-random scalars manages the range of spectrum over which the ratio $\widehat{\varphi}_{1,n}(\omega|z)/\widehat{\varphi}_{0,n}(\omega|z)$ is integrated to gradually expand as the sample size grows.

The formal statement of the conditions that are necessarily required for the weighting kernel and bandwidth:

Condition 1.8. *Let $\varphi_\xi(\omega)$ be a real-valued function over $[-1, 1]$ that satisfies the following conditions:*

(i) $\varphi_\xi(\omega)$ *is continuously differentiable and symmetric around zero*

(ii) $\varphi_\xi(\omega) = 1 + o(|\omega|^m)$ *for some* $m > 2$ *as* $|\omega| \to 0$.

Condition 1.9. *Let h_n be a positive sequence that satisfies the following conditions:*

(i) $h_n = o(1)$,

(ii) $\kappa_n^{1/2} n^{-1/2} z_\kappa h_n^{-3} r_n^{(3-4\alpha_0)/2} = o(1)$, *and*

(iii) $h_n^{-1} r_n^{-\frac{2\alpha_0+\alpha_1}{4}} = o(1)$.

Condition 1.10. *Suppose that the following conditions hold:*

(i) $\overline{\Upsilon}_1(1/h_n) \min\{\kappa_n^{-1/2} n^{1/2} z_n^{-3/2}, r_n^{\alpha_1} h_n\} = o(1)$

(ii) $\underline{\Upsilon}_0(1/h_n) \kappa_n^{3/2} n^{-3/2} z_\kappa r_n^{9/2} h_n^{-1} = o(1)$.

Condition 1.8 imposed on the spectral kernel φ_ξ is equally or less restrictive to that in other types of deconvolution estimators. For example, earlier work of Fan (1991b) restricts φ_ξ to be more than twice differentiable over $\mathbb{R}$ which is more restrictive than what is given in Condition 1.8. Such an assumption is overly restrictive if the parameters of interest are the density and the distribution functions of Δ. Therefore, I only consider the case where a regularization kernel is continuously differentiable over the spectral domain. Johannes (2009) suggests truncating the ratio of ECFs to be zero for all $|\omega| \geq a_n$ with the threshold a_n tends to infinity as the sample size grows. The case is also covered by Condition 1.8 as truncating ECFs over the spectral domain is represented by multiplying with the uniform kernel such as $\varphi_\xi(\omega/a_n) = \mathbf{1}(|\omega/a_n| \leq 1)$.

Conditions 1.9 and 1.10 are more specific to the potential outcome framework. Unlike the classical statistical deconvolution method, the model discussed in this chapter involves the empirical approximation in the first stage of the estimation process to recover the distributions of potential outcomes. Therefore, a proper bandwidth h_n should depend on the parameters that are chosen to estimate the propensity score functions as stated in Condition 1.9. In addition, the asymptotic behavior of nonparametric estimator of the distribution of causal effects depend on the tail behavior of the characteristic functions of potential outcome distributions. Condition 1.10 guides how the bandwidth h_n should be adjusted according to the level of smoothness of the underlying data generating process.

The nonparametric estimator for the conditional distribution of causal effects Δ is constructed in the same manner. Recall that the conditional distribution of Δ

is identified by (1.11) as stated in Proposition 1.3. By replacing the conditional characteristic functions with their empirical counterparts and penalizing with spectral kernel φ_ξ, we have the following formula:

$$\widehat{F}_{\Delta,n}(\tau|z) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \omega^{-1}\Im\left(\exp(-\iota\omega\tau)\varphi_\xi(h_n\omega)\frac{\widehat{\varphi}_{1,n}(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)}\right)d\omega \tag{1.17}$$

where $\Im(\cdot)$ denotes the imaginary part of complex-valued functions indexed by τ. A benefit of using formula (1.17) is that, as discussed in previous section, it involves numerical integration only once when integrating the ratio of ECFs over the spectral domain $\omega \in \mathbb{R}^+$. On the other hand, suppose that we compute $\widehat{F}_{\Delta,n}(\tau|z)$ by integrating $\widehat{f}_{\Delta,n}(\tau|z)$ over $(-\infty, \tau]$. Then the estimation procedure should involve numerical integration twice which makes it computationally intense and less accurate than using the formula (1.17).

Finally, the corresponding estimator for the quantile effect function of Δ is defined as the left inverse of the estimator (1.17). That is,

$$\widehat{Q}_{\Delta,n}(u|z) = \inf_{\tau\in\mathcal{T}}\{\tau : \widehat{F}_{\Delta,n}(\tau|z) \geq u\} \tag{1.18}$$

for $u \in [0, 1]$.

While some of the asymptotic properties of nonparametric estimators (1.16), (1.17), and (1.18) such as the uniform consistency is discussed in the following section, it is generally unknown whether the estimator (1.16) will converge to a well-known type of stochastic process. Therefore, I compute the uniform confidence bounds for (1.16) using the recent result of Chernozhukov, Fernández-Val, Melly, and Wüthrich (2016). In the paper, they propose a bootstrap-based algorithm to obtain uniform confidence bands for functional estimators. Implementation details can be found in Appendix A.3.

1.4 Asymptotic Properties

In this section, I show that the nonparametric density estimator (1.16) is uniformly consistent with the true density of Δ. The rate of convergence is shown to be a function of tuning parameters and it is slower than the conventional $\sqrt{n}$-rate.

With some additional technical assumptions imposed, it can be shown that the estimator (1.16) is uniformly consistent with the parameter of interest, $f_{\Delta}(\tau|z)$. The key is to find a proper balance between tuning parameters that are used in either series approximation of the propensity score function or the range of spectral cut-off. The former one is denoted by κ_n, which implies the length of polynomial basis used to approximate the unknown propensity score function $p(x)$. The latter one is denoted by h_n as in Section 1.3. Other determinants of the speed of convergence are the size of basis functions, denoted by z_{κ}, and the length of basis r_n which controls the approximation error in ECFs given by (1.14) and (1.15). I choose $z_{\kappa} = \kappa^{d_x/2}$, which is obtained by various types of basis functions including, but not limited to, Hermite polynomial and splines. In addition, I set r_n to be proportional to κ_n in the sense that r_n/κ_n will be a constant for all n. This will reduce complexity in the resulting rate of convergence.

The following theorem states the result on consistency of density estimator. The proof can be found in Appendix A.2.

Theorem 1.2. *Suppose that the assumptions in Proposition 1.2 and Theorem 1.1 hold. In addition, $\varphi_\xi(\omega)$ and h_n satisfy Condition 1.8, 1.9 and 1.10, respectively. Then for each $z \in [\underline{p}, \overline{p}]$, $\widehat{f}_{\Delta,n}(\tau|z)$ defined as in* (1.16) *is uniformly consistent for $f_\Delta(\tau|z)$ on $\mathcal{T}$. Moreover,*

$$\sup_{\tau\in\mathcal{T}} \|\widehat{f}_{\Delta,n}(\tau|z) - f_\Delta(\tau|z)\| = O_p\left(\xi_f(\frac{1}{h_n})\right)$$

where $\xi_f(b) = \left(\frac{\kappa_n}{n}\right)^{\frac{3}{2}} z_\kappa r_n^{\frac{9}{2}} b \underline{\Upsilon}_0(b)^{-1} + r_n^{-(2\alpha_0+\alpha_1)} b^4$ as $n \to \infty$.

Theorem 1.2 shows that the estimator (1.16) is uniformly consistent with the true density of causal effects (1.9). Notice that the speed of convergence is no longer equal to the square-root of the sample size. Such a property is typically observed in any type of nonparametric estimators that involves deconvolution of density functions. However, the rate is even slower than, for example, a nonparametric density estimator for the classical measurement error model. For example, Fan (1991b) shows that for the classical deconvolution estimator, optimal rate of convergence is proportional to the log of sample size. The reason for having a slower rate in Theorem 1.2 is that the estimation procedure involves an additional source of error by approximating the conditional characteristic functions with their empirical counterparts. Similar result can be found in Taupin (2001) and Johannes (2009) where they present a semi-parametric estimator that involves the statistical deconvolution formula in the context of measurement error models when the error distribution is approximated by the empirical distribution of proxy observations.

By construction, formula (1.17) yields a point-wise estimator of $F_\Delta(\tau|z)$ that is monotonically increasing as long as the point estimates $\widehat{f}_\Delta(\tau|z)$ are weakly positive for all $\tau \in \mathcal{T}$ and $z \in [\underline{p}, \overline{p}]$. Such a property makes the interpretation of the distribution estimator (1.17) straightforward as it preserves natural properties of the distribution functions. However, in practice, $\widehat{F}_\Delta(\tau|z)$ can exceed 1 which violates the axiom of distribution functions. In a later section, I consider truncating the distribution and report the normalized estimate. That is, $\min\{\widehat{F}_\Delta(\tau|z), 1\}$ instead of $\widehat{F}_\Delta(\tau|z)$ itself. Chernozhukov et al. (2016) argues that truncating the distribution function estimates should not harm consistency while improving finite-sample efficiency by shrinking the distance between $\widehat{F}_\Delta$ and the true distribution F_Δ.

The following result states that the estimator $\widehat{F}_{\Delta,n}(\tau)$ constructed as above is uniformly consistent with the target parameter $F_\Delta(\tau)$. The proof is given in Appendix A.2.

Theorem 1.3. *Suppose that the assumptions in Proposition 1.3 and Theorem 1.1 hold. In addition, $\varphi_\xi(\omega)$ and h_n satisfy Condition 1.8, 1.9, 1.10 and assume that $\overline{\Upsilon}_1(1/h_n)/(\underline{\Upsilon}_0(1/h_n)h_n) = o(1)$. Then, for each $z \in [\underline{p}, \overline{p}]$,*

$$\sup_{\tau \in \mathcal{T}} \|\widehat{F}_{\Delta,n}(\tau|z) - F_\Delta(\tau|z)\| = O_p\left(\xi_F(\frac{1}{h_n})\right)$$

where $\xi_F(b) = \left(\frac{\kappa_n}{n}\right)^{\frac{3}{2}} z_\kappa r_n^{\frac{9}{2}} b(\ln(1/b)\underline{\Upsilon}_0(b))^{-1} + r_n^{-(2\alpha_0+\alpha_1)} b^4$ as $n \to \infty$.

Notice that (1.18) is by construction a monotonic function with respect to u for each $z \in [\underline{p}, \overline{p}]$. The property is directly followed by the monotonicity of the distribution estimator (1.17). Natural monotonicity in $\widehat{Q}_\Delta(u|z)$ gives a practical advantage over the semi-parametric quantile effect function estimators identified via rank invariance assumption. Bassett and Koenker (1982) point out that empirical quantile effect functions can be locally non-monotonic even though such irregular behavior should be negligible as the sample size approaches to infinity. One resolution for the problem is proposed by Chernozhukov, Fernández-Val, and Galichon (2010). They consider re-arranging point-wise estimates of the quantile function in ascending order. This type of additional operation is not necessary for the estimator given as expression (1.18).

Finally, the estimator of the quantile function of heterogeneous treatment effects is shown to be uniformly consistent. Corollary 1.3 is the formal statement for the uniform consistency of the quantile effects estimator (1.18). The result is directly followed by the argument of Theorem 1.3 and the fact that $\widehat{Q}_{\Delta,n}$ converges at the same rate as of $\widehat{F}_{\Delta,n}$.

Corollary 1.3. *Suppose that the Conditions in Theorem 1.3 hold. Then, for each* $z \in [\underline{p}, \overline{p}]$,

$$\sup_{u\in[0,1]} \|\widehat{Q}_{\Delta,n}(u|z) - Q_{\Delta}(u|z)\| = o_p(1)$$

1.5 Monte Carlo Experiment

In this section, I present results from a Monte Carlo experiment under various scenarios to analyze the small sample performance of the nonparametric estimator developed in this chapter. The data generating process (hereby DGP) for the experiment is given as follow. A 2×1 random vector of covariates X is given as $X = (X_1, X_2)'$ where $X_k \sim N(0, 0.25)$ for $k = 1, 2$ with $X_1 \perp X_2$. Treatment status D is assumed to be a binary random variable which is randomly drawn, conditional on X. Specifically, let $D = \mathbf{1}(\alpha + X_1 + X_2 + X_1^2 + X_1 X_2 + \varepsilon \geq 0)$ where $\varepsilon \sim N(0, 0.3)$. Note that parameter α governs the size of the treatment group relative to the control group in the simulated sample. For a given number of total observations, a larger α would yield a relatively smaller number of treated compared to that of untreated. Setting $\alpha = 0$ as a reference, I experiment with different values of α to investigate relationships between the relative size of the treated group versus the untreated group with respect to performance of the nonparametrically estimated distribution of causal effects in terms of its precision and efficiency.

The propensity score function in the experimental model described above takes the form $p(X) = \Phi(-(\alpha + X_1 + X_2 + X_1^2 + X_1 X_2)/\sigma_\varepsilon)$ where $\Phi(\cdot)$ is the standard normal CDF. The potential outcomes are generated by the following DGP:

$$Y(0) = \mu_0 + \delta_{01} X_1 + \delta_{02} X_2 + U_0$$

$$Y(1) = \mu_1 + \delta_{11} X_1 + \delta_{12} X_2 + U_1$$

where $U_0 \perp U_1$ are the idiosyncratic errors and $\mu_1,\ \mu_0,\ \delta_{0k}, \delta_{1k}$ for $k = 1, 2$ are parameters for which values will be set later. It is straightforward to see that the heterogeneous treatment effect is given by

$$\Delta \equiv Y(1) - Y(0) = \mu_1 - \mu_0 + (\delta_{11} - \delta_{01}) X_1 + (\delta_{12} - \delta_{02}) X_2 + U_1 - U_0 \qquad (1.19)$$

The model in (1.19) is simple enough to distinguish within- and between-group heterogeneity into two additively separable terms. Between-group heterogeneity is fully captured by the mean effect which is a linear function of X_1 and X_2. Within-group heterogeneity, on the other hand, is represented by the dispersion in the two error terms, U_1 and U_0. The key is to estimate not only the mean effect, which is parameterized by μ_1, μ_0, and δ_{jk} for $j = 0, 1$, $k = 1, 2$, but also the shape of heterogeneity represented by the distribution of $U_1 - U_0$. For this purpose, I target not only the mean but also the variance of Δ to evaluate how precisely the nonparametric estimator developed in this chapter captures the shape of heterogeneity in causal effects. Actual values for both mean and variance of the heterogeneous causal effect are set to be equal to 1.

I consider three different choices for the distributional model of the heterogeneity in the causal effect: normal, Laplace, and exponential distributions. The case with normal distribution is a benchmark where the shape of the distribution of Δ is symmetric around its mean and decays at an exponential rate. The Laplace distribution also represents a symmetric distribution while the density is more concentrated around its mean and decays slower than the normal distribution. Finally, exponential distribution represents a case when the distribution of causal effect is asymmetric and right-skewed. In all three distributional models, parameters are adjusted so that the variance of $U_1 - U_0$ is equal to 1.

To begin with, I present the result of distribution estimates as Figure A.2 in Appendix A.4. It consists of 12 panels which is a combination of four sample sizes (300, 500, 1000, 2000) and three different values of α (0, 0.15, 0.3). In every case, the underlying distribution of the random component $U_1 - U_0$ is assumed to be normal with mean and variance equal to one. In each panel, I overlay distribution estimates for each of 1000 Monte Carlo iterations in black solid lines. As illustrated in the figure, there is a larger dispersion of distribution estimates at tails of distribution.

This may suggest that even though the nonparametric estimator (1.17) is uniformly consistent with its true distribution as shown in Theorem 1.3, the rate of convergence can be uneven between the tails and center of the distribution.

Estimated quantile effect functions of heterogeneous causal effect under various scenarios are presented as figures in Appendix A.4. Figure A.3 shows the result when the true distribution of Δ is assumed to be normal. The figure also contains $4 \times 3 = 12$ panels as in the case of distribution estimates shown in Figure A.2. In each panel, a blue solid line represent the point-wise average of quantile effect estimates among 1000 experiments. Red dashed lines show the 95% range of quantile effect estimates for the same number of repetitions. The result shows that for all 12 cases, the true quantile function, which is depicted with a black solid line, stays in the confidence interval and close to the average estimates. However, one may observe that there are small upward/downward biases at the bottom/upper quantiles. This is mainly due to the fact that distribution estimator is less precise at its tails as shown by Figure A.2.[8]

Figures A.4 and A.5 present results from the same experiment as seen by Figure A.3 with different distributional assumptions. Estimates in Figure A.4 are obtained with samples generated by assuming that Δ is drawn from the Laplace distribution. While the difference is not significant, we may find the quantile function estimates are downward/upward biased at the bottom/upper half. This may suggest that

[8] Chernozhukov (2005) finds that the asymptotic distribution of quantile function estimators are different at extreme quantiles compared to the center. Although his result is based on a different setting, similar argument can be made for larger deviations at extreme quantiles in this exercise.

precision of the nonparametric estimator is lower when the underlying distribution of heterogeneous causal effects decays slower than the normal distribution. However, the size of the bias becomes smaller as the sample size grows. The bias seems to be larger in Figure A.5 where the true distribution of Δ is assumed to be the exponential distribution. In this case, quantile effects are biased upwards in general, although the true quantile effect function is within the 95% interval.

To further investigate the overall precision of the nonparametric density estimator (1.16), I compute Mean Integrated Squared Error (MISE) of the estimates. The formula for the MISE of density estimator is given as

$$MISE(f) \equiv \int_{\mathcal{T}} (\widehat{f}_{\Delta,n}(\tau) - f_{\Delta}(\tau))^2 f(\tau) d\tau$$

Similarly, precision of the distribution estimator (1.17) is also measured by the MISE of $\widehat{F}_{\Delta}$ which is obtained as follows

$$MISE(F) \equiv \int_{\mathcal{T}} (\widehat{F}_{\Delta,n}(\tau) - F_{\Delta}(\tau))^2 f(\tau) d\tau$$

Table A.1 shows the resulting MISE under various scenarios. The estimator matches the true distribution function well in the case of normal and Gumbel distribution. However, the experiment with the Laplace distribution may suggest that the estimated density loses precision if the underlying distribution has heavy tails at both ends. In addition, the first three columns suggest that the overall estimation errors in $\widehat{f}_{\Delta,n}(\cdot)$ tend to decrease as the relative size of control groups to that of treated groups increase.

1.6 Conclusion

In this chapter, I introduced a flexible nonparametric approach to estimate the heterogeneous causal effects of unknown form in a partial observations framework. Identification of the causal effect is achieved by introducing conditional independence of the treatment effect assumption, the case in which gains from treatment are independent from the baseline outcome, conditional on a set of observable characteristics. I discussed that the independence between the baseline outcome and the gains from treatment imply that the counterfactual outcome distribution, conditional on observable characteristics is stochastically increasing with respect to the observed outcomes. This property relaxes the restrictive implication of the rank invariance by allowing the possibility for ranks to be different between pre- and post-treatment.

Chapter 2: Nonparametric Estimation of Heterogeneous Earnings Losses of Displaced Workers

2.1 Introduction

In this chapter, I study the heterogeneity in earnings losses of displaced workers using a newly-developed nonparametric estimation method. Recent empirical studies show that the effect of displacement on (permanent) income is substantially different across time and individuals.[1] I revisit this question focusing on the heterogeneity of earnings losses due to displacement within worker groups. Using the Current Population Survey (CPS) individual-level data from 1998 to 2016, I show that the decline in labor incomes of displaced workers is not only substantial compared to their non-displaced counterparts, but also significantly dispersed within groups characterized by observable characteristics such as tenure. A few examples are the followings. While less tenured workers lose more on average, the distribution of the effect is

[1]For example, Davis and von Wachter (2011) estimate workers lose on average 40/23% in recessionary/expansionary periods compared to non-displaced counterparts. On the other hand, Abraham, Haltiwanger, Sandusky, and Spletzer (2016) find negative impacts of unemployment duration on post-displacement earnings even after controlling for work histories.

more dispersed among more tenured workers. In addition, the size of within-group heterogeneity of earnings losses is larger when the unemployment rate is high.

Previous researches find that workers involuntarily separated from her previous position suffer substantial loss in their potential earnings. Earliest results can be found in, for example, Ruhm (1991) and Jacobson, LaLonde, and Sullivan (1993). Ruhm (1991) is, to my knowledge, the first to formally estimate the amount of permanent loss in incomes due to displacement using survey data. On the other hand, Jacobson et al. (1993) use administrative data from Pennsylvania to estimate on average 25% loss in permanent income of displaced workers compared to those who were able to continue in their position throughout the sample period. While the size of earnings loss may differ in each study, their main argument is supported by various subsequent studies using both administrative data (Couch and Placzek, 2010; Davis and von Wachter, 2011) and survey data (Carrington, 1993; Stevens, 1997; Farber, 2011, 2015).[2]

I revisit the question using matched Monthly CPS data from 1996 to 2016. CPS has comprehensive demographic information of workers which is used to control for heterogeneity in individual-specific returns to tenure and construct a fair comparison group to displaced workers. Especially the Displaced Workers Survey

[2] Couch and Placzek (2010) finds above 10% loss in earnings after displacement using the administrative data of Connecticut in 2000s. They argue that Jacobson et al. (1993)'s result may be biased upward due to industrial reform in Pennsylvania during 1980s. Farber (2015) estimates the effect using merged CPS data from 1984 to 2010, while the information related to employment status is augmented from DWS supplement.

(DWS) supplement provided in every January or February samples contain detailed information on displaced workers including the reason of separation and earnings prior to displacement.[3] Such information provides better identification strategy by restricting both treated and controlled group to be the workers having similar profiles.

Identification of the causal effect of displacement is based on the assumption that worker-specific draw of returns to general experience is conditionally independent to both observed and unobserved factors in potential earnings. The intuition follows from the argument by Kletzer (1989) that in case of a permanent job loss, bias in relationship between post-displacement earnings and pre-displacement tenure is small as the match quality is randomly drawn. I formalize the intuition and establish a random coefficient model featuring the following properties. First, workers who have been re-employed following displacement draw match-specific returns to tenure from an unknown distribution which is the objective of interest. Second, worker-specific returns to tenure is conditionally independent to unobserved match quality given her pre-displacement earnings. Then I show that the distribution of counterfactual earnings losses following the displacement is nonparametrically identified by comparing potential wage distribution of displaced workers to that of non-displaced workers with similar observed characteristics.

This chapter is to shed new light on the topic of permanent income losses of

[3] DWS consists of supplemental questions asked to individual participants for every even numbered years. Additional questions have been asked to eligible survey participants for every January surveys after 2000 and February surveys until 1998.

displaced workers by looking at the distributional characteristics of earnings losses. Since the earlier work of Ruhm (1991), many studies have shown a significant size of the causal effect of temporary unemployment on wage profile especially on whether the effect is permanent. While some studies address the differences in the earnings loss by demographic groups of workers (Jacobson et al., 1993; Davis and von Wachter, 2011), little has been known about the heterogeneity in earnings losses within each group. Such disparity may have contributed to the rise in wage inequality in the U.S. economy which has been well documented by, for example, Autor, Katz, and Kearney (2008) and Acemoglu and Autor (2011). Especially after 2008, mass layoff events has been occurred more frequently than previous years. In a broader context, this exercise aims to identify the impact of labor market dynamics on wage distribution in a less parametric framework than existing work.

Post-displacement earnings losses largely due to the loss of human capital which is not transferable. In particular, size of the wage loss following the displacement is much worse if larger portion of human capital is specific to the previously-held job (Kletzer, 1989). The argument is supported by a number of subsequent studies including, Jacobson et al. (1993), Carrington (1993), Neal (1995), and Stevens (1997) to name a few. However, there are disagreements on where an individual worker's accumulated skills are the most attached to. Jacobson et al. (1993) and Neal (1995) argue that the skills are mostly attached to their industry classified at various levels and therefore, switching across different industries may result in larger earnings loss after displacement. (Carrington, 1993) show that losses in job-specific skills is the most relevant source of the effect of displacement while general labor market condition does not have significant effect.

The question of how to correctly measure the effect of displacement is closely related with problem of identifying the returns to tenure. Difficulty in estimating the returns to tenure lies in the fact that workers endogenously decide to accumulate their human capital. Earlier theories on human capital accumulation process has been argued that there exists positive returns to tenure and experience (Becker, 1964; Mincer, 1974; Mortensen, 1978; Mincer and Jovanovic, 1979). However, empirical studies have also shown that the causality can be reversed since higher match quality could lead to a position that has large wage and longer lifespan at the same time (Abraham and Farber, 1987).

Studies have adopted various approaches to control unobserved heterogeneity in wage process. One way is to introduce proper instrumental variables and implement two-stage estimation procedure where first stage regression is to control potentially confounding factors between job-specific returns and potential wage offers. Altonji and Shakotko (1987) estimate on average 24% larger wage for 10 years of tenure in the same position compared to newly hired workers. Subsequent studies including Topel (1991) and Altonji and Williams (2005) found that the effect can be smaller. More recently, Dustmann and Meghir (2005) use control function approach under exclusion restriction which assumes that the endogeneity in unobserved heterogeneous skill levels are independent with age if the sample is restricted to young workers.

On the other hand, properly constructed sample of non-displaced workers can help identify the counterfactual earnings loss by matching wage profiles across displacement status. Jacobson et al. (1993) use a comprehensive payroll data of Pennsylvanian workers to find that the displaced workers on average suffer wage loss at least 3 Months prior to separation. From this evidence, we may argue that without controlling workers' pre-displacement earnings, the effect of losing job-specific capital will be under-estimated. Motivated by their approach, I estimate the causal effect of displacement by matching earnings distribution of displaced workers with its counterpart of non-displaced workers via observable characteristics including past wage levels. However, due to the limited sample, I can only control for the wage of previous year which may not reflect the decline of wage just before the date of displacement.

I contribute to the literature by estimating the distribution of earnings losses following the displacement. I begin with a theoretical model of wage process featuring heterogeneous returns to tenure as a random coefficient. A nonparametric estimation strategy developed in Chapter 1 is implemented. The theory extends distribution estimators for random coefficient models as in, for example, Horowitz and Markatou (1996), Evdokimov (2010), Canay (2011), and Arellano and Bonhomme (2012). Major difference is that the observed wages are weighted by estimated propensity scores to recover the distribution of potential earnings of displaced and non-displaced workers. Under this assumption, the estimated distribution of the difference between potential earnings has causal interpretation.

The rest of the chapter consists of the following sections. Section 2.2 shows summary statistics from the sample used in the empirical analysis focusing on earnings dynamics, demographics, and decision of displaced and non-displaced workers.

In section 2.3, I present a simple model for potential earnings dynamics and propose a set of identifying assumptions along with their implications. In section 2.4, I describe nonparametric estimation method implemented to estimate the distribution of earnings losses and present the result. Section 2.5, I conclude.

2.2 Data

I use the series of January samples of the Monthly Current Population Survey (CPS) from 1996 to 2016. The major benefit of using CPS data is that I can exploit a rich set of information from additional questions available from the Displaced Workers Survey (DWS) and Job Tenure (JT) supplements. The DWS is collected biannually and asks detailed questions regarding the labor market status of survey takers, such as whether the person has been displaced within the last three years, the reason for displacement. In the same survey Months, JT supplemental questions are also collected. The supplemental questions has information on previously held job which includes the type, industry, and length of tenure. The survey has been widely used in the context of estimation of earnings losses of displaced workers because of its natural advantages. Some of the noticeable examples are Ruhm (1991), Carrington (1993), Neal (1995), and Farber (2011, 2015).

There are limitations of using CPS to identify the causal effect of displacement. One of which is that by the structure of survey method, detailed history of wage profile of an individual can only be tracked back to one and a half years. For this reason, growing number of studies use administrative data for their rich set of information on track history of earnings and matches between employer and its employees. Jacobson et al. (1993), to my knowledge, is the first study to address importance of constructing well-represented control group using a comprehensive observations of Pennsylvanian workers to estimate causal effect of displacement on earnings. Dustmann and Meghir (2005) use administrative data from Germany to estimate the returns to experience and tenure while controlling heterogeneity bias via two-stage method. Davis and von Wachter (2011) use the data of U.S. taxpayers which contains extensively detailed history of earnings. Abraham et al. (2016) construct a novel dataset by merging Longitudinal Employer-Household Dynamics (LEHD) and CPS via unique individual identifiers to incorporate detailed information on both employer-employee matches and worker characteristics. More recently, Heredia, Rucci, Saltiel, and Urzúa (2017) estimate heterogeneous returns to experience and tenure across different labor market institutions using administrative data from Brazil and Chile.

I address this issue in two ways–one from a theoretical perspective and the other one in empirical approach. For a theoretical resolution, I build a simple model to describe the stochastic process for potential wage. The model features random coefficients that represent worker-specific returns to tenure which is assumed to be different across workers in both observed and unobserved dimensions. It is shown that the unknown distribution of returns to tenure is a sufficient statistic to describe the heterogeneity in earnings losses by displacement. The key to identify the distribution of earnings losses is to assume that it is independent to the unobserved match quality for a new job conditional on the pre-displacement wage. Under this assumption, I show that the distribution of earnings losses is nonparametrically identified by matching the wage distribution of displaced workers to that of non-displaced workers having similar past wage profile (refer Section 2.3 for details).

The treatment group consists of workers who have been displaced within one year from the survey date as to match with the sample of non-displaced workers whose wage of previous year is available. I only consider workers displaced because of the plant closure which is considered to be a fairly exogenous reason and does not necessarily reflect workers' innate ability such as skill level and productivity. However, it is still possible to have endogenous selection in samples as some plants may suffer from decrease in productivity at the aggregate level. Therefore, I introduce regional and industry dummies while estimating the propensity score to displacement to control for heterogeneity across region and industry.[4] The region is controlled at the state level and industry is classified by 3-digit NAICS code.

The control group is constructed by non-displaced workers who have been continuously working within the same period when displaced workers have been unemployed. I rule out the case if a worker has switched her job or temporarily unemployed within a year from the survey date. Such information can be verified from JT supplement in which workers are asked if they have been working with the same employer since last year. The sample I use in this chapter can be potentially

[4]This is related with the concern raised by Couch and Placzek (2010). They argue that the estimated earnings losses in Jacobson et al. (1993) could be upward biased as Pennsylvania in late 1980s have suffered economic downturn in overall which leads to larger loss in earnings and smaller re-employment rates.

biased. As pointed out by Jacobson et al. (1993), workers may start to lose their wage at least 3 Months prior to displacement due to financial distress that the firm has. Therefore, the estimated average earnings loss presented in this paper is possibly over-estimated by not fully incorporating the heterogeneity in firm-specific characteristics.

The sample used in this exercise consists of about 779,000 worker-year observations. The outcome variable of interest is log weekly earnings. The numbers are based on self-reported amount on the wage a worker earned in the past week. Therefore, earnings reflects not only the hourly wage but also the number of hours worked within one week.[5] The reported wages are normalized by 2010 dollars using the personal consumption expenditure price index. I use inverse hyperbolic sine (IHS) transformation of the real wage so as to approximate the log wage without losing zero entries.[6] The treatment variable is defined as a binary indicator which takes 1 if a worker has been displaced at least once within one year. Specifically, only

[5] I abstract the case in which previous displacement status could alter the decision of working hours and assume that hours do not vary between the current and previous jobs. This is a potentially important question if a large portion of displaced workers move from full time to part time jobs. I exclude such cases as there are only small number of observations relative to the entire sample.

[6] IHS transformation of a variable y is defined as $IHS(y) \equiv \log(y + \sqrt{1 + y^2})$. A benefit of using IHS instead of log transformation of earnings data is that IHS transformation can naturally incorporate zero entries by $IHS(0) = 0$ while overall the function well approximates the log transformation. This is useful especially for the displaced workers study as the data naturally contains a large number of zero earnings.

workers displaced from their previous positions due to a plant closure are included in the treated group. By doing so, I can further eliminate potentially confounding factors in treated and controlled groups.

Throughout the sample period, gaps between the average earnings of displaced and non-displaced workers have been consistently large. To illustrate the annual variations in wage gaps between displaced and non-displaced workers, I plot within-year averages of log wages among workers in both groups for every two years from 1996 to 2016. In Figure B.1, the red solid line indicates the evolution of average weekly wages of non-displaced workers while blue dashed line represents that of displaced workers. Each data point reflects the within-year average of self-reported weekly wages in the sample. I also depicted with vertical lines the size of standard errors for each within-year average. For the last 20 years, differences between the average wages of both groups has been persistent. Displaced workers earned about 10% less than non-displaced workers on average while the size of wage gap varies across time. For example, average loss by displacement was larger in 2012 and 2014 compared to other years in the sample.

One possible way to explain the observed wage differences between displaced and non-displaced workers is that the chance of finding a new job as good as the pre-displaced position varies across time and individuals. For example, the average wage level among displaced workers has been steadily decreasing since 2008. This may have resulted in a lower rate of re-employment among displaced workers. I briefly investigate this conjecture by summarizing the labor market decision of displaced workers by survey year.

Table B.3 summarizes the labor market status of displaced workers for each year. The labor market status of displaced workers are classified into three groups. First are the workers who currently-employed, which includes workers who are employed either full-time or part-time as of the date of survey. While every year about 70% of displaced workers found jobs, the numbers vary across time. Specifically, there were significant drops in 2010 and 2012. This may be a reason that the average wage gaps between displaced and non-displaced workers widened following the recession in 2008 as shown in Figure B.1. Workers who have not been re-employed after 2008 are still in the labor market. As shown in the second and third columns in Table B.3, the fraction of displaced workers who were unemployed in 2010-2014 is larger compared to previous years while that of workers who are out of the labor force was relatively stable throughout the sample period. This evidence suggests that changes in macroeconomic factors mostly affect the unemployment rate of displaced workers and do not alter the decision to stay out of the labor force.

One potential concern is that there could be other factors that affect the decision of workers following displacement. In tables B.4 and B.5, I present the fraction of worker who choose to find a new job, be unemployed, or stay out of the labor market after being exogenously separated, according to their educational attainment and age, respectively. As can be seen in table B.4, the ratio of workers employed following displacement differ by education levels. The ratio of workers who stay unemployed and out of the labor market is largest among high school dropouts. This suggests that both quality and frequency of potential job offers that a displaced worker would face highly depends on her educational attainment. On the other hand, the decision of displaced workers does not differ significantly according to age of displacement except if a worker is over 55. Older workers tend to exit the labor market rather than finding a new job after being displaced, according to the evidence shown in table B.5. Based on such observations, I characterize worker types by their educational attainment and not age to capture the observable heterogeneity in potential wage offers.

Table B.6 summarizes observed demographic characteristics of workers by their displacement status. I compare averages of individual characteristics including age, sex, race, educational attainment as well as job-related variables, such as the length of tenure and weekly wage between displaced and non-displaced workers. Pre-determined demographic characteristics are largely similar between two groups. On average, workers in both groups are 38-39 years old, about 70% are white, and about 26% of workers have a bachelor's degree or higher. Major differences between displaced and non-displaced workers are found in job-related characteristics. Displaced workers have significantly fewer years of tenure in the current position as the sample restricts focus to workers who have been separated from their previous position within the past three years. Displaced workers earn 50 dollars fewer per week on average compared to the non-displaced workers. Such a difference may imply that there exists substantially negative effects of displacement.

2.3 Theoretical Framework

In this section, I present a simple model to describe the wage process of an individual worker. The model will provide insights for the set of identifying conditions that are necessary to specify the sources of heterogeneous effects of displacement on earnings losses. Key features of the theoretical framework introduced in this section are the following. First, wage difference between displaced and non-displaced workers is assumed to be a random variable. Second, the distribution of wage difference is independent with the previous match quality, conditioning on a set of observed characteristics.

2.3.1 Income Process

Consider a simple model to describe accumulation of human capital of an individual worker. Denote h_{ijt} to be the worker i's stock of human capital at period t given that she has been matched with firm j in period t. The stock of human capital is assumed to consist of three components. The first is the transferable capital which can be carried over different jobs. Transferable capital is often represented as a function of general experience (for example, Dustmann and Meghir, 2005). However, due to the limited information available in CPS and its supplements, I assume that it is a function of age and education level, while leaving the functional form to be flexible. The second part is the firm-specific component that cannot be transferred across different positions in different firms. Non-transferable skill level is assumed to be a function of the firm-specific experience, or seniority, denoted by s_{ijt}. Lastly,

μ_{ijt} denotes the match-specific capital which evolves independently with the tenure within a job.

I assume that the log of human capital evolves following the two features: (1) increases non-linearly according to the seniority s_{ijt}, and (2) generic and job-specific components are additively separable. The form specifically motivated from Dustmann and Meghir (2005) and Heredia et al. (2017).[7] Specifically, the log of human capital is assumed to be governed by the following functional form:

$$\ln h_{ijt} = g(a_{it}|\phi_i) + f(s_{ijt}|\phi_i) + \rho_{ijt}s_{ijt} + \mu_{ijt} \tag{2.1}$$

where a_{it} is the age of worker i at period t and ϕ_i is a set of time-invariant characteristics including sex and educational attainments. In equation (2.1), log of human capital is decomposed largely into two parts. Term $g(a_{it}|\phi_i)$ is the component corresponding to transferable human capital while $f(s_{ijt}|\gamma_i) + \rho_{ijt}s_{ijt}$ is the portion of human capital which is specific to the firm. Note that the returns to both general and firm-specific experiences consist of two parts. Terms $g(a_{it}|\phi_i)$ and $f(s_{ijt}|\phi_i)$ represent the deterministic component of the returns to general and firm-specific experiences, respectively. Heterogeneity in returns to experiences are governed by worker types which is reflected by the specification that both functionals f and g are conditioned on the individual-specific parameter ϕ_i. While functional forms for f and g will remain unspecified in the fully nonparametric approach, I consider quadratic functions as a reference for benchmark regressions presented in Appendix B.2.

[7]Equation 2.1 simplifies the form assumed in Dustmann and Meghir (2005) by excluding sector-specific components in human capital accumulation process.

Besides the deterministic component in returns to experiences, there exists stochastic components in (2.1). Term ρ_{ijt} denotes the unobserved heterogeneity in returns to general and firm-specific experiences, respectively. In a later section, I assume that ϕ_i is fully specified by observable characteristics while ρ_{ijt} is assumed to be random coefficients. These random coefficients become the source of within-group heterogeneity in returns to experiences among the workers with the same observed characteristics.

Suppose that the observed wage of worker i at period t is given as $w_{ijt} = r^e_{it} h_{ijt} \exp(\varepsilon_{it})$. The term r^e_{it} represents the market returns to human capital of which heterogeneity across workers is fully captured by one's education level e. Term ε_{it} is an idiosyncratic error. Taking log transformation on both sides, we have the following specification for the stochastic process of observed wage:

$$\ln w_{ijt} = \ln r^e_{it} + g(a_{it}|\phi_i) + f(s_{ijt}|\phi_i) + \rho_{ijt} s_{ijt} + \mu_{ijt} + \varepsilon_{it} \tag{2.2}$$

There are two sets of parameters that jointly formulate the heterogeneity in returns to experience. The first is the worker type ϕ_i which governs the deterministic component of the earnings growth. I assume that worker type is fully characterized by the observed demographic information that has been pre-determined before entering the labor market. This includes sex, race, and education level. The second is the unobserved heterogeneity ρ_{ijt} which is assumed to be a random coefficient. This represents remaining heterogeneity that is not captured by the differences in observed characteristics.

There are at least two major challenges in estimating the wage equation (2.2).

The first is that the accumulation of general and job-specific experiences are results of endogenous decisions. As pointed out earlier by Abraham and Farber (1987), job seniority is correlated with unobserved factors including the quality of a worker and employer-employee match. Such confounding factors lead to a bias in cross-sectional estimates of returns to tenure, as measured by $g(a_{it}|\phi_i)$ and $f(s_{ijt}|\phi_i)$ in the equation (2.2). While previous studies by Altonji and Shakotko (1987), Altonji and Williams (2005), and Dustmann and Meghir (2005), among others, use instrumental variables to control endogeneity bias, I consider matching via observables based on a sufficient statistic.

It is also important to notice that the returns to tenure in wage equation (2.2) are heterogeneous across individual workers. Therefore, it is natural to target the distributions of effects, rather than the mean effect, of displacement on earnings. One possible approach is to assume a well-known parametric family of distributions (e.g. normal distribution) for the stochastic component ρ_{ijt} and implement maximum likelihood estimation method. This approach inevitably assumes a strong set of restrictions which is difficult to justify without the prior knowledge about the shape of heterogeneity in returns to skills. For example, distribution of the marginal returns to seniority is possibly more right-skewed among managerial workers than sales persons.[8] Regarding the flexibility in distributional characteristics of heterogeneous returns to experience, I use the nonparametric approach developed in Chapter 1 to estimate the unknown distribution of the random parameter without imposing

[8]Studies found that significant portion of human capital is attached to workers' previous occupation or task. See, for example, Kambourov and Manovskii (2009) and Pavan (2011).

parametric assumptions.

2.3.2 Identification of the Heterogeneous Effects of Displacement

The objective of interest is the counterfactual difference in earnings between displaced and non-displaced workers. The effect is different across workers due to the heterogeneity in accumulated human capital and match-specific components in observed earnings. In this section, I show how the sample of displaced workers helps characterizing heterogeneous returns to experience. The result shows that counterfactual wage loss following displacement is decomposed into unobserved determinants of potential wage level and a function of the worker's experience and observed characteristics.

Suppose that we restrict the sample to displaced workers. Specifically, consider those who have been displaced for exogenous reasons (e.g. plant closure) in period $t-1$ and re-employed in t. Denote $D_{it-1}=1$ for these workers. If such a worker is currently employed (at period t) and observed to have wage $w_{ij't}$ for $j' \neq j$, then the observed log wage is decomposed according to the wage process given as (2.2):

$$\ln w_{ij't}(D_{it-1}=1) = \ln r^e_{it} + g(a_{it}|\phi_i) + f(0|\phi_i) + \mu_{ij't} + \varepsilon_{it} \tag{2.3}$$

On the other hand, consider a worker who has been working continuously between $t-1$ and t in the same position. The case is denoted by $D_{it-1}=0$. Her observed wage is written as follows:

$$\ln w_{ijt}(D_{it-1}=0) = \ln r^e_{it} + g(a_{it}|\phi_i) + f(s_{ijt}|\phi_i) + \rho_{ijt}s_{ijt} + \mu_{ijt} + \varepsilon_{it} \tag{2.4}$$

The difference between (2.3) and (2.4) consists of two parts. The first is that if a worker was displaced in period $t-1$, she would have lost her wage components related to her previous work experience. The second is that the match quality of her post-displacement job will be different from that of her previous job. The counterfactual difference between observed wages, conditional on the worker's experience, is given by

$$\begin{aligned}\Delta_{it} &\equiv \ln w_{ij't}(D_{it-1}=1) - \ln w_{ijt}(D_{it-1}=0) \\ &= f(0|\phi_i) - f(s_{ijt}|\phi_i) - \rho_{ijt}s_{ijt} + \mu_{ij't} - \mu_{ijt} \end{aligned} \tag{2.5}$$

Suppose that the worker type ϕ_i is fully determined by the observed characteristics. Then the heterogeneous effects of displacement (2.5) are given as a function of tenure, observed characteristics, and stochastic components ρ_{ijt}, μ_{ijt}, and $\mu_{ij't}$.

One way to identify the stochastic components separately is to estimate the returns to tenure from the wage equation. Recall the wage process (2.2). Dustmann and Meghir (2005) introduce a set of assumptions under which the deterministic component of the individual-specific returns to tenure is identified if the stochastic returns ρ_{ijt} is mean independent with age, conditional on the observed characteristics of workers. Such a restriction is justified given that the authors only look at the sample of young workers and therefore, potential wage offers are independent for their ages.

Instead, I consider identifying the aggregate loss by displacement as a whole. This is done by imposing two types of identifying restrictions. One is to match displaced workers to non-displaced workers with similar earnings and work history from their previous jobs. The other is to impose a certain type of assumption on the stochastic relationship between heterogeneous returns to experience and worker types. The following is the formal statement of the identifying conditions.

(C1) For all j, $\{\rho_{ijt}, \mu_{ijt}\}$ is independent of D_{it-1} conditional on $\phi_i, a_{it},$, and w_{ijt-1}

(C2) ρ_{ijt} is independent of $\mu_{ij't}$ for all $j' \neq j$ conditional on ϕ_i, a_{it}, and s_{ijt}

(C3) ε_{it} is independent of D_{it-1}

Condition (C1) implies that the unobservable factors in potential earnings (2.2) drawn after getting a new job are independent with the displacement status, conditional on observed characteristics and prior wage level. The assumption is justified if there is no discrimination against displaced workers conditioning on that they have the same observed characteristic to non-displaced workers. More specifically, I assume that workers' previous wage level w_{ijt-1} is a sufficient statistic to capture the quality of match in the previously-held position. The argument is plausible if a worker was separated from the previous job for a reason she cannot control and her potential employers are aware of that.

Condition (C2) implies that the unobserved heterogeneity in returns to seniority is uncorrelated with the match-specific component from a potential new job, conditioning on the observed components that are collectively characterizing the labor market condition, worker type, and tenure. This assumption is in general restrictive given that the re-employment probability differs across individuals. In this case, the match quality of a new job reflects the worker's innate ability and productivity.[9] However, in this paper, I restrict the sample to displaced workers who

[9]If a worker has been unemployed longer than a year, studies find that the duration of unem-

have been unemployed for a relatively short period (less than a year) to minimize the selection bias due to the heterogeneity in unemployment duration.

The idiosyncratic error term ε_{it} is assumed to be random as stated in the condition (C3). It is a necessary assumption to separate stochastic components in potential wage distribution from one that represents worker heterogeneity and that reflects purely idiosyncratic errors. One way to interpret this assumption is that ε_{it} only represents variations in potential wage due to factors that are exogenous from a worker's point of view. The factors may include, but are not limited to, volatility in the local labor market due to aggregate business cycles. In a later section, I compare estimated distributions of earnings losses by workers with different labor market status characterized by the year and local unemployment rate.

From the three conditions stated previously, it can be shown that the underlying distribution of heterogeneous effects of displacement is identified as a function of marginal distributions of potential earnings of displaced and non-displaced workers using the method developed in Chapter 1. The first step is to identify the potential wage distribution of both groups of workers. Under conditions (C1) and (C3), the potential wages of both displaced and non-displaced workers are independent from the displacement status, conditional on observables. Thus the potential wage distribution of a displaced worker is comparable to a non-displaced worker with the same work experience, observed characteristics, and the same wage at $t-1$ as the displaced worker earned from the previous position. This implies that the po-

ployment may have a negative impact on the projected wage profile. See, for example, Abraham et al. (2016).

tential wage distribution is identified via matching on observables. Specifically, let $p(\phi_i, a_{it}, s_{ijt-1}, w_{ijt-1}) \equiv \text{Prob}(D_{it-1} = 1|\phi_i, a_{it}, s_{ijt-1}, w_{ijt-1})$ be the propensity score to displacement. Then the potential wage distribution of displaced workers as defined in (2.3), conditional on a set of observable characteristics $\{\phi_i, a_{it}, s_{ijt-1}, w_{ijt-1}\}$ is identified by the observed wage distribution of displaced workers weighted by $1/p(\phi_i, a_{it}, s_{ijt-1}, w_{ijt-1})$.

Notice that the condition (C1) is not strong enough to separately identify the marginal distributions of unobserved heterogeneous factors ρ_{ijt}, and μ_{ijt}. However, it is not necessary to impose a stronger assumption as the objective of interest is the distribution of the wage difference given as (2.5), not the marginal distributions of every unobserved factor. Instead, it is sufficient to consider an additional restriction which is weaker than imposing pairwise stochastic relationships for every pair among unobserved factors ρ_{ijt}, and μ_{ijt}.

A sufficient condition to achieve identification of the distribution of earnings losses (2.5) is the condition (C2). Note that the wage difference (2.5) is a function of the tenure, worker type, and unobserved heterogeneity in returns ρ_{ijt}. As the deterministic component is fully identified as a function of observables, the only thing needed is to have the unobserved heterogeneity drawn independently from other stochastic components in potential wage offers. Condition (C2) implies that the match-specific components in potential wage offer is uncorrelated with the returns to tenure in the previous job, conditional on the aggregate returns to human capital.

Having identified the potential wage distributions of displaced and non-displaced workers, we may identify the distribution of wage differences as follows. For simplicity, I abstract the index i and t to illustrate the identification strategy. Denote $\ln w_1$ be the potential wage earned by displaced workers while $\ln w_0$ be that of non-displaced workers. The conditional characteristic functions of potential wage distributions are defined as the exponential transformations of wages. Denote $\varphi_j(\omega|\phi, a, s) = E[\exp(\iota\omega \ln w_j)|\phi, a, s]$ for $j = 0, 1$ as the conditional characteristic functions of potential wage distributions where $\iota = \sqrt{-1}$ and $\omega \in \mathbb{R}$. From the condition (C1), we may identify $\varphi_1(\omega|\phi, a, s)$ and $\varphi_0(\omega|\phi, a, s)$ separately by observed wage distribution weighted by the inverse of propensity score to displacement. Denote $\varphi_\Delta(\omega|\phi, a, s)$ as the characteristic function of Δ which is defined as (2.5). Note that, conditional on ϕ, a, and s, the only random component remaining in Δ is ρ. What condition (C2) implies is that Δ is independent to $\ln w_0$, conditional on ϕ, a, and s. Then we may write the characteristic function of potential wage distribution of displaced workers as a product of the characteristic functions of potential wage distribution of non-displaced workers and earnings losses. That is, $\varphi_1(\omega|\phi, a, s) = \varphi_0(\omega|\phi, a, s)\varphi_\Delta(\omega|\phi, a, s)$. Refer to the proof of Proposition 2.1 in Appendix B.1 for derivation. Suppose that the characteristic function of potential earnings distribution of non-displaced workers is non-vanishing in the sense that $\varphi_0(\omega|\phi, a, s) \neq 0$ for all ω, ϕ, a, s. Then the characteristic function of the causal effect of displacement is identified as $\varphi_\Delta(\omega|\phi, a, s) = \varphi_1(\omega|\phi, a, s)/\varphi_0(\omega|\phi, a, s)$. Consequently, the distribution of Δ is uniquely identified as $F_\Delta(\tau|\phi, a, s) = 1/2 + (1/2\pi)\int(\iota\omega)^{-1}\exp(-\iota\omega\tau)\varphi_\Delta(\omega|\phi, a, s)d\omega$.[10] The following proposition formally states

[10]The inversion formula is derived in the earlier work by Gil-Pelaez (1951) and later adopted in the context of statistical deconvolution literature. Refer Hall and Lahiri (2008) and Dattner

the identification result.

Proposition 2.1. *Suppose that conditions (C1), (C2), and (C3) hold. Assume that* $\varphi_0(\omega|\phi, a, s) \neq 0$ *for all* ϕ, a, s, *and* $\omega \in \mathbb{R}$, *then the conditional distribution of* (2.5) *is identified.*

The proof of Proposition 2.1 is given in Appendix B.1. In following section, I propose a flexible estimation method which exploits the intuition from the identification result discussed in this section.

2.4 Nonparametric Estimation of Effects of Displacement

2.4.1 Implementation of Estimation Strategy

I implement a nonparametric method developed in the previous chapter to estimate heterogeneous effects of displacement without having to impose a prior parametric restriction on the distribution of effects. In this section, I briefly illustrate the estimation strategy followed by remarks on how the propositions shown in the previous section correspond to the generic set of identification assumptions.

To estimate the distribution of unobserved components in heterogeneous returns to experience, I first compute the adjusted wage after controlling for the deterministic growth rate. The stochastic components in wage equation (2.2) remain after subtracting market returns $\ln r^e_{it}$ and deterministic growth rates $g(a_{it}|\phi_i)$ and $f(s_{ijt}|\phi_i)$. In the benchmark model, $\ln r^e_{it}$ is approximated only with time-fixed

et al. (2011), for example. A more technically detailed discussion on identification via conditional characteristic functions can be found in chapter 1.

effects. The model is compared with a more complex version by introducing a set of dummy variables specific to region, industry, and occupation, interacted with the time-specific effects.[11] For the deterministic component of growth rate, I consider a quadratic function of the experience. Lastly, note that we restrict our focus to displaced workers and non-displaced workers who continue to work for at least one year. For these cases, the mean level of returns to tenure only depends on their observed characteristics which is parameterized by $f(0|\phi_i)$. Therefore, I use a flexible series estimator for the set of observable characteristics of a worker to estimate the term $f(0|\phi_i)$.

In sum, the observed wage adjusted after the estimated deterministic component is given as

$$\widehat{\ln w_{ijt}} = \ln w_{ijt} - \widehat{\ln r^e_{it}} - g\widehat{(a_{it}|\phi_i)} - \widehat{f(0|\phi_i)}$$

where $\widehat{\ln r^e_{it}}$, $g\widehat{(a_{it}|\phi_i)}$, and $\widehat{f(0|\phi_i)}$ are empirical counterparts of $\ln r^e_{it}$, $g(a_{it}|\phi_i)$, and $f(0|\phi_i)$, respectively. Given that the empirical approximations of the three components are consistent with their population counterparts in the wage equation (2.2), the residual wage approximately equals to

$$\widehat{\ln w_{ijt}} \approx \rho_{ijt} s_{ijt} + \mu_{ijt} + \varepsilon_{it} \tag{2.6}$$

The next step is to find a consistent estimator of the propensity score function. Denote X_{it} for the $k \times 1$ vector of covariates associated with worker i at year t. This includes years of tenure in the previous job, the previous wage, and a set

[11]Region is classified by states. Industries and occupations are classified with 3-digit NAICS codes.

of worker-specific characteristics such as age, sex, race, and educational attainment. Denote $p(x)$ for the propensity score function at a specific value $X_{it} = x$. That is, $p(x) = \text{Prob}(D_{it} = 1|X_{it} = x)$. I implement a series approximation to compute an empirical estimate of the propensity score function which allows flexible approximation without having to impose a certain parametric restriction. Precisely, let $\Psi^{\kappa}(x) = (\psi_1(x), \psi_2(x), \cdots, \psi_{\kappa}(x))'$ be a vector of basis functions. I use polynomials up to second order including products among the components in X_{it}. Denote the logistic cdf by $L(u) = \exp(u)/(1 + \exp(u))$ and the sieve approximation of $p(x)$ is defined by $\widehat{p}_n(x) = L(\Psi^{\kappa}(x)'\widehat{\gamma}_{\kappa})$ where

$$\widehat{\gamma}_{\kappa} = \arg\max_{\gamma} \sum_{i=1}^{n} \Big(D_{it} \ln L(\Psi^{\kappa}(X_{it})'\gamma) + (1 - D_{it}) \ln(1 - L(\Psi^{\kappa}(X_{it})'\gamma)) \Big)$$

Technical results regarding the consistency of series estimators of an unknown function are discussed in, for example, Newey (1997) and Hirano et al. (2003). In particular, Hirano et al. (2003) discuss conditions under which the conditional average treatment effect estimator implemented by a semi-parametrically estimated propensity score function achieves efficiency bound proposed by Hahn (1998). One limitation of the existing theory is that Hirano et al. (2003) restrict the support of covariates to be bounded. This may not be a plausible assumption as X_{it} includes previous wage which has a right-skewed distribution. Therefore, I follow recently developed theory by Belloni et al. (2015) and Hansen (2015) on series estimation of smooth functions over possibly unbounded support. See the Appendix to Chapter 1 for further detail on technical assumptions imposed on the propensity score function.

Having a uniformly consistent approximation of the propensity score function, the next step is to find sample counterparts of the conditional characteristic functions of potential wage distributions which is referred as the conditional characteristic functions as empirical characteristic functions (hereby ECF). Recall that the characteristic functions are defined as the population average of exponential transformation of observed wages. Theorem 1.1 argues that the characteristic functions of potential outcomes are consistently estimated by their sample counterparts. Using this principle, I construct the ECFs of both displaced and non-displaced workers using the adjusted wage as given by (2.6).

Note that the condition (C2) implies that the distribution of earnings losses is independent from the potential wage, conditional on the propensity score. Thus, the objective that must be estimated is the ECF, conditional on the propensity score values. I consider an additional step to project empirical characteristic functions onto the propensity score estimates. Let $P^r(z)$ be a vector of B-spline basis functions of arbitrary order $r \geq 2$. Specifically, let $\{b_1, b_2, \cdots, b_{r-2}\}$ be equally-spaced nodes over the interval $[0,1]$. The B-spline series is defined as $P^r(z) = (1, z, \max\{z - b_1, 0\}, \max\{z - b_2, 0\}, \cdots, \max\{z - b_{r-2}, 0\})'$. I construct a vector of basis using the estimated propensity score value which is denoted by $\widehat{P}^r_{it} \equiv P^r(\widehat{p}_n(X_{it}))$. Let $\gamma_1(D, X) = D/\widehat{p}_n(X)$ and $\gamma_0(D, X) = (1-D)/(1-\widehat{p}_n(X))$ be the inverse propensity score weights associated with the treatment and control groups, respectively. Then the ECF is constructed as follows:

$$\widehat{\varphi}_{j,n}(\omega|z) = P^r(z)'\Big(\sum_{i=1}^{n} \widehat{P}^r_{it}\widehat{P}^{r\prime}_{it}\Big)^{-1}\Big(\sum_{i=1}^{n} \widehat{P}^r_{it}\gamma_j(D_{it}, X_{it}) \exp(\iota\omega\widehat{\ln w_{ijt}})\Big) \tag{2.7}$$

for $j = 0, 1$. In Chapter 1, I present a set of technical assumptions under which the empirical approximation (2.7) is uniformly consistent with its population counterpart.

Precision of the approximated characteristic functions depends on the length of the basis functions which is parameterized with a non-negative integer r. If r is small, empirical characteristic functions do not approximate their population counterparts well enough, even with a large sample size. On the other hand, choosing a too large r may result in over-fitting and therefore, the estimated distribution of heterogeneous effects will be biased. I choose the optimal degree of approximation via the cross-validation method which is widely-used in semi-parametric models and statistical learning models. The idea is to evaluate a measurement of estimation bias with different choice of degrees and find r that minimizes the bias. Evaluation of the model is based on the mean integrated squared error which implies the average squared bias over the point estimates of the distribution of Δ_{it}. Details regarding the implementation of the cross-validation method are illustrated in Appendix B.3.

A nonparametric estimator for the conditional distribution of causal effects of displacement is constructed by applying the statistical deconvolution method. Note that the conditional distribution of heterogeneous effects is independent from the potential wage distribution, conditional on the covariates. As shown in Proposition 2.1, the conditional distribution of the counterfactual wage difference is identified by the ratio of the conditional characteristic functions, assuming that $\varphi_0(\omega|z)$ is bounded away from zero. Then a naive estimator of the distribution of heterogeneous effects is constructed by replacing the conditional characteristic functions with their empirical counterparts. However, simply replacing the population characteristic functions with their estimators results in a biased estimate due to the ill-posed inverse problem (Refer, for example, Carroll and Hall, 1988; Fan, 1991a; Taupin, 2001). Thus I introduce kernel weights to penalize empirical estimates of the ratio of characteristic functions over the spectrum ω. The distribution function is estimated by the following formula:

$$\widehat{F}_{\Delta,n}(\tau|z) = \frac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty}(\iota\omega)^{-1}\exp(-\iota\omega\tau)\varphi_{\xi}(h_n\omega)\frac{\widehat{\varphi}_{1,n}(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)}d\omega \tag{2.8}$$

On the other hand, the corresponding estimator for the quantile effect function of heterogeneous effects is defined as the left inverse of the estimator (2.8). That is,

$$\widehat{Q}_{\Delta,n}(u|z) = \inf_{\tau\in\mathcal{T}}\{\tau : \widehat{F}_{\Delta,n}(\tau|z) \geq u\} \tag{2.9}$$

for $u \in [0, 1]$.

2.4.2 Propensity Score Specification

The key to eliminating potential selection bias is to find a control group that matches the distribution of the characteristics of the treated group. The control group is composed of the workers who have been working continuously during the same period that displaced workers have been unemployed. Using the individual-level identifier, I merge Monthly CPS data across years to track the past wage profile of workers back to one year. In addition, I collect workers who have been continuously working for at least one year from the survey date, as identified from the JT supplement.

Five potential models for propensity score specifications are tested prior to the full estimation step. Table B.7 shows the results from the test of reverse causality between a set of explanatory variables and residuals of treatment after controlling for the propensity score. Each column indicates different model specifications for the propensity score function, from the smallest to the largest models. An important thing to notice is that inclusion of unemployment rates substantively increases the significance of propensity score models compared to other variables. Therefore, in the following exercise, I choose the last specification which controls for individual characteristics, unemployment rates, and past wages.

Treated and controlled groups are matched via the predicted probability of displacement. One way to investigate the quality of the match is to compare the distribution of estimated propensity to displacement within each group. Figure B.2 shows that the distribution of characteristics within both the treated and controlled group are similar to each other. The distribution of estimated propensity score values largely overlap between displaced and non-displaced workers. This may suggest that the sample distributions of two groups are well matched.

2.4.3 Distribution of Earnings Losses

I estimate the distribution of the heterogeneous effects of displacement using the nonparametric approach described in this chapter to investigate how earnings losses are dispersed across workers. Figure B.3 shows the estimated distribution function of heterogeneous effects of displacement on earnings losses. The blue solid line in Figure B.3 indicates the point estimates of the probability distribution function of effects of displacement on earnings estimated nonparametrically by the deconvolution method presented in the previous section. The horizontal line indicates the location of the mean effect which indicates that workers lose about 19% of their earnings following the displacement compared to their non-displaced counterparts.

The nonparametric estimation strategy provides information beyond the mean effect. For a more precise comparison, I present the quantile effects at various levels in Table B.8. The first column presents the mean effect obtained by numerical integration of the nonparametrically estimated density. The quantile effects are computed by the left inverse of the estimated distribution function of the causal effect, following the procedure described previously. The result shows that the mean effect estimated via the nonparametric strategy proposed in this chapter matches well with the benchmark estimates shown in Table B.1 in Appendix B.2.

The following are some of the notable findings. First, there exists a non-negligible heterogeneity in the distribution of earnings losses caused by displacements. The estimated distribution of Δ suggests that, for example, the lower 10% of population would suffer an approximately 70% loss in their wages compared to their non-displaced counterparts. In addition, one may find that the estimated distribution of Δ is slightly right-skewed. This is suggested by the fact that the mean estimate is larger than the median estimate. This implies that if we only measure the effect of displacement by its conditional mean effect, we may ignore the fact that larger fraction of workers may have greater loss than the average estimate of the effect of displacement.

I further investigate the heterogeneity by looking at the differences between groups of workers with the same characteristics. First, I look at the differential impact of displacement within and between groups of workers who are categorized by their tenure. Table B.9 presents the resulting estimates. The first column shows that the average losses following the displacement increase with respect to the seniority. That is, workers with less tenure prior to displacement tend to have larger average losses compared to those with longer experience.

Table B.10 shows heterogeneous effects by workers' education levels. By looking at the mean effects, it is difficult to find differences across groups. Major differences between workers with different education levels, however, are in the within-group heterogeneity. It can be found that the effect of displacement is more dispersed among lower educated workers compared to workers with higher educations. This may suggest that workers with higher education are more likely to find a new position after being displaced that fits well with their set of skills, resulting in smaller dispersion in post-displacement earnings.

On the other hand, I present the differential effects of displacement across the labor market condition represented by the unemployment rate. I consider unemployment rates at disaggregated level in three ways. First is the regional level. I use state level unemployment provided by Local Area Unemployment Statistics. For industry- and occupation-specific unemployment rates, I compute them by merging Annual Social and Economic Supplement sample across the sample period. The first four rows of Table B.11 show estimates of mean and quantile effects by separating the samples into pre- and post-2008. While the mean effect is slightly larger in absolute value after 2008, the differences are smaller in quantile effects.

For a better comparison, I look at the other dimensions by separating the groups by local and industry-specific unemployment rates. The latter part of Table B.11 shows the result, while the threshold for distinguishing high and low unemployment in state and industry is 6% which is the median in the sample used in this application. Unlike the case where the effects of displacement are compared across years, patterns are more clearly stated when the market condition is defined with either local or industry-specific unemployment rates. For example, workers in states with higher unemployment rates suffer 8 percentage points more on average, compared to those who live in states with lower unemployment rates. The same comparison yields a 19 percentage point difference between industries with higher and lower unemployment rates.

In addition, dispersion of the effects of displacement is larger in states and industries with higher unemployment rates than those with lower unemployment rates. The argument can be verified by comparing, for example, 10 to 90 percentile ranges of the cases with either high or low unemployment rates. Among the workers who have been re-employed in states with high unemployment rates, dispersion of the effect is $0.1497-(-0.7596)=0.9093$, which is larger than $0.1518-(-0.7353)=0.8871$ of the states with lower unemployment rates. A similar pattern is found by comparing across industries. The dispersion of the effect of displacement is $0.0657-(-0.8845)=0.9501$ among the workers who have been working in the industries with high unemployment rate, larger than the $0.2153-(-0.6774)=0.8927$ of the industries with lower unemployment rate. Another way to look at this is to compare extreme cases. The difference between the lower 10% quantiles of earnings losses is 34 percentage points when we compare across industry-specific unemployment rates. However, the difference is much smaller for the upper 10%, which is about 9 percentage points. This implies that the lower mean effect of displacement in the tighter labor market is mostly driven by workers who are at the left tail of the earnings distribution.

In Appendix B.2, I present a sensitivity analysis by estimating the heterogeneous effects of displacement on earnings losses via linear regression models. While the intuition remains the same, regression models yield less significant results when it comes to the comparison of differential effects across groups of workers. This may suggest that there exists a highly non-linear relationship between observed characteristics and the distribution of individual-specific effects of displacement. Such a finding suggests that it is useful to consider a flexible nonparametric method while analyzing heterogeneous causal effects.

2.5 Conclusion

In this chapter, I investigated the size of within- and between-group heterogeneity in the earnings losses of displaced workers using a new nonparametric estimation method. Empirical findings suggest that causal consequences of unexpected displacement are significantly different across individuals. For example, less experienced workers tend to suffer less on average compared to more experienced workers, while the within-group dispersion of earnings losses among those young workers is substantially larger. In addition, I also find that a tighter labor market condition represented by higher unemployment rates can lead to a larger dispersion in the heterogeneous effect of displacement. This suggests that measuring the heterogeneity in the causal effect of displacement by only the conditional mean effect may lead to a biased conclusion as it ignores the probability of having significantly larger losses than the average effect.

Chapter 3: Semiparametric Estimation of Non-Linear Impulse Responses with Discrete Policy Shocks (with Guido Kuersteiner)

3.1 Introduction

In this chapter, we develop a semi-parametric estimator for the non-linear impulse-response functions with discrete policy shocks. Empirical investigation of effects of macroeconomic policy across time often relies on the underlying structural assumptions such as Dynamic Stochastic General Equilibrium (DSGE) models (See, for example, Christiano, Eichenbaum, and Evans, 1999; Christiano, Trabandt, and Walentin, 2010). The estimation method proposed in this chapter complements the existing approaches based on structural assumptions by developing an identification and estimation strategy that is robust to underlying models generating impulse-response functions.

The method is based on the discussion by Angrist, Jord, and Kuersteiner (2013) and Angrist, Jordà, and Kuersteiner (2016). In the paper, they propose flexible econometric method to estimate the dynamic impact of monetary policy on macroeconomic outputs including real GDP and treasury yields. The estimation of the causal effect of monetary policy is implemented by inverse policy score weighted estimator. They suggest that if the researcher correctly specified the information set which correctly describes the policy score function, the causal effect of monetary policy is estimated consistently via inverse policy score weighted average of a set of forward-looking outcomes. Using the method, they re-investigate the effect of changes in federal fund rates on aggregate outputs and treasury yields to find asymmetric impulse-response functions which is difficult to be captured with parametric models.

We generalize the method proposed by Angrist et al. (2016) to allow more flexible types of policy score functions. Asymptotic properties for the inverse policy score estimator are studied when the underlying stochastic process is assumed to be weakly dependent. We provide a set of conditions to achieve uniform consistency and asymptotic normality of the semi-parametric estimator while leaving the specification of policy score functions to be flexible. The result is particularly useful in following cases. The first is that when a researcher knows about the set of variables that affect the policy decision while it is hard to specify a certain rule. By using semi-parametric method, the model automatically determines the best fit for the model of policy rule. In addition, the method can be useful to test external validity of impulse-response functions estimated under specific parametric assumptions.

The method proposed in this paper is in line with the theory of inverse propensity score estimators which has been developed extensively for the last few decades. Regarding the asymptotic properties of propensity score matching estimators, Hahn

(1998) is the first to present efficiency bounds of semi-parametric estimators of the causal effect. Further extension to various semi-parametric estimators include, for example, Hirano et al. (2003) and Abadie and Imbens (2006). Some of the noticeable recent studies extend the framework to multi-valued treatments (Cattaneo, 2010) or apply it to dynamic causal effect of monetary policy (Angrist and Kuersteiner, 2011; Angrist et al., 2013, 2016).

The asymptotic properties of the semi-parametric estimator discussed in this paper is heavily rely on the earlier results on series estimators. Some of the examples are Geman and Hwang (1982), Newey (1994), Andrews (1994), and Chen and Shen (1998). More recently, Belloni et al. (2015) and Hansen (2015) extend existing theories to continuous and differentiable functions over non-compact support. Especially on the inferential problem, Newey and West (1987) and Andrews (1991) propose consistent kernel-weighted estimators for asymptotic variance matrix. Later studies including Ackerberg, Chen, and Hahn (2012), Ackerberg, Chen, Hahn, and Liao (2014), and Chen and Liao (2015) extends the method to two-stage GMM estimators under weakly dependent process. We adapt and modify the existing results to construct consistent estimators for the covariance matrix of semiparametric impulse response estimators of finite lags.

The rest of this chapter consists of following sections. In section 3.2, we present stochastic framework and discuss identification conditions for the dynamic policy effects. In section 3.3, we list a set of conditions to achieve consistency and asymptotic normality of the semi-parametric estimator and present key results. In section 3.5, we propose inference theory by constructing feasible test statistics. In section 3.6, we implement Monte Carlo exercise to show finite-sample performance of the estimator in various scenarios. Finally in section 3.7, we conclude.

3.2 Potential Outcomes and Causal Effects

The stochastic environment is determined by the observed vector stochastic process χ_t. Denote by y_t a k_y-vector of outcome variables and by D_t a vector of policy variables that takes on a discrete number of possible values $\mathcal{D} = \{d_0, ..., d_J\}$. Both y_t and D_t are elements of χ_t. A leading example is the case of monetary policy where D_t is the change in the target interest rate the central bank sets. We assume that the information used by policy makers at time t, is known to the public but not necessarily observed by researchers. Formally, the relevant information is assumed to be described by a finite dimensional vector z_t which may depend on past and current observations of y_t and D_t and other current and lagged elements of χ_t. As an example in the case of monetary policy, z_t could be an index constructed from employment and inflation data as would be the case if the Fed followed a Taylor rule.

We assume that there are innovations u_t that drive the process χ_t. In formal statement, we impose the the following probabilistic structure. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the common probability space on which the sequence u_t is defined. Let χ_0 be the initial condition of χ_t at time 0. Let $\mathcal{F}_t = \sigma(\chi_0, u_1..., u_t)$ and assume that there are measurable mappings F_t such that

$$F_t\left(\chi_0, u_1..., u_t, \varphi\right) = \chi_t \tag{3.1}$$

where $\varphi \in \Theta$ is a non-random parameter. The parameter φ determines the mappings F_t and can be finite or infinite dimensional. It follows that χ_t is adapted to $\mathcal{F}_t$. Partition $u_t = (\eta_t, \varepsilon_t)$ and assume that $D_t = D\left(z_t, \varphi, \varepsilon_t\right)$ where z_t is measurable with respect to $\mathcal{F}_{t-1}$. In this notation ε_t is called the policy shock. We assume that ε_t affects χ_t only through D_t. This restriction is formalized as follows. Define the set $\mathcal{E}^d\left(z_t\right) = \{e \in \mathcal{E} | D\left(z_t, \varphi, e\right) = d\}$ and assume that $F_{t+l}\left(\chi_0, u_1..., \left(\eta_t, e\right), u_t, \cdots, u_{t+l}\right)$ is constant for $e \in \mathcal{E}^d\left(z_t\right)$ and each $d \in \mathcal{D}$. We now give a definition of potential outcomes.

Definition 3.1. *Potential outcomes* $y^{\varphi}_{t,l}\left(d\right)$ *are defined by*

$$y^{\varphi}_{t,l}\left(d\right) = F^y_{t+l}(\chi_0, u_1, \cdots, u_{t-1}, (\eta_t, e), u_{t+1}, \cdots, u_{t+l}, \varphi)$$

for $e \in \mathcal{E}^d\left(z_t\right)$ *where* $F^y_{t+l}(\cdot)$ *is the component corresponding to* y *of the vector valued function* $F_t(\cdot)$ *defined in* (3.1).

The causal effect of a policy is defined as the difference $y^{\varphi}_{t,l}(d_j) - y^{\varphi}_{t,l}(d_0)$. Counterfactual outcomes indexed against regime as well as a policy change allow us to distinguish the effects of systematic and unexpected changes, though only the latter are identified in our framework. Note that the notation $y^{\varphi}_{t,l}\left(d\right)$ accounts for both the timing of the intervention relative to the initial conditions as well as the delay l at which the effect of the intervention is measured.

The variation that identifies causal relationships is characterized by the selection on observables identifying assumption:

Condition 3.1. *Selection on observables:* $y_{t,l}^{\varphi}(d) \perp D_t | z_t, \varphi$ *for all* $l > 0$, $d \in \mathcal{D}$, *and* $\varphi \in \Theta$.

Given the structure imposed, Condition 3.1 is identical to the assumption that ε_t is independent of $u_1, \cdots, u_{t-1}, \eta_t, u_{t+1}, \cdots, u_{t+l}$, conditional on z_t. If we strengthen Condition 3.1 to requiring $y_{t,l}^{\varphi}(d_j) \perp D_t | (z_t, \chi_{t-1},)$ the then it is enough to require that ε_t is indpendent of $\eta_t, u_{t+1}, \cdots, u_{t+l}$, conditional on (z_t, χ_{t-1}). Condition 3.1 is somewhat weaker than requiring that ε_t, η_t are independent sequences of random variables. Consider the effect of setting $D_t = d_j$ on a vector of outcome variables y_t for a given regime, φ. Let $Y_t = (y_t', \cdots, y_{t+L}')'$ and define the vector of potential outcomes $Y_t^{\varphi}(d) = \left(y_{t,0}^{\varphi}(d), \cdots, y_{t,L}^{\varphi}(d)\right)$. Define the dummy variables $D_{t,j} = 1\{D_t = d_j\}$. Given a fixed regime the potential outcome of a policy variable is related to the observed outcome through the following latent variables model:

$$Y_t = \sum_{d \in \mathcal{D}} Y_t^{\varphi}(d) 1\{D_t = d\} \tag{3.2}$$

Using Equation (3.2) and Condition 3.1, average policy effects, conditional on covariates z_t, are identified as follows

$$E\left[Y_t^{\varphi}(d_j) - Y_t^{\varphi}(d_0) | z_t\right] = E\left[Y_t | D_t = d_j, z_t\right] - E\left[Y_t | D_t = d_0, z_t\right]. \tag{3.3}$$

Estimation of the conditional expectations in (3.3) can be considerably simplified if a parametric model for the policy variable D_t is available. In Angrist and Kuersteiner (2004, 2011), such a policy model was termed the policy propensity score. Assuming that $\mathbb{P}(D_t = d_j|z_t) = p^j(z_t, \varphi)$ we construct estimates from the fact that

$$E[Y_t D_{t,j}|z_t] = E[Y_t^{\varphi}(d_j)|z_t]\, p^j(z_t, \varphi).$$

Therefore,

$$E\Big[Y_t^{\varphi}(d_j) - Y_t^{\varphi}(d_0)|z_t\Big] = \frac{E[Y_t D_{t,j}|z_t]}{p^j(z_t, \varphi)} - \frac{E[Y_t D_{t,0}|z_t]}{p^0(z_t, \varphi)}, \tag{3.4}$$

a formulation that first appeared in Horvitz and Thompson (1952). The LHS of (3.4) is the impulse response function, generated from the unknown and potentially nonlinear process for outcomes. In the cross-section literature, (3.4) would be called an average treatment effect. In Angrist et al. (2013) discuss the relationship of 3.4 to the literature on non-linear impulse response functions, in particular Gallant, Rossi, and Tauchen (1993).

One advantage of our approach to non-linear impulse response functions is that estimation and testing is extremely simple and reduces to computing a multivariate weighted average. Joint confidence sets for the impulse coefficients are readily available. We define

$$\theta_{l,j}(z_t) = E[y_{t+l}|D_t = d_j, z_t] - E[y_{t+l}|D_t = d_0, z_t]. \tag{3.5}$$

The causal effect, or average impulse of the policy innovation at time t on the outcome variable at time $t+l$ then is defined as

$$E\left[y_{t,l}^{\varphi}(d_j) - y_{t,l}^{\varphi}(d_0)\right] = \theta_{l,j}$$

where $\theta_{l,j}$ has the interpretation of a generalized impulse response of an unexpected policy change from d_0 to d_j at time t on the outcome variable at time $t+l$. The policy d_0 is a common reference policy. Using the vector notation introduced in Section 3.2 the impulse response function of the unexpected policy change can be compactly represented in vector form as

$$E\left[Y_t^{\varphi}(d_j) - Y_t^{\varphi}(d_0)\right] \equiv \theta_j$$

where now $\theta_j = (\theta_{0,j}', ..., \theta_{L,j}')'$ is the impulse response function at horizons 0 to L. We can further summarize the effects of all possible policy changes by defining the vector of effects θ as $\theta = (\theta_1', ..., \theta_J')'$. The dimension of θ is $k = k_y(L+1)J$ with k_y the number of outcome variables, L the horizon of the impulse response function and J the number of distinct policy choices.

In addition, we may be interested in the effects of the policy variable conditional on the policy being enacted. This effect is defined as

$$E\left[Y_t^{\varphi}(d_j) - Y_t^{\varphi}(d_0) | D_t = d_j\right] \equiv \theta_{j,TOT}.$$

This parameter is known as the treatment effect on the treated in the cross-sectional literature.[1] It follows from well known arguments that

$$E\left[Y_t\left(\frac{D_{t,j}}{p^j(z_t,\varphi)} - \frac{D_{t,0}}{p^0(z_t,\varphi)}\right)\frac{p^j(z_t,\varphi)}{\mathbb{P}(D_t = d_j)}\right] = \theta_{j,TOT}$$

The key ingredient in our estimation strategy is the weighting function,

$$\tau_{t,j}(\varphi) = \left(\frac{D_{t,j}}{p^j(z_t,\varphi)} - \frac{D_{t,0}}{p^0(z_t,\varphi)}\right) g^j(z_t,\varphi)$$

where $g(z_t,\varphi) = 1$ for the unconditional impulse response function and $g(z_t,\varphi) = p^j(z_t)/\mathbb{P}(D_t = d_j, \varphi)$ for the impulse response function conditional on the policy

[1] See, for example, Hirano et al. (2003).

maker taking action d_j. For the estimator of θ_j we define

$$h_{t,j}(\varphi) = Y_t \tau_{t,j}(\varphi) \tag{3.6}$$

and $h_t(\varphi) = \left(h'_{t,1}, ..., h'_{t,J}\right)'$. The functions h_t are evaluated at φ such that $\hat{h}_t = h_t(\varphi)$. Then θ is estimated directly as the sample average of $\hat{h}_t$,

$$\hat{\theta} = T^{-1} \sum\nolimits_{t=1}^{T} \hat{h}_t. \tag{3.7}$$

Estimation of θ can be understood as a minimum distance procedure where $\hat{\theta}$ solves

$$\arg\min_{\theta} \left(T^{-1} \sum\nolimits_{t=1}^{T} \hat{h}_t - \theta\right)' \Omega^{-1} \left(T^{-1} \sum\nolimits_{t=1}^{T} \hat{h}_t - \theta\right). \tag{3.8}$$

Because there are as many parameters as moment conditions the choice of Ω does not affect the solution to (3.8) which is always (3.7). However, considering (3.8) is useful in parctice because we may want to constrain the parameter estimates in θ. There are at least two scenarios that are of parctial importance. Consider the case where $d_j = -d_{j'}$, for example d_j could stand for a 25 basis point increase in the Fed target rate while $d_{j'}$ is a 25 basis point decrease in the target. In a linear model, such as a conventional VAR, the impulse response for d_j has the oposite sign of the impulse response to $d_{j'}$. To impose this constraint in our setting we impose the restriction $\theta_j = -\theta_{j'}$. More generally, we can consider general linear constraints of the form $\theta = R\alpha$ with R a fixed and known matrix of dimension $k \times q$ and α is a $q \times 1$ vector of free parameters.

A second possibility consists in modeling θ_j as a function of a lower dimensional parameter α. One might for example want to approximate the impulse response function with a low order polynomial to achieve a parsimonious specification. We thus consider more general cases where $\theta = \theta(\alpha)$ and $\theta(\cdot)$ is a known function. Clearly a special case is $\theta(\alpha) = R\alpha$. The case $R = I$ covers (3.8). Thus, estimation of α and θ is based on

$$\arg\min_{\alpha} \left(T^{-1}\sum\nolimits_{t=1}^{T} h_t - \theta(\alpha)\right)' \hat{\Omega}^{-1} \left(T^{-1}\sum\nolimits_{t=1}^{T} h_t - \theta(\alpha)\right). \tag{3.9}$$

In this case the model is generally overidentified or in other words the dimension of α is less than k, the dimension of h_t. The choice of Ω now does matter for efficiency of the estimator. The optimal Ω in this case is the spectral density matrix of h_t at zero frequency which can be estimated by standard methods such as Newey and West (1987). Once we obtain an estimate $\hat{\alpha}$ we estimate $\hat{\theta} = \theta(\hat{\alpha})$.

The asymptotic variance covariance matrix Ω_θ of $\hat{\theta}$ is given by

$$\Omega_\theta = \sum_{j=-\infty}^{\infty} E\left[v_t(\varphi_0)\, v_{t-j}(\varphi_0)'\right] \tag{3.10}$$

where $v_t(\varphi_0) = h_t(\varphi_0) - \theta_0 + \dot{h}(\varphi_0)\Omega_\varphi^{-1} l(D_t, z_t, \varphi_0)$ and $\dot{h}(\varphi_0) = E\left[\partial h_t(\varphi_0)/\partial\varphi'\right]$. The formula for Ω_θ takes into account that the 'observations' $\hat{h}_t$ used to compute the sample averages are based on estimated, rather than observed data. It now follows under regularity conditions detailed below that $T^{1/2}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N(0, \Omega_\theta)$. Confidence intervals for θ can be constructed from this distributional approximation in standard ways, using an estimator for Ω_θ. In order to estimate Ω_θ we start by forming the sample averages

$$\widehat{\dot{h}}(\varphi_0) = T^{-1}\sum_{t=1}^{T} \partial h_t(\hat{\varphi})/\partial\varphi', \hat{\Omega}_\varphi = T^{-1}\sum_{t=1}^{T} l(D_t, z_t, \hat{\varphi}) l(D_t, z_t, \hat{\varphi})'$$

and letting $v_t(\hat{\varphi}) = h_t(\hat{\varphi}) - \hat{\theta} + \widehat{\dot{h}}(\varphi)\hat{\Omega}_\varphi^{-1} l(D_t, z_t, \hat{\varphi})$. Methods to estimate Ω_θ have been developed in Newey and West (1987, 1994), Andrews (1991) and Andrews and

Monahan (1992), amongst others.

3.3 Assumptions

This section provides an analysis of the the asymptotic properties for the estimators proposed in Section 3.2. We make use of results in Andrews (1994), Newey (1994), and Chen and Shen (1998) for asymptotic approximations of semiparametric estimators. We start with a list of regularity conditions we impose on the process χ_t which is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that $\{\chi_t\}_{t=1}^{\infty}$ is strictly stationary with values in the measurable space $(\mathbb{R}^r, \mathcal{B}^r)$ where $\mathcal{B}^r$ is the Borel σ-field on $\mathbb{R}^r$ and r is fixed with $2 \leq r < \infty$. Let P be the marginal distribution function of χ_t. Let $\mathcal{A}_1^l = \sigma(\chi_1, ..., \chi_l)$ be the sigma field generated by $\chi_1, ..., \chi_l$. The sequence χ_t is β-mixing or absolutely regular if

$$\beta_m = \sup_{l \geq 1} E\left[\sup_{A \in \mathcal{A}_{l+m}^{\infty}} \left|\mathbb{P}\left(A|\mathcal{A}_1^l\right) - \mathbb{P}(A)\right|\right] \to 0 \text{ as } m \to \infty. \tag{3.11}$$

Condition 3.2. *Let χ_t be a stationary, absolutely regular process such that for some $2 < p < \infty$ the β-mixing coefficient of χ_t satisfies $\beta_m \leq m^{-p/(p-2)}$. There exists a $\delta > 0$ such that $E\left[\|\chi_t\|^{2+\delta}\right] < \infty$.*

We consider the general case where $E[\mathbf{1}(D_t = d_j)|z_t] = p^j(z_t)$ and $p^j(z) \in \mathcal{P}$ where $\mathcal{P}$ is a class of functions that satisfies the constraint $\sum_{j=1}^{J} p^j(z) = 1$ for all z. As is the case when the propensity score is parameteric, we need to take estimation uncertaint of estimating the propensity score into account when deriving the limiting distribution of our causal parameter θ. Newey (1994) provides formulas for computing the effect of non-parametric estimation in plug-in procedures. Expressions for the correction terms were obtained by Hahn (1998) and Hirano et al. (2003) for the binary treatment case and by Cattaneo (2010) for the average treatment effect in the case of multinomial treatments. Let

$$\tau_{t,j}\left(p^0, p^j\right) = \left(\frac{D_{t,j}}{p^j\left(z_t\right)} - \frac{D_{t,0}}{p^0\left(z_t\right)}\right) g^j\left(z_t\right)$$

with $g^j\left(z_t\right) = p^j\left(z_t\right)$. Then $h_{t,j}\left(p^0, p^j\right) = Y_t \tau_{t,j}\left(p^0, p^j\right)$, and

$$h_t\left(p\right) = \left(h'_{t,1}\left(p^0, p^1\right), \cdots, h'_{t,J}\left(p^0, p^J\right)\right)'$$

with $p = \left(p^0, ..., p^J\right)$. Now let $\xi_t = \left(Y_t, D_t, z_t\right)$ and define the function $m\left(\xi_t, \alpha, p\right) = h_t\left(p\right) - \theta\left(\alpha\right)$. The GMM estimator of α is now defined as

$$\hat{\alpha} = \arg\min_{\alpha} \left(T^{-1} \sum\nolimits_{t=1}^{T} h_t\left(\hat{p}\right) - \theta\left(\alpha\right)\right)' \hat{V}^{-1} \left(T^{-1} \sum\nolimits_{t=1}^{T} h_t\left(\hat{p}\right) - \theta\left(\alpha\right)\right) \tag{3.12}$$

where $\hat{p}$ is some non-parametric estimator of p. The influence function of the estimator α then is given by

$$v_t\left(\xi_t, \alpha, p\right) = m\left(\xi_t, \alpha, p\right) + \gamma\left(\xi_t\right)$$

where $\gamma\left(\xi_t, \alpha, p\right) = \left(\gamma^1\left(\xi_t\right), ..., \gamma^J\left(\xi_t\right)\right)$ and explicit formulas for $\gamma^j\left(\xi_t\right)$ are given in the appendix.

To establish an asymptotic distribution for our dynamic treatment effect estimators we impose the following additional high level regularity condition. Low level conditions under which these assumptions hold are discussed in Section 3.4.

Condition 3.3. *The functions $p^j(z_t) = E[\mathbf{1}(D_t = d_j)|z_t]$ for $j = 0, ..., J$ satisfy the following conditions:*

(i) $\inf_z p^j(z) > \underline{p} > 0$ *for all* $j = 0, .., J$.

(ii) Assume that $\widehat{p} = (\widehat{p}^0(z_t), \cdots, \widehat{p}(z_t))$ *are nonparametric estimators of* p *such that* $\|\widehat{p}(z) - p(z)\|_{4+\delta,P} = o(T^{-1/4})$.

(iii) Let $b^{0,j}(z_t) = E[Y_t(d_0)|z_t]g_0^j/p_0^0$, $b^j(z_t) = -E[Y_t(d_j)|z_t]g_0^j/p_0^j + (E[Y_t(d_j)|z_t] - E[Y_t(d_0)|z_t])$, $b(z) = \operatorname{diag}(b^1(z), \cdots, b^J(z))$ *and* $b^0(z) = (b^{0,1}(z), \cdots, b^{0,J}(z))'$. *Let* $B(z) = b(z) - b^0(z)\mathbf{1}'$ *and* $\dot{p}_\kappa(z_t, \varphi_\kappa) = \partial p_\kappa(z_t, \varphi_\kappa)/\partial\varphi_\kappa'$. *Then, there exists a sequence* Π_κ *of* $J \times J\kappa$ *matrices of constants such that*

$$E\left[\|\Pi_\kappa (I_J \otimes \Psi^\kappa(z_1)) - B(z_1)\|^2\right] = o\left(\kappa^{-\alpha}\right).$$

(iv) Let $\frac{1}{T}\frac{\partial^2(L_{T,\kappa}(\varphi_\kappa))}{\partial\varphi_\kappa\partial\varphi_\kappa'} = H_T(\varphi_\kappa)$. *Then,* $\|H_T(\varphi_\kappa^*) - H(\varphi_\kappa^*)\|_2 = O_p\left(\zeta(\kappa)/\sqrt{T}\right)$.

(v) $\|H(\varphi_\kappa^*) - \Omega_{\varphi,\kappa}\|_2^2 = O(\kappa^{-\alpha}\zeta(\kappa))$.

(vi) $\|\hat{\varphi}_\kappa - \varphi_\kappa^*\| = O\left(\zeta(\kappa)/\sqrt{T} + \kappa^{-\alpha}\zeta(\kappa)\right)$.

Note that the functions $p^j(z)$ satisfy the constraint $\sum_{j=0}^J p^j(z) = 1$ by definition and because of the properties of D_t. The constraint $0 \leq p^j(z) \leq 1$ also holds by defintion while Condition 3.3 imposes additional constraints on the joint distribution of D_t and z_t.

The following result shows asymptotic normality of the semiparametric estimator when the propensity score is estimated nonparametrically. The high level conditions given here are the same as part of Newey (1994, Assumption 5.1). More specifically, a high level assumption on the estimation of p is imposed with the remaining conditions in Newey (1994) following in a similar way as shown in Hahn (1998) when Conditions 3.2 and 3.3 hold.

THEOREM 3.1. *Assume that Conditions 3.2 and 3.3 hold. In addition, assume that $\theta_0 = \theta(\alpha_0)$ holds and that $\hat{V} \to_p V$ where V is given as*

$$V = \sum_{l=-\infty}^{\infty} E\left[\left(m(\xi_t, \alpha, p) + \gamma(\xi_t)\right)\left(m(\xi_{t-l}, \alpha, p) + \gamma(\xi_{t-l})\right)'\right]$$

and where $\dot{\theta}(\alpha_0) = \partial\theta(\alpha_0)/\partial\alpha'$ and $\dot{\theta}(\alpha)$ is assumed to be continuous and Ω_θ to be positive definite. Let $\kappa = T^\beta$. Then, it follows that for $\hat{\alpha}$ as in (3.12) satisfies

$$\sqrt{T}(\hat{\alpha} - \alpha) \to^d N\left(0, \dot{\theta}(\alpha)' V^{-1} \dot{\theta}(\alpha)\right).$$

3.4 Semiparametric Estimation of the Propensity Score

Denote $\Psi^\kappa(z) = (\psi_{1\kappa}(z), ..., \psi_{\kappa\kappa}(z))'$ and $\varphi_{j\kappa}$ a $\kappa \times 1$ dimensional vector of parameters. Define $\Psi_{j,\kappa}(z_t, \varphi_\kappa) = \Psi^\kappa(z)' \varphi_{j\kappa}$ and $\varphi_\kappa = (\varphi'_{1\kappa}, ..., \varphi'_{J\kappa})'$. An example for $\Psi^\kappa(z)$ is a polynomial which can be defined by letting $\lambda = (\lambda_1, ..., \lambda_K)$ be a vector of non-negative integers and define $z_t^\lambda = \prod_{l=1}^K z_{t,l}^{\lambda_l}$ where $z_{t,l}$ are the elements of the vector z_t and $|\lambda| = \sum_{l=1}^K \lambda_l$ is the order of the polynomial. Let $\lambda(q)$ be a sequence of vectors indexed by κ. The approximating polynomial then is given by $\Psi_{j,\kappa}(z_t, \varphi_\kappa) = \sum_{q=1}^\kappa z_t^{\lambda(q)} \varphi_{q,j\kappa}$. The probability $p^j(z_t, \varphi)$ is obtained by applying the logistic transfromation of Hirano et al. (2003) and Cattaneo (2010) to $\Psi_{j,\kappa}(z_t, \varphi_\kappa)$ and is given by

$$p_\kappa^j(z_t, \varphi_\kappa) = \frac{\exp(\Psi_{j,\kappa}(z_t, \varphi_\kappa))}{1 + \sum_{j=1}^J \exp(\Psi_{j,\kappa}(z_t, \varphi_\kappa))}, \; p_\kappa^0(z_t, \varphi_\kappa) = \frac{1}{1 + \sum_{j=1}^J \exp(\Psi_{j,\kappa}(z_t, \varphi_\kappa))}$$

Let

$$p_\kappa(z_t, \varphi_\kappa) = \left(p_\kappa^1(z_t, \varphi_\kappa), \cdots, p_\kappa^J(z_t, \varphi_\kappa)\right)'$$

$$\Psi_\kappa(z_t,\varphi_\kappa) = (\Psi_{1,\kappa}(z_t,\varphi_\kappa), ..., \Psi_{J,\kappa}(z_t,\varphi_\kappa))$$

such that

$$\Psi_\kappa(z_t,\varphi_\kappa) = \left(I_J \otimes \Psi^\kappa(z_t)'\right)\varphi_\kappa. \tag{3.13}$$

Denoting by $\Gamma_j(\Psi_\kappa(z_t,\varphi_\kappa))$ the mapping $\Psi_\kappa(z_t,\varphi_\kappa) \to p_\kappa^j(z_t,\varphi_\kappa)$, the inverse of Γ_j is given by

$$\Psi_{j,\kappa}(z_t,\varphi_\kappa) = \log\left(p_\kappa^j(z_t,\varphi_\kappa)/p_\kappa^0(z_t,\varphi_\kappa)\right) \tag{3.14}$$

where $p_\kappa^0(z_t,\varphi_\kappa) = 1 - \sum_{j=1}^J p_\kappa^j(z_t,\varphi_\kappa)$. Thus, for

$$\Gamma(\Psi_\kappa(z_t,\varphi_\kappa)) = (\Gamma_1(\Psi_\kappa(z_t,\varphi_\kappa)), ..., \Gamma_J(\Psi_\kappa(z_t,\varphi_\kappa)))'$$

the inverse is given by

$$\Gamma^{-1}(p_\kappa(z_t,\varphi_\kappa)) = \left(\log\left(p_\kappa^1(z_t,\varphi_\kappa)/p_\kappa^0(z_t,\varphi_\kappa)\right), ..., \log\left(p_\kappa^J(z_t,\varphi_\kappa)/p_\kappa^0(z_t,\varphi_\kappa)\right)\right)'. \tag{3.15}$$

The approximate multinomial log likelihood is then $L_{T,\kappa}(\varphi_\kappa) = \sum_{t=1}^T l_\kappa(D_t,z_t,\varphi_\kappa)$ where $l_\kappa(D_t,z_t,\varphi_\kappa) = \sum_{j=0}^M D_{t,j}\log(p_\kappa^j(z_t,\varphi_\kappa))$ which can be maximized to obtain parameter estimates for φ_κ. Let $L(\varphi_\kappa) = E[l_\kappa(D_t,z_t,\varphi_\kappa)]$ and define $\varphi_\kappa^* = \arg\max L(\varphi_\kappa)$. The estimator $\hat{\varphi}_\kappa$ satisfies

$$\hat{\varphi}_\kappa = \arg\max_{\varphi_\kappa} L_{T,\kappa}(\varphi_\kappa).$$

Estimated probabilities are obtained from $\hat{\varphi}_\kappa$ by using a truncating sequence $\tau_\kappa > 0$ where $\tau_\kappa \to 0$ as $\kappa \to \infty$. Then define

$$\widehat{p}_t^j = \max\left(p_\kappa^j(z_t,\hat{\varphi}_\kappa), \tau_\kappa\right).$$

The following result extends Cattaneo (2010) to the case of stationary, weakly dependent processes with unbounded support. The conditions imposed here are comparable Newey (1997), Hirano et al. (2003), and Cattaneo (2010), yet adjusted to the specific proof strategy employed. In particular, L^p norms rather than uniform bounds are imposed on the qaulity of the series approximation.

Condition 3.4. *i) For $s > 0$, s not integer, the Hölder space is the space $C^s\left(\mathbb{R}^d\right)$ of all $\lfloor s \rfloor$-times differentiable functions f with finite norm*

$$\|f\|_{s,\infty} = \sum_{0 \leq |\alpha| \leq \lfloor s \rfloor} \|D^\alpha f\|_\infty + \sum_{|\alpha| = \lfloor s \rfloor} \sup_{x \neq y} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x-y|^{s-\lfloor s \rfloor}}.$$

Assume that that for each j, $p^j(z) \in C^s(\mathbb{R}^d)$, $b^j(z) \in C^s(\mathbb{R}^d)$ and $b^{0,j}(z) \in C^s(\mathbb{R}^d)$.

ii) $E[\|\Psi^\kappa(z_t)\|^{4+\delta}] = O(\zeta(\kappa))$,

iii) for some δ it holds that the smallest eigenvalue $\lambda_{\min} E[\Psi^\kappa(z_t)\Psi^\kappa(z_t)'] > \delta > 0$ uniformly in κ.

iv) for all $f(z) \in L_2(P)$ there exists a constant C such that, for all κ,

$$\sum_{l=1}^{\kappa} \left(\int f(z) \psi_{l\kappa}(z) \, dP(z) \right)^2 \leq C < \infty.$$

Theorem 3.2. *Let $\tau_\kappa = \kappa^{-\gamma}$. If Conditions 3.2, 3.3(i) and 3.4 are satisfied then*

$$\|\hat{\varphi}_\kappa - \varphi^*_\kappa\| = O\left(\zeta(\kappa)/\sqrt{T} + \kappa^{-\alpha}\zeta(\kappa)\right) \tag{3.16}$$

and for $p \leq 4 + \delta$,

$$\int \left\| p(z) - p^j_\kappa(z, \hat{\varphi}_\kappa) \right\|^p dP_0(z) = O_p\left(\frac{\zeta(\kappa)^{p+1}}{T^{p/2}} + \kappa^{-\alpha p}\zeta(\kappa)^{p+1} \right) \tag{3.17}$$

In addition,

$$\|H_T(\varphi^*_\kappa) - H(\varphi^*_\kappa)\|_2 = O_p\left(\zeta(\kappa)/\sqrt{T}\right), \qquad \|H(\varphi^*_\kappa) - \Omega_{\varphi,\kappa}\|_2^2 = O\left(\kappa^{-\alpha}\zeta(\kappa)\right).$$

Theorem 3.2 establishes the high level assumptions made in Condition (3.3) under more primitive assumptions.

3.5 Feasible Tests

3.5.1 Kernel-weighted Long-run Variance Estimator

In this section, we discuss a feasible testing procedure associated with a general type of joint hypothesis imposed on the parameter of interest, denoted by α. To begin with, we first construct a consistent estimator of the long-run variance V shown in Theorem 1. The idea is to apply the plug-in estimator of Ackerberg et al. (2012) while adjusting for weak dependence of the underlying process.

We begin with illustrating the steps to construct variance estimator. Denote $\widehat{v}_t = \widehat{v}_t(\xi_t, \widehat{\alpha}, \widehat{\varphi}_\kappa)$ as the sample analog of the score function. As the score function itself, its empirical counterpart $\widehat{v}_t$ also consists of two parts. First, the mean deviation $m(\xi_t, \alpha, p) = h_t(\xi_t, \alpha, p) - \theta(\alpha)$ is approximated by $\widehat{h}_t(\xi_t, \widehat{\alpha}, \widehat{\varphi}_\kappa) - \theta(\widehat{\alpha})$ where $\widehat{h}_t(\xi_t, \alpha, \varphi_\kappa) = \big(\widehat{h}_{t,1}(\xi_t, \alpha, \varphi_\kappa), \cdots, \widehat{h}_{t,J}(\xi_t, \alpha, \varphi_\kappa)\big)'$ and

$$\widehat{h}_{t,j}(\xi_t, \alpha, \varphi_\kappa) = Y_t\Big(\frac{D_{t,j}}{p^j(z_t, \varphi_\kappa)} - \frac{D_{t,0}}{p^0(z_t, \varphi_\kappa)}\Big)g^j(z_t, \varphi_\kappa),$$

for $j = 1, \cdots, J$.

Second, for the influential function, we approximate the linear functional $D(\xi_t, p)$ which is explicitly derived in the appendix with sample analog counterparts. Following the argument in Chen and Liao (2015), we use a plug-in estimator to construct the empirical counterpart of $D(\xi_t, p)$. First, we begin with the directional derivative, denoted by $D(\xi_t, p_0)[p] \equiv \frac{\partial l(D_t, z_t, p_0 + \tau p)}{\partial \tau}\Big|_{\tau=0}$, which is a linear functional in p. The empirical counterpart of $D(\xi_t, p_0)[p]$ can be written as $\widehat{D}(\xi_t, \widehat{p})[p] = \frac{\partial l(D_t, z_t, \widehat{p} + \tau p)}{\partial \tau}\Big|_{\tau=0}$ where $\widehat{p}(\cdot)$ is the sieve approximant of the true propensity score function $p_0(\cdot)$. To complete characterization of the influence function, we need to replace p within $\widehat{D}(\xi_t, \widehat{p})[p]$ with the sieve estimator $\widehat{p}(\cdot)$. For an arbitrary orthornormal sieve space $\mathcal{P}_\kappa = \{P(\cdot) = \Psi^\kappa(\cdot)'\varphi_\kappa, E[\Psi^\kappa(z)\Psi^\kappa(z)'] = I_\kappa\}$, we have a Riesz representor $\widehat{p}(\cdot) = \Psi^\kappa(\cdot)'\widehat{\varphi}_\kappa = \Psi^\kappa(\cdot)'\widehat{\Omega}_\varphi^{-1}\widehat{\dot{h}}(\alpha, \widehat{\varphi}_\kappa)$ where $\widehat{\Omega}_\varphi$ is an estimator of Hessian $E\big[\frac{\partial l(D_t, z_t, p_0)}{\partial p}\big(\frac{\partial l(D_t, z_t, p_0)}{\partial p}\big)'\big]$ and $\widehat{\dot{h}}(\alpha, \varphi_\kappa)$ is an estimator of gradient $E\big[\frac{\partial h_t(\xi_t, \alpha_0, p_0)}{\partial p}\big]$.

In sum, we have

$$\widehat{\upsilon}_t(\xi_t, \widehat{\alpha}, \widehat{\varphi}_\kappa) = \widehat{h}_t(\xi_t, \widehat{\alpha}, \widehat{\varphi}_\kappa) - \theta(\widehat{\alpha}) + \widehat{\dot{h}}(\widehat{\alpha}, \widehat{\varphi}_\kappa)\widehat{\Omega}_\varphi^{-1}\frac{\partial l_\kappa(D_t, z_t, \widehat{\varphi}_\kappa)}{\partial \varphi_\kappa}$$

where

$$\begin{aligned}
\widehat{\dot{h}}(\alpha, \varphi_\kappa) &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \widehat{h}_t(\xi_t, \alpha, \varphi_\kappa)}{\partial \varphi_\kappa'} = \frac{1}{T}\sum_{t=1}^{T}\Bigg(\frac{\partial \widehat{h}_{t,1}(\xi_t, \alpha, \varphi_\kappa)}{\partial \varphi_\kappa} \quad \cdots \quad \frac{\partial \widehat{h}_{t,J}(\xi_t, \alpha, \varphi_\kappa)}{\partial \varphi_\kappa}\Bigg)' \\
\frac{\partial \widehat{h}_{t,j}(\xi_t, \alpha, \varphi_\kappa)}{\partial \varphi_\kappa} &= Y_t\Big(\frac{D_{t,0}}{p^0(z_t, \varphi_\kappa)^2}\frac{\partial p^0(z_t, \varphi_\kappa)}{\partial \varphi_\kappa} - \frac{D_{t,j}}{p^j(z_t, \varphi_\kappa)^2}\frac{\partial p^j(z_t, \varphi_\kappa)}{\partial \varphi_\kappa}\Big)g^j(z_t, \varphi_\kappa) \\
&\quad + Y_t\Big(\frac{D_{t,j}}{p^j(z_t, \varphi_\kappa)} - \frac{D_{t,0}}{p^0(z_t, \varphi_\kappa)}\Big)\frac{\partial g^j(z_t, \varphi_\kappa)}{\partial \varphi_\kappa} \\
\widehat{\Omega}_\varphi &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_\kappa(D_t, z_t, \widehat{\varphi}_\kappa)}{\partial \varphi_\kappa}\frac{\partial l_\kappa(D_t, z_t, \widehat{\varphi}_\kappa)}{\partial \varphi_\kappa'} \\
\frac{\partial l_\kappa(D_t, z_t, \varphi_\kappa)}{\partial \varphi_\kappa} &= \big(D_t - p(z_t, \varphi_\kappa)\big) \otimes \begin{pmatrix} 1 \\ P_\kappa(z_t) \end{pmatrix}.
\end{aligned}$$

A consistent estimator of the long-run variance is constructed by using a symmetric kernel as in Newey and West (1987, 1994) and Andrews (1991). Let $K(\cdot)$ be a kernel function that satisfies a certain set of assumptions. Then the kernel-based estimator of V is given by

$$\widehat{V} = \widehat{\Omega}_0 + 2\sum_{h=1}^{B} K\left(\frac{h}{B}\right)(\widehat{\Omega}_h + \widehat{\Omega}_h'), \tag{3.18}$$

where

$$\widehat{\Omega}_h = \frac{1}{T}\sum_{t=h+1}^{T} \widehat{v}_t \widehat{v}_{t-h}',$$

for $h = 0, \cdots, B$ while B indicating the bandwidth. We may also construct a variance estimator with centered residual such as

$$\tilde{V} = \tilde{\Omega}_0 + \sum_{h=1}^{B} K\left(\frac{h}{B}\right)(\tilde{\Omega}_h + \tilde{\Omega}_h') \tag{3.19}$$

where $\tilde{\Omega}_h = T^{-1}\sum_{t=h+1}^{T}(\widehat{v}_t - \overline{v}_T)(\widehat{v}_{t-h} - \overline{v}_T)'$, $\overline{v}_T = T^{-1}\sum_{t=1}^{T}\widehat{v}_t$.

Regarding the consistency of $\widehat{V}$, we consider an additional set of assumptions that are sufficient to guarantee the consistency of $\widehat{V}$ and $\tilde{V}$ regarding their target V. To begin with, we need a more restrictive case then the Condition 3.2 to regularize the degree of autocorrelation.

Condition 3.5. *Let χ_t be a stochastic process with the following properties: i) Let $\{\alpha_m\}_{m=1}^{\infty}$ and $\{\beta_m\}_{m=1}^{\infty}$ be the strong and uniform mixing coefficients, respectively. For some $r \in (2, 4]$ and $p > r$, either one of the followings is true:*

$$\sum_{m=0}^{\infty} \alpha_m^{2(1/r-1/p)} < \infty, \qquad \sum_{m=0}^{\infty} \beta_m^{1-2/p} < \infty$$

ii) For some $\delta > 0$, $E\|\chi_t\|^{\eta} < \infty$ where $\eta = \max\{4+\delta, p\}$

The symmetric kernel $K(\cdot)$ and bandwidth B are chosen to satisfy the following regulatory conditions:

Condition 3.6. *The kernel function $K(\cdot)$ satisfies the following properties: i) $K(\cdot)$ is symmetric around zero, i.e. $K(u) = K(-u)$, ii) $\sup_{u\in[0,1]} |K(u)| \leq 1$, (iii) $\int_{\mathbb{R}} |K(u)|du < \infty$, and iv) $\int_{\mathbb{R}} |K(u)||u|du < \infty$.*

Condition 3.7. *$B = B_T$ such that for some positive sequence $\mu_{\kappa,T} \to 0$, i) $B_T T^{-1} = o(1)$, ii) $B_T \mu_{\kappa,T} = o(1)$, and iii) $B_T \mu_{\kappa,T} \|\widehat{\varphi} - \varphi^*\| = o_p(1)$.*

With the additional assumptions stated above, one can show that the long-run variance estimators (3.18) and (3.19) are both consistent with V.

Theorem 3.3. *Assume that Theorem 3.1–3.2 hold. In addition, if Conditions 3.5, 3.6, and 3.7 are satisfied, then $\|\widehat{V} - V\| = o_p(1)$ and $\|\tilde{V} - V\| = o_p(1)$.*

The proof of Theorem 3.3 is given in the Appendix. While it is sufficient to choose any type of symmetric kernel satisfying the Condition 3.6 with a sequence of bandwidth B_T that is mildly increasing in the sense of Condition 3.7, we use Bartlett kernel and optimal bandwidth of Newey and West (1994) in the numerical example.

3.5.2 Inference Based on Consistent Variance Estimator

Once we have a consistent estimator for the long-run variance, inference problem becomes straightforward. Using the estimates $\widehat{\alpha}$ and $\widehat{V}$, one can construct a test statistics as follow. Let $\Omega^{-1} = \dot{\theta}(\alpha_0)'V^{-1}\dot{\theta}(\alpha_0)$ be the inverse of the long-run variance of α and $\widehat{\Omega} = \dot{\theta}(\widehat{\alpha})'\widehat{V}^{-1}\dot{\theta}(\widehat{\alpha})$ be the estimate of it. As $\dot{\theta}(\cdot)$ is assumed to be continuous, one can expect that $\|\widehat{\Omega} - \Omega\| = o_p(1)$ by the result of Theorem 3.3 and continuous mapping theorem. Then the following theorem holds by the result of asymptotic normality of $\widehat{\alpha}$ (Theorem 3.1), consistency of $\widehat{V}$ (Theorem 3.3), and the Cramér convergence theorem.

PROPOSITION 3.1. *Suppose that $\alpha_0 \in \mathcal{A}$ uniquely satisfies $E[h_t(\xi_t, \alpha_0, p_0)] = \theta(\alpha_0)$. If the regularity conditions in Theorem 3.3 are satisfied, we have*

$$\sqrt{T}\widehat{\Omega}^{-1/2}(\widehat{\alpha} - \alpha_0) \ \stackrel{d}{\to} \ N(0, I_{d_\alpha})$$

where $d_\alpha = \dim(\mathcal{A})$.

The result is useful in building standard types of test statistics such as t-test and Wald test. First, we may construct the t-statistic for the significance test of individual responses by picking an arbitrary component in $\widehat{\alpha}$. Let e_k be a unit vector in $\mathbb{R}^{d_\alpha}$ with one in k-th component and zero elsewhere. Also, denote $\widehat{\alpha}_k$ as the k-th component in $\widehat{\alpha}$. That is, $\widehat{\alpha}_k = e_k' \widehat{\alpha}$. Then by continuous mapping theorem, we get the following result:

COROLLARY 3.1 (t-Statistic). *Given that Proposition 3.1 holds, we have*

$$t_T = \frac{\widehat{\alpha}_k - \alpha_{0,k}}{\sqrt{\frac{1}{T} e_k' \widehat{\Omega} e_k}} \ \stackrel{d}{\to} \ N(0, 1),$$

for all $k = 1, \cdots, d_\alpha$.

In addition to this, we also construct the Wald statistic. Denote χ^2_d for the chi-square distribution with the degrees of freedom d. Then the result follows by continuous mapping theorem:

COROLLARY 3.2 (Wald-Statistic). *Given that Proposition 3.1 holds, we have*

$$W_T = T(\widehat{\alpha} - \alpha_0)'\widehat{\Omega}^{-1}(\widehat{\alpha} - \alpha_0) \ \xrightarrow{d}\ \chi^2_{d_\alpha}.$$

Finally, consider the case $d_\theta > d_\alpha$ where $d_\theta = \dim(\Theta) = k_y \times (L+1) \times J$. By construction, we have the number of moment restrictions more than necessary to identify parameters in the second-stage estimation. Such case allows us to apply the J-test of over-identification by Hansen (1982). The definition of J-statistic follows that of Ackerberg et al. (2012) and Chen and Liao (2015), which is a natural extension of Hansen (1982) into semi-parametric framework. Specifically, the test statistic is written as follow:

$$J_T = \Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{h}_t(\xi_t, \widehat{\alpha}, \widehat{\varphi}_\kappa) - \theta(\widehat{\alpha})\Big)'\widehat{V}^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{h}_t(\xi_t, \widehat{\alpha}, \widehat{\varphi}_\kappa) - \theta(\widehat{\alpha})\Big) \tag{3.20}$$

Then the limiting distribution of J_T will be a chi-square distribution with degrees of freedom equal to the difference between the dimension of θ, number of moment conditions, and that of α, the number of free parameters.

PROPOSITION 3.2. *Suppose that the conditions in the above theorem holds. If the correct null specification is* $E[h_t(\xi_t, \alpha_0, p_0)] = \theta(\alpha_0)$ *and* $d_\theta > d_\alpha$*, then we have* $J_T \to^d \chi^2_{d_\theta - d_\alpha}$.

It is easy to interpret the J-test as a test of the validity of (possibly non-linear) restriction given by $\theta = \theta(\alpha)$. For example, suppose that we want to test the symmetry of inflation response to the monetary policy shock. We may use a series of de-trended aggregate prices changes for the outcome variable y_t. And for discrete policy variables, we may create a set of dummy variables indicating the changes in policy rates. Specifically, let D_t be a 3×1 vector with j-th element denoted by $D_{t,j}$. For each $j \in \{0, 1, 2\}$, we set $D_{t,0} = 1\{\Delta PR_t = 0\}$, $D_{t,1} = 1\{\Delta PR_t = -0.025\}$, and $D_{t,2} = 1\{\Delta PR_t = 0.025\}$ where ΔPR_t implies the changes in policy rate. Regarding the moment restriction, for any positive integer L, we may consider a matrix of constants $R = (-1, 1)' \otimes I_{L+1}$ and set $\theta(\alpha) = R\alpha$ which is equivalent to $\theta = (-\alpha', \alpha')'$. Then we can test the symmetry of impulse-response functions using the limiting distribution of J-statistic which is, according to the Proposition 3.2, equivalent to χ^2_{L+1}.

3.6 Monte Carlo

3.6.1 Data Generating Process

We consider a simple AR(1) process augmented with an exogenous policy intervention. More precisely, the DGP is written as follows:

$$y_t = \tau d_t + \rho y_{t-1} + \varepsilon_t \tag{3.21}$$

where $d_t \in \{0, 1\}$ indicates the exogeneously given discrete policy intervention and $\varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$ is a random innovation. Discrete policy intervention is assumed to be a Bernoulli random process with "success" probability conditional on a covariate z_t is denoted by $p(z_t)$. Policy-score covariate z_t is assumed to be an iid process with Gaussian distribution with mean μ_z and variance σ_z^2. Then conditioning on z_t, the probability structure of the the discrete policy variable d_t is characterized as follow:

$$d_t|\mathcal{F}_t = \begin{cases} 1 & \text{with probability } p(z_t) \\ \\ 0 & \text{with probability } 1 - p(z_t) \end{cases},$$

where the conditional probability is explicitly written by $p(z_t) = 1 - \Phi(\delta - z_t\beta)$ with standard normal CDF $\Phi(\cdot)$.

The lag coefficient ρ is chosen among three candidates to reflect various degrees of persistence, while keeping the y_t process to be stationary. Later we discuss numerical result regarding the changes implied by y_t getting closer to the unit root case.

3.6.2 Benchmark Parametric Estimation

We begin with deriving the explicit formula of the actual dynamic causal effect implied by the DGP given as (3.21). For any positive integer l, we get the following decomposition by recursive substitution

$$\begin{aligned} y_{t+l} &= \tau d_{t+l} + \rho y_{t+l-1} + \varepsilon_{t+l} \\ &= \tau d_{t+l} + \rho(\tau d_{t+l-1} + \rho y_{t+l-2} + \varepsilon_{t+l-1}) + \varepsilon_{t+l} \\ &= \tau d_{t+1} + \rho\tau d_{t+l-1} + \rho^2 y_{t+l-2} + \varepsilon_{t+l} + \rho\varepsilon_{t+l-1} \\ &= \cdots \\ &= \tau \sum_{j=0}^{l-1} \rho^j d_{t+l-j} + \rho^l y_t + \sum_{j=0}^{l-1} \rho^j \varepsilon_{t+l-j}. \end{aligned}$$

By taking the conditional expectation on both sides, we have the predicted value of y_{t+l} as follow:

$$E[y_{t+l}|\mathcal{F}_t] = \tau \sum_{j=0}^{l-1} \rho^j E[d_{t+l-j}|\mathcal{F}_t] + \rho^l y_t$$

The dynamic causal effect is defined by the difference between two predicted outcomes, one is conditioned on $d_t = 1$ while the other one is conditioned on $d_t = 0$. That is,

$$\theta_l \equiv E[y_{t+l}|d_t = 1] - E[y_{t+l}|d_t = 0] = \rho^l \tau \tag{3.22}$$

and by accumulating the effects, we get the cumulative IRF as $\theta_l^c \equiv \sum_{j=0}^{l} \theta_j$.

As a benchmark case, we include two parametric estimates for comparison of the performance with the semi-parametric estimates. Given that we know the parametric specification of outcome process (3.21), it is easy to think of estimating the structural parameters, τ and ρ, directly and get an estimate of θ_l by plug-in method. More precisely, estimation is implemented with a simple two-step procedure. First, regress y_t on d_t and its lagged observation, y_{t-1}, to get the estimates $\widehat{\tau}$ and $\widehat{\rho}$. Then construct the estimator of dynamic causal effect by replacing ρ and τ in (3.22) for their corresponding estimator, such as $\widehat{\theta}_l = \widehat{\rho}^l \widehat{\tau}$. Obviously, cumulative effect can be easily obtained by accumulating $\widehat{\theta}_l$'s.

In addition to this, we also consider the local projection method by Jordà (2005). The estimation is done by the following steps. First, regress y_t's on its own lags and d_t's to get the residuals. Denote $\widehat{\varepsilon}_t = y_t - \widehat{\tau} d_t - \widehat{\rho} y_{t-1}$ where $\widehat{\tau}$ and $\widehat{\rho}$ are the OLS estimates of τ and ρ, respectively. Then project y_t onto $\{1, \widehat{\varepsilon}_t, \widehat{\varepsilon}_{t-1}, \cdots, \widehat{\varepsilon}_{t-L}\}$ where L is a pre-fixed integer which implies the maximum lag of the impulse-response

Table 3.1: Parameter values used in Monte Carlo experiment.

						ρ		
τ	δ	β	μ_z	σ_z	σ_ε	low	mid	high
0.33	0.90	0.85	1.00	0.20	0.04	0.70	0.85	0.98

function. Let $\pi = (\pi_0, \tilde{\pi}')'$ be the projection coefficient of size $(L+1) \times 1$, with π_0 indicating the projection coefficient corresponding to the intercept. Jordà (2005) argues that the l-th component in $\tilde{\pi}$, denoted by $\tilde{\pi}_l$, is a local approximation of $Cov(y_t, \varepsilon_{t-l})$, which is coherent with the l-th period impulse-response function. Then the estimate of l-period impulse-response to the discrete policy shock is given by $\widehat{\tau}\tilde{\pi}_l$.

3.6.3 Numerical Results

We simulate 287 observations per each repetition among the total of 1000 based on to the DGP given by (3.21) and estimate $L = 36$ periods of impulse-response functions. The first 37 observations are used as initial conditions, implying that the effective sample size would be 250. The set of parameter values used to simulate the data is summarized in Table 3.1.

Point estimates of impulse-response functions are given in the table C.1–C.3. Each columns indicates the estimates obtained by different options including two parametric cases and semi-parametric estimation with first-stage policy score estimates computed by ordered probit model, multinomial logit model, and sieve approximation via polynomial, trigonometric polynomial, Hermite polynomial, and wavelet basis. Table C.1 is the result with date simulated under lowest persistence level, $\rho = 0.70$. As expected, estimates in all of the different types approach to zero fast due to a low level of persistence. While every estimates show mild degree of under estimation, not surprisingly, two parametric benchmarks produce more accurate result compared to semi-parametric estimates. Monte Carlo standard deviations, numbers shown in the parenthesis, also suggests that the estimates are more dispersed across different simulations when we run semi-parametric estimation.

The precision seems to be worsen as we increase the persistence level in time series process. The result in table C.2 is analogous to that of table C.1 while the level of persistence is increased to $\rho = 0.85$. We still observe the tendency to under estimate impulse-responses while the degree of under-estimation is worsen as we compute semi-parametrically. Monte Carlo standard deviations also increase at about 0.02 points compared to the case with the lowest persistence level. Then it is easy to expect that the problem will become more serious as we increase the persistence level further close to unit root case. Table C.3 proves such conjecture. Now the Monte Carlo standard deviation is increased roughly upto 0.15 points in ordered probit case and 0.185 in sieve estimation with wavelet basis.

One of the possible reasons for the inaccuracy comes from the bias in the first-stage estimates of policy score functions. Figure C.1 compares policy score estimates in various semi-parametric estimation methods. While parametric first-stage estimates–that is, ordered probit and multinomial logit–relatively well-approximates the true conditional probabilities, estimation bias appears to be more serious in some of the sieve approximations. According to the numerical result shown in our numerical example, it may be safer to use simple geometric polynomial and Hermite polynomial to minimize approximation bias in finite sample.

Table C.4–C.6 show the mean squared errors for each of the point estimates of impulse-response functions while underlying process varies in terms of its level of persistence. In all of the three tables, two parametric benchmark outperform semi-parametric estimates. By comparing the accuracy among different semi-parametric estimates, we observe stable numbers across different columns in table C.4–implying that there is minor information loss in using sieve approximation in the first stage instead of estimation policy scores under correct parametric specification. However, in table C.5, we observe slight increases in mean squared errors for the sieve estimates, especially in trigonometric polynomial and wavelet basis. The result corresponds to the fact shown in figure C.1 where we see larger bias in policy score estimates using trigonometric polynomial and wavelet basis. The mean squared errors increase even further as persistence level approaches one. In table C.6, we see the numbers are roughly doubled between semi-parametric specifications and local projection.

Figure C.2 helps finding one of the source of inaccuracy in semi-parametric impulse-response estimators. The left column in the figure plots evolution of mean squared errors at each time periods while the right column plots their bias in absolute level. In the first row, we see that although there is relatively larger fluctuations in absolute bias, mean squared error is more consistent across the periods. It suggests that mean squared error is mostly determined by the variance of estimators, not by their bias. However, as we increase the level of persistence, bias becomes much larger and stable across the periods. Hence, the result shows that it is worthwhile to consider a bias correction technique especially when the underlying process is believed to have higher level of persistence.

Lastly, we illustrate the empirical size to check if asymptotic result shown in section 3.5 hold in our numerical example. Numbers in table C.7–C.9 are the fractions of estimates that are rejected via t-test.[2] The size is controlled at 5% implying that one should expect, if the test is well-specified, actual number of rejections should be close to 1 for each 20 simulations. In table C.7, although we see some mild deviations from 0.05, empirical size is roughly consistent with the pre-fixed level. However, in table C.8, slight increases in actual number of rejections can be found in some cases. The increase in empirical size is especially noticeable in the earlier periods with estimates via sieve approximations. The problem becomes worse in table C.9 of which underlying process has the highest level of persistence. In some cases, empirical size reaches upto 0.15 which is the triple of the actual size.

3.7 Conclusion

In this chapter, we developed a semi-parametric estimator for dynamic causal effects with discrete policy shocks. The estimator takes the form of inverse propensity score weighted average which is robust to non-linear and asymmetric impulse-response functions. Using a recently proposed asymptotic theory, we show that the semi-parametric estimator is uniformly consistent and asymptotically normal when the

[2]The test statistic is constructed following the result in corollary 3.1 while α_0 is set to be the actual impulse-response function implied by the parametric specification 3.22.

underlying data is weakly dependent. As for the inference, we exploit the asymptotic normality and construct Wald and F-statistics that can be used to test both linear and non-linear hypotheses. Test statistics are constructed by substituting a semi-parametric estimator for the asymptotic variance matrix which is shown to be consistent under weakly dependent process. Monte Carlo experiments show that the estimator performs well even when the true impulse-response function is non-linear and asymmetric.

Chapter A: Appendix to Chapter 1

A.1 Series Estimation of Propensity Score Function

As the propensity score function $p(x)$ is unknown and possibly nonlinear in x, I consider the series approximation to estimate $p(x)$ in more flexible setting. Asymptotic properties of series estimators are discussed earlier by, for example, Geman and Hwang (1982) and Newey (1997). Hirano et al. (2003) develop the series estimation theory to construct the efficient semi-parametric estimator for causal parameters. While previous studies rely on the assumption that the support $\mathcal{X}$ is compact subset of $\mathbb{R}^{d_x}$, I adopt new asymptotic theory recently developed by Belloni et al. (2015) and Hansen (2015) to extend the case to smooth functions over possibly non-compact support.

Let $R^{\kappa}(x) = (r_1(x), r_2(x), \cdots, r_{\kappa}(x))'$ so that $R^{\kappa}(x)$ becomes a triangular array of the collection of basis functions $r_j : \mathbb{R}^{d_x} \to \mathbb{R}$ for $j = 1, 2, \cdots, \kappa$. The idea is to approximate a generic, unknown function $g : \mathbb{R}^{d_x} \to \mathbb{R}$ by a linear combination of the basis functions, which is given by $P^{\kappa}(x)'\gamma_{\kappa}$ for some vector of constants $\gamma_{\kappa} \in \mathbb{R}^{\kappa}$. Note that for any unitary matrix $H_{\kappa} \in \mathbb{R}^{\kappa \times \kappa}$, we get $R^{\kappa}(x)'\gamma_{\kappa} = R^{\kappa}(x)'H_{\kappa}^{-1}H_{\kappa}\gamma_{\kappa}$.

Therefore, we may choose H_κ properly so that $\Psi^\kappa(x) = H_\kappa R^\kappa(x)$ satisfies

$$\int_{\mathcal{X}} \|\Psi^\kappa(x)\Psi^\kappa(x)'\| d\mu(x) = I_\kappa$$

Approximation Theory By choosing a properly normalized basis, I restrict the size of triangular array $\Psi^\kappa(x)$. Specifically, consider $\Psi^\kappa(x)$ such that $\|\Psi^\kappa(x)\|_\eta = O(z_\kappa)$ for some positive sequence z_κ which is weakly increasing with respect to κ. This condition replaces the usual sup-norm based condition that is not well-defined in case the support $\mathcal{X}$ is possibly non-compact. The bounding sequence z_κ is specific to the type of basis functions. Some of the known examples in one dimensional space are the following. For polynomial basis $\Psi^\kappa(x) = (1, x, x^2, \cdots, x^{\kappa-1})'$ and $z_\kappa = O(\kappa)$. For Fourier series, wavelets, and splines, $z_\kappa = O(\kappa^{1/2})$.[1]

Validity of linear approximation is supported by a classical result in functional analysis. The key is to find a vector $\gamma_\kappa \in \mathbb{R}^\kappa$ which satisfies the following equation:

$$\left(\int_{\mathcal{X}} \|g(x) - \Psi^\kappa(x)'\gamma_\kappa\|^\eta d\mu(x)\right)^{\frac{1}{\eta}} = O(b_\kappa) \tag{A.1.1}$$

where $b_\kappa \to 0$ as $\kappa \to \infty$. Hansen (2015) lists a set of necessary conditions under which the linear approximation is valid. Suppose that a function g over $\mathbb{R}^{d_x}$ is s times continuously differentiable in the following sense: for all $|\mathbf{s}| \leq s$, $(\int \|\nabla^{\mathbf{s}} g(x)\|^\eta \exp(-A\|x\|_\alpha)dx)^{1/\eta} < \infty$ for some $A > 0$ and $\alpha \geq 2$. Then the largest deviation of a linear approximation $\Psi^\kappa(x)'\gamma_\kappa$ from $g(x)$ is bounded as fol-

[1] See Belloni et al. (2015) and Hansen (2015) for the list of known basis functions and their bounding sequences.

lows (Mhaskar, 1996; Maiorov and Meir, 1998):

$$\inf_{\gamma_\kappa \in \mathbb{R}^\kappa} \Big(\int \| g(x) - \Psi^\kappa(x)'\gamma_\kappa \|^\eta \exp(-A\|x\|_\alpha) dx \Big)^{1/\eta} = O(\kappa^{-\frac{s}{d_x}(1-\frac{1}{\alpha})})$$

Then it follows that if $\mu(x) \leq C \exp(-A\|x\|_\alpha)$ for some $C < \infty$, equation (A.1.1) is satisfied with $b_\kappa = O(\kappa^{-\frac{s\eta}{d_x}(1-1/\alpha)})$.

Series Estimation of Propensity Score To ensure that the target of approximation stays in between $[0, 1]$, consider approximating the log odds ratio. Define γ_κ as a vector of constant which satisfies

$$\Big(\int_{\mathcal{X}} \Big\| \ln \frac{p(x)}{1 - p(x)} - \Psi^\kappa(x)'\gamma_\kappa \Big\|^\eta d\mu(x) \Big)^{\frac{1}{\eta}} = O(\kappa^{-\frac{s}{d_x}(1-\frac{1}{\alpha})}) \tag{A.1.2}$$

Note that the transformation $\rho(z) = \ln \frac{z}{1-z}$ is monotonically increasing and smooth over a closed interval $z \in [\underline{p}, \overline{p}] \subset (0, 1)$. Therefore, it follows from the previous discussion that γ_κ satisfying (A.1.2) exists only if $p(x)$ is s times differentiable and $\nabla^{\mathbf{s}} p(x)$ is absolutely continuous and bounded in $\mathcal{L}_\eta(\mu)$-norm.

It is natural to find the best linear approximation by maximizing a criterion function implied by the log-likelihood. Let $L(u) = \exp(u)/(1 + \exp(u))$, the logistic CDF. Two versions of the series estimators are defined as follows. First, denote the empirical estimator of $p(x)$ by $\widehat{p}_\kappa(x) \equiv L(\Psi^\kappa(x)'\widehat{\gamma}_\kappa)$ for $\widehat{\gamma}_\kappa = \arg\max_{\gamma \in \mathbb{R}^\kappa} L_n(\gamma)$ with

$$L_n(\gamma) \equiv \sum_{i=1}^{n} \Big(D_i \ln L(\Psi^\kappa(X_i)'\gamma) + (1 - D_i) \ln(1 - L(\Psi^\kappa(X_i)'\gamma)) \Big)$$

On the other hand, we may also define the population-level estimator as $p^*_\kappa(x) \equiv L(\Psi^\kappa(x)'\gamma^*)$ for $\gamma^*_\kappa = \arg\max_{\gamma \in \mathbb{R}^\kappa} L^*(\gamma)$ with

$$L^*(\gamma) \equiv E\Big[p(X) \ln L(\Psi^\kappa(X)'\gamma) + (1 - p(X)) \ln(1 - L(\Psi^\kappa(X)'\gamma)) \Big]$$

The first step is to show that the propensity score function $p(x)$ is well-approximated by the population-level estimator $p^*_\kappa(x)$. The following lemma is the formal statement of the claim:

Lemma A.1.1. *Suppose that the following conditions hold: (i) $\mathcal{X} \subseteq \mathbb{R}^{d_x}$, (ii) for all $|\mathbf{s}| \leq s$, $\nabla^{\mathbf{s}} p(x)$ is absolutely continuous and bounded–i.e. for some $A > 0$, $(\int \|\nabla^{\mathbf{s}} p(x)\|^\eta \exp(-A\|x\|_2)dx)^{1/\eta} < \infty$, (iii) $\mu(x)$ is bounded away from zero on $\mathcal{X}$, and (iv) $\mu(x) \leq C\exp(-A\|x\|_2)$. Then, for γ_κ satisfying* (A.1.2),

$$\Big(\int_{\mathcal{X}} \|p(x) - p^*_\kappa(x)\|^\eta d\mu(x)\Big)^{\frac{1}{\eta}} = O(\kappa^{-\frac{s}{2d_x}} z_\kappa)$$

Proof. The proof follows from the result of Hansen (2015, Theorem 6) by showing that the difference between actual and approximated propensity score functions is bounded at the same rate of (A.1.2). Let $\xi = \sup_{x\in\mathcal{X}} \|p(x)(1-p(x))\|$ which satisfies $\xi \in (0,1)$ by Condition 1.2. Note that since $L(u)$ is continuously differentiable over $\mathbb{R}$. For any $u_1, u_2 \in \mathbb{R}$ such that $u_1 \leq u_2$, there exists $\bar{u} \in [u_1, u_2]$ satisfying $\|L(u_1) - L(u_2)\| \leq \|L'(\bar{u})\|\|u_1 - u_2\|$ by the mean value theorem. Then, for all $x \in \mathcal{X}$,

$$\|p(x) - p^*_\kappa(x)\| = \Big\|L\Big(\ln\frac{p(x)}{1-p(x)}\Big) - L(\Psi^\kappa(x)'\gamma^*_\kappa)\Big\| \leq \xi\|g(x) - \Psi^\kappa(x)'\gamma^*_\kappa\|$$

Hence the result follows for any $\eta \geq 1$. □

It also needs to be shown that the $\widehat{p}_\kappa(x)$ is close enough to the population-level approximation which is denoted by $p^*_\kappa(x)$ previously. The claim can be verified by the following lemma which argues $\widehat{\gamma}_\kappa$ converges to γ^*_κ with a proper choice of bandwidth κ.

LEMMA A.1.2. *Suppose that the Conditions 1.1, 1.2, and that stated in Lemma A.1.1 hold. In addition, let $\kappa = \kappa_n$ where κ_n is a non-stochastic sequence satisfying $\kappa_n \to \infty$ and $n^{-1} z_\kappa^{\frac{2\eta}{\eta-2}} \log \kappa = o(1)$ as $n \to \infty$. Then,*

$$\|\widehat{\gamma}_{\kappa_n} - \gamma^*_{\kappa_n}\| = O_p(\sqrt{\frac{\kappa_n}{n}})$$

Proof. The proof follows the argument in Lemma 1 of Hirano et al. (2003). Let $\widehat{M}_\kappa = \frac{1}{n}\sum_{i=1}^n \Psi^\kappa(X_i)\Psi^\kappa(X_i)'$ which satisfies $E[\widehat{M}_\kappa] = I_\kappa$. Following result is from Lemma 2 of Hansen (2015):

$$\|\widehat{M}_\kappa - I_\kappa\| = O_p\Big(z_\kappa^{\frac{\eta}{\eta-2}}\sqrt{\frac{\log\kappa}{n}}\Big) \tag{A.1.3}$$

Since the rate $z_\kappa^{\frac{\eta}{\eta-2}}\sqrt{\log\kappa/n}$ converges to zero as stated, the result (A.1.3) also implies that $\|\widehat{M}_\kappa - I_\kappa\| = o_p(1)$. Then for an arbitrarily small constant $\varepsilon > 0$, $\|\lambda_{\min}(\widehat{M}_\kappa) - 1\| \geq \varepsilon$ with probability approaching to one. Without loss of generality, assume that $\lambda_{\min}(\widehat{M}_\kappa) \leq 1$ and pick $\varepsilon = 1/2$.

Note that $L_n(\gamma)$ is concave over all $\gamma \in \mathbb{R}^\kappa$ and, by definition, $\frac{1}{n}\frac{\partial L_n(\widehat{\gamma}_\kappa)}{\partial\gamma} = 0$. Following the proof of Lemma 1 in Hirano et al. (2003), I show that the first order condition evaluated at $\gamma = \gamma^*_\kappa$ also satisfies the first order condition with probability approaching to one. It can be shown that

$$\begin{aligned} E\Big[\Big\|\frac{1}{n}\frac{\partial L_n(\gamma^*_\kappa)}{\partial\gamma}\Big\|^2\Big] &= \frac{1}{n}E\Big[\mathrm{tr}\Big((D_i - L(\Psi^\kappa(X_i)'\gamma^*_\kappa))^2\Psi^\kappa(X_i)\Psi^\kappa(X_i)'\Big)\Big] \\ &\leq \frac{1}{n}\mathrm{tr}(E[\Psi^\kappa(X_i)\Psi^\kappa(X_i)']) = \frac{\kappa}{n} \end{aligned}$$

For an arbitrary constant C, the Markov inequality yields

$$\mathrm{Prob}\Big(\Big\|\frac{1}{n}\frac{\partial L_n(\gamma^*_\kappa)}{\partial\gamma}\Big\| \geq C\sqrt{\frac{\kappa}{n}}\Big) \leq \frac{1}{C}\frac{n}{\kappa}E\Big[\Big\|\frac{1}{n}\frac{\partial L_n(\gamma^*_\kappa)}{\partial\gamma}\Big\|^2\Big] = \frac{1}{C} \tag{A.1.4}$$

which implies that $\left\|\frac{1}{n}\frac{\partial L_n(\gamma_\kappa^*)}{\partial\gamma}\right\| = O_p\Big(\sqrt{\frac{\kappa}{n}}\Big)$. Let

$$\xi \equiv \inf_{x\in\mathcal{X}} L(\Psi^\kappa(x)'\gamma_\kappa^*)(1 - L(\Psi^\kappa(x)'\gamma_\kappa^*)) \tag{A.1.5}$$

which is strictly positive from Condition 1.2 and the result of Lemma A.1.1. For arbitrary $\epsilon > 0$, we may choose a constant C and large enough n such that

$$\text{Prob}\Big(\Big\|\frac{1}{n}\frac{\partial L_n(\gamma_\kappa^*)}{\partial\gamma}\Big\| < \xi C\sqrt{\frac{\kappa}{n}}\Big) \geq 1 - \epsilon \tag{A.1.6}$$

Consider the second-order expansion of $L_n(\gamma)$ for an arbitrary γ within a shrinking neighborhood $\Gamma_n \equiv \{\gamma \in \mathbb{R}^\kappa : \|\gamma - \gamma_\kappa^*\| \leq C\sqrt{\kappa/n}\}$. That is,

$$\frac{1}{n}L_n(\gamma) = \frac{1}{n}L_n(\gamma_\kappa^*) + \frac{1}{n}\frac{\partial L_n(\gamma_\kappa^*)}{\partial\gamma}(\gamma - \gamma_\kappa^*) + \frac{1}{2n}(\gamma - \gamma_\kappa^*)'\frac{\partial^2 L_n(\bar{\gamma})}{\partial\gamma\partial\gamma'}(\gamma - \gamma_\kappa^*)$$

for some $\bar{\gamma}$ satisfying $\|\bar{\gamma} - \gamma_\kappa^*\| \leq \|\gamma - \gamma_\kappa^*\| \leq C\sqrt{\kappa/n}$. Note that

$$\begin{aligned}\frac{1}{n}\frac{\partial^2 L_n(\bar{\gamma})}{\partial\gamma\partial\gamma'} &= -\frac{1}{n}\sum_{i=1}^n L(\Psi^\kappa(X_i)'\bar{\gamma})(1 - L(\Psi^\kappa(X_i)'\bar{\gamma}))\Psi^\kappa(X_i)\Psi^\kappa(X_i)' \\ &\leq -\frac{2\xi}{n}\sum_{i=1}^n \Psi^\kappa(X_i)\Psi^\kappa(X_i)' = -2\xi\widehat{M}_\kappa \end{aligned} \tag{A.1.7}$$

where the inequality (A.1.7) holds since ξ is the uniform lower bound as defined by (A.1.5). Given that $\lambda_{\min}(\widehat{M}_\kappa) \geq 1/2$ with probability approaching to one, we get $\|\frac{1}{n}\frac{\partial^2 L_n(\bar{\gamma})}{\partial\gamma\partial\gamma'}\| \leq -\xi$ for n large enough. Therefore, for any γ satisfying $\|\gamma - \gamma_\kappa^*\| = C\sqrt{\kappa/n}$,

$$\begin{aligned}\frac{1}{n}L_n(\gamma) - \frac{1}{n}L_n(\gamma_\kappa^*) &\leq \Big\|\frac{1}{n}\frac{\partial L_n(\gamma_\kappa^*)}{\partial\gamma}\Big\|\|\gamma - \gamma_\kappa^*\| + \frac{1}{2}\Big\|\frac{1}{n}\frac{\partial^2 L_n(\bar{\gamma})}{\partial\gamma\partial\gamma'}\Big\|\|\gamma - \gamma_\kappa^*\|^2 \\ &\leq \Big\|\frac{1}{n}\frac{\partial L_n(\gamma_\kappa^*)}{\partial\gamma}\Big\|\|\gamma - \gamma_\kappa^*\| + \xi\|\gamma - \gamma_\kappa^*\|^2 \\ &\leq \Big(\Big\|\frac{1}{n}\frac{\partial L_n(\gamma_\kappa^*)}{\partial\gamma}\Big\| - \xi C\sqrt{\frac{\kappa}{n}}\Big)\|\gamma - \gamma_\kappa^*\| < 0\end{aligned}$$

Therefore, $L_n(\gamma) < L_n(\gamma_\kappa^*)$ with probability greater than $1-\varepsilon$ for all $\gamma \in \Gamma_n \setminus \text{int}\Gamma_n$. Thus we have the maximum $\widehat{\gamma}_\kappa$ must be in the interior of Γ_n for n sufficiently large which implies $\|\widehat{\gamma}_\kappa - \gamma_\kappa^*\| = O_p(\sqrt{\kappa/n})$. □

Suppose that the number of terms in the series expansion $\kappa = \kappa_n$ increases with respect to the sample size n. The rate of increment should be carefully managed with a choice of basis functions. In a later part of this section and the results in Section 1.4, I consider multi-dimensional polynomial series for $\Psi^\kappa(x)$. This implies a bounding sequence of $z_\kappa = O(\kappa^{d_x})$ as discussed above. In this case, we may further specify the rate condition to be $\kappa_n^{4d_x} = o(n)$.

Denote $\widehat{\gamma}_n = \widehat{\gamma}_{\kappa_n}$, $\gamma_n^* = \gamma_{\kappa_n}^*$, $\widehat{p}_n(x) = \widehat{p}_{\kappa_n}(x)$, $p_n^*(x) = p_{\kappa_n}^*(x)$ to reduce complexity in notation. The formal statement of the uniform consistency of series estimator is given in Lemma 1.6 in the main text. I present the proof of the statement below.

Proof of Lemma 1.6 By Minkowski inequality,

$$\begin{aligned}\|\widehat{p}_n(x) - p(x)\|_\eta &= \|\widehat{p}_n(x) - p_n^*(x) + p_n^*(x) - p(x)\|_\eta \\ &\leq \|\widehat{p}_n(x) - p_n^*(x)\|_\eta + \|p_n^*(x) - p(x)\|_\eta\end{aligned}$$

From the result of Lemma A.1.1, we know that $\|p_n^*(x) - p(x)\|_\eta = O(\kappa^{-\frac{s}{2d_x}} z_\kappa)$. In addition,

$$\begin{aligned}\|\widehat{p}_n(x) - p_n^*(x)\|_\eta &= \Big(\int_{\mathcal{X}} \|\Psi^\kappa(x)'\widehat{\gamma}_n - \Psi^\kappa(x)'\gamma_n^*\|^\eta d\mu(x)\Big)^{\frac{1}{\eta}} \\ &\leq \|\widehat{\gamma}_n - \gamma_n^*\| \Big(\int_{\mathcal{X}} \|\Psi^\kappa(x)\|^\eta d\mu(x)\Big)^{\frac{1}{\eta}} = O_p\left(\sqrt{\frac{\kappa_n}{n}} z_\kappa\right)\end{aligned}$$

as $\|\widehat{\gamma}_n - \gamma_n^*\| = O_p(\sqrt{\kappa_n/n})$ from Lemma A.1.2 and $\|\Psi^\kappa\|_\eta = O_p(z_\kappa)$ by the assumption given in the statement. Then the result follows by $n^{-1/2}\kappa_n^{1/2(s/d_x+1)} = o(1)$.

A.2 Proof of Results

A.2.1 Proof of Proposition 1.1

Let $F_{1|0}$ denote the conditional distribution of $Y(1)$ given $Y(0)$ and X. That is, $F_{1|0}(y_1|y_0, x) = \text{Prob}(Y(1) \le y_1 | Y(0) = y_0, X = x)$ for $y_1 \in \mathcal{Y}_1$, $y_0 \in \mathcal{Y}_0$, and $x \in \mathcal{X}$. For given x, Y_1 is stochastically increasing with respect to $Y(0)$ iff $F_{1|0}(y_1|y_0, x)$ is decreasing in y_0 for all $y_1 \in \mathcal{Y}_1$. With the Condition 1.3, we have

$$
\begin{aligned}
F_{1|0}(y_1|y_0, x) &= \text{Prob}(Y(0) + \Delta \le y_1 | Y(0) = y_0, X = x) \\
&= \text{Prob}(\Delta \le y_1 - y_0 | Y(0) = y_0, X = x) = \text{Prob}(\Delta \le y_1 - y_0 | X = x) \\
&= F_\Delta(y_1 - y_0 | x)
\end{aligned}
$$

where F_Δ is the CDF of Δ. Consider an arbitrary pair of $y_0, y_0' \in \mathcal{Y}_0$ such that $y_0 < y_0'$. Since F_Δ is a weakly increasing function over the support, we have $F_\Delta(y_1 - y_0) \ge F_\Delta(y_1 - y_0')$ for all y_1. Then the conclusion follows.

A.2.2 Proof of Lemma 1.1

Note that Conditions 1.1 and 1.2 collectively imply the following: for $j \in \{0, 1\}$, $Y(j) \perp D | p(X)$ (Rosenbaum and Rubin, 1983, Theorem 3). In addition,

$$
\begin{aligned}
E[D \exp(\iota\omega Y)|p(X)] &= E[\exp(\iota\omega Y)|D = 1, p(X)]\text{Prob}(D = 1|p(X)) \\
&\qquad + 0 \cdot \text{Prob}(D = 0|p(X)) \\
&= E[\exp(\iota\omega Y(1))|D = 1, p(X)]\text{Prob}(D = 1|p(X)) && \text{(A.2.1)} \\
&= E[\exp(\iota\omega Y(1))|p(X)]\text{Prob}(D = 1|p(X)) && \text{(A.2.2)}
\end{aligned}
$$

while (A.2.1) is equal to (A.2.2) because of the independence between $Y(1)$ and D conditional on $p(X)$. Then the result follows by dividing both sides of (A.2.2) with respect to $\text{Prob}(D = 1|p(X))$ which is bounded away from zero. Similarly, it can be easily shown that the characteristic function of $Y(0)$ is identified as follow:

$$E[(1 - D)\exp(\iota\omega Y)|p(X)] = E[\exp(\iota\omega Y(0))|p(X)]\text{Prob}(D = 0|p(X))$$

and the result follows from the fact that $\text{Prob}(D = 0|p(X)) = 1 - \text{Prob}(D = 1|p(X))$.

A.2.3 Proof of Lemma 1.2

Let $\tau, y \in \mathbb{R}$ and $z \in [\underline{p}, \overline{p}]$ be arbitrary values. Conditional distribution of Δ being less than or equal to τ given $Y(0) = y$ and $p(X) = z$ is given by

$$\begin{aligned}
\text{Prob}(\Delta \leq \tau|Y(0) = y, p(X) = z) &= \int_{\{x:p(x)=z\}} \text{Prob}(\Delta \leq \tau|Y(0) = y, X = x)d\mu(x) \\
&= \int_{\{x:p(x)=z\}} \text{Prob}(\Delta \leq \tau|X = x)d\mu(x) \\
&= \text{Prob}(\Delta \leq \tau|p(X) = z)
\end{aligned}$$

while the second equality is followed by the Condition 1.3. Therefore, the result follows.

A.2.4 Proof of Lemma 1.3

By definition, we have $Y(1) = Y(0) + \Delta$. Using the Condition 1.3, we may decompose the conditional characteristic function of $Y(1)$ as the product of the characteristic functions of $Y(0)$ and Δ. That is, $\varphi_1(\omega|z) = E[\exp(\iota\omega Y(1))|p(X) = z] = E[\exp(\iota\omega(Y(0) + \Delta))|p(X) = z] = E[\exp(\iota\omega Y(0))|p(X) = z]E[\exp(\iota\omega\Delta)|p(X) =$

$z] = \varphi_0(\omega|z)\varphi_\Delta(\omega|z)$. Given that $\varphi_0(\omega|z)$ is non-vanishing, we have $\varphi_\Delta(\omega|z) = \varphi_1(\omega|z)/\varphi_0(\omega|z)$. Then the result follows from the result of Lemma 1.1 that $\varphi_1(\omega|z)$ and $\varphi_0(\omega|z)$ are identified.

A.2.5 Proof of Lemma 1.4

Let q be a distribution which satisfies $\int_{\underline{p}}^{\overline{p}} dq(s) = 1$ and for any $z \in [\underline{p}, \overline{p}]$, $\int_{\underline{p}}^{z} dq(s) = \text{Prob}(p(X) \le z)$. First, we need to show that the unconditional characteristic function of the causal effect, denoted by $\varphi_\Delta(\omega)$, is identified.

$$
\begin{aligned}
E[\exp(\iota\omega\Delta)] &= \int_{\underline{p}}^{\overline{p}} E[\exp(\iota\omega\Delta)|p(X) = z]dq(z) \\
&= \int_{\underline{p}}^{\overline{p}} \frac{E[\exp(\iota\omega Y(1))|p(X) = z]}{E[\exp(\iota\omega Y(0))|p(X) = z]} dq(z) \\
&= \int_{\underline{p}}^{\overline{p}} \frac{1 - \text{Prob}(D = 1|p(X) = z)}{\text{Prob}(D = 1|p(X) = z)} \frac{E[D\exp(\iota\omega Y)|p(X) = z]}{E[(1-D)\exp(\iota\omega Y)|p(X) = z]} dq(z) \\
&= E\Big[\frac{1 - p(X)}{p(X)} \frac{E[D\exp(\iota\omega Y)|p(X)]}{E[(1-D)\exp(\iota\omega Y)|p(X)]}\Big]
\end{aligned}
$$

A.2.6 Proof of Proposition 1.2

Note that Condition 1.4 guarantee that $\varphi_0(\omega|z)$ is non-vanishing almost everywhere. Therefore, $\varphi_\Delta(\omega|z)$ is uniquely identified by Lemma 1.1. Then the remaining is to show that the expression (1.9) is well-defined, i.e. $\|f_\Delta(\tau|z)\| < \infty$. Note that,

$$
\begin{aligned}
\|f_\Delta(\tau|z)\| &\le \sup_{\tau\in\mathcal{T}} \Big\| \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-\iota\omega\tau) \frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)} d\omega \Big\| \\
&\le \frac{1}{2\pi} \int_{\mathbb{R}} \sup_{\tau\in\mathcal{T}} \|\exp(-\iota\omega\tau)\| \Big\| \frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)} \Big\| d\omega \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \Big\| \frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)} \Big\| d\omega
\end{aligned}
$$

by using the fact that $\|\exp(-\iota\omega\tau)\| = 1$ for all $\omega, \tau \in \mathbb{R}$ and the dominant convergence theorem. Hence, the result satisfies if and only if $\|\varphi_1(\omega|z)/\varphi_0(\omega|z)\|$ is integrable over $\mathbb{R}$. Since $\inf_{z\in[\underline{p},\overline{p}]} \|\varphi_0(\omega|z)\| \geq \underline{\Upsilon}_0(\omega)$, we have $\|\varphi_1(\omega|z)/\varphi_0(\omega|z)\| \leq \underline{\Upsilon}_0(\omega)^{-1}\|\varphi_1(\omega|z)\|$ almost everywhere on $\mathbb{R}$. By Hölder inequality,

$$\begin{aligned}\int_{\mathbb{R}}\Big\|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\Big\|d\omega &\leq \int_{\mathbb{R}}\|\varphi_1(\omega|z)\|\underline{\Upsilon}_0(\omega)^{-1}d\omega \\ &\leq \Big(\int_{\mathbb{R}}\|\varphi_1(\omega|z)\|^2 d\omega\Big)^{\frac{1}{2}}\Big(\int_{\mathbb{R}}\underline{\Upsilon}_0(\omega)^{-2}d\omega\Big)^{\frac{1}{2}}\end{aligned}$$

Note that $(\int_{\mathbb{R}}\underline{\Upsilon}_0(\omega)^{-2}d\omega)^{\frac{1}{2}} < \infty$ by assumption. Then the result holds if and only if $\int_{\mathbb{R}}\|\varphi_1(\omega|z)\|^2 d\omega$ is shown to be finite. Suppose that $\varphi_1(\omega|z)$ is ordinary smooth. From Condition 1.4 and the property of characteristic functions, there exists $M < \infty$ such that $\|\varphi_1(\omega|z)\| \leq 1$ for $|\omega| \leq M$ and $\|\varphi_1(\omega|z)\| \leq A|\omega|^{-\gamma}$ for $|\omega| > M$. Therefore,

$$\int_{\mathbb{R}}\|\varphi_1(\omega|z)\|^2 d\omega \leq 2\int_0^M 1d\omega + 2\int_M^\infty A|\omega|^{-2\gamma}d\omega = 2M + \frac{4A}{1-2\gamma}M^{1-2\gamma} < \infty$$

Suppose that $\varphi_1(\omega|z)$ is super-smooth. From Condition 1.4, there exists $M < \infty$ such that $\|\varphi_1(\omega|z)\| \leq A|\omega|^c\exp(-B|\omega|^\gamma)$ for $|\omega| > M$. Therefore,

$$\begin{aligned}\int_{\mathbb{R}}\|\varphi_1(\omega|z)\|^2 d\omega &\leq 2\int_0^M 1d\omega + 2\int_M^\infty A\omega^{2c}\exp(-2B\omega^\gamma)d\omega \\ &= 2M + 2^{1-\frac{2c+1}{\gamma}}\frac{A}{\gamma}B^{-\frac{2c+1}{\gamma}}\int_{2BM^\gamma}^\infty \omega^{\frac{2c+1}{\gamma}-1}\exp(-\omega)d\omega \\ &= 2M + 2^{1-\frac{2c+1}{\gamma}}\frac{A}{\gamma}B^{-\frac{2c+1}{\gamma}}\Gamma\Big(\frac{2c+1}{\gamma}, 2BM^\gamma\Big) < \infty\end{aligned}$$

where $\Gamma(s,x) \equiv \int_x^\infty t^{s-1}\exp(-t)dt$ is the incomplete gamma function which is well-defined for all $s, x \geq 0$.

A.2.7 Proof of Proposition 1.3

Given that the Condition 1.4 implies that $\varphi_0(\omega|z)$ is non-vanishing almost everywhere and Conditions 1.1, 1.2, and 1.3 are satisfied, we have $\varphi_\Delta(\omega|z) = \varphi_1(\omega|z)/\varphi_0(\omega|z)$ uniquely identified by Lemma 1.1. Then the formula (1.11) follows from Gil-Pelaez (1951) and its modification by Dattner et al. (2011).

A.2.8 Proof of Proposition 1.4

For simplicity, I suppress covariate X in notation. Consider transformed variables $U = F_1^{-1}(Y(1))$ and $V = F_0^{-1}(Y(0))$, both distributed as Uniform$[0,1]$. Note that as they are rank-preserving transformations, joint distribution of (U,V) is equal to the copula of $(Y(1),Y(0))$ (Nelsen, 2007). Let $C^{RI}(u,v)$ be the bivariate CDF of (U,V). From the rank invariance property, we have

$$\frac{C^{RI}(u,v)}{v} = \frac{\Pr(U \leq u, V \leq v)}{\Pr(V \leq v)} = \Pr(U \leq u | V \leq v) = \begin{cases} \frac{u}{v} & \text{if } u < v \\ 1 & \text{if } u = v \\ 1 & \text{if } u > v \end{cases}$$

$$\frac{C^{RI}(u,v)}{u} = \frac{\Pr(U \leq u, V \leq v)}{\Pr(U \leq u)} = \Pr(V \leq v | U \leq u) = \begin{cases} 1 & \text{if } u < v \\ 1 & \text{if } u = v \\ \frac{v}{u} & \text{if } u > v \end{cases}$$

Thus, we have $C^{RI}(u,v) = \min\{u,v\}$.

A.2.9 Proof of Theorem 1.1

The B-spline series of order $r \geq 2$ is defined as $P^r(z) = (1, z, \max\{z-b_1, 0\}, \max\{z-b_2, 0\}, \cdots, \max\{z-b_{r-2}, 0\})'$ for $z \in [0,1]$ where $\{b_1, b_2, \cdots, b_{r-2}\}$ are equally-spaced nodes in $[0,1]$. Denote $\widehat{Z}_i = \widehat{p}_n(X_i)$ and $Z_i = p(X_i)$. Suppose that Lemma 1.6 holds with $\eta = 1$. Then one can show that the following lemma holds:

LEMMA A.2.1. *For $r = r_n$, $E[\|P^r(\widehat{Z}_i) - P^r(Z_i)\|] = O\Big(\sqrt{\frac{\kappa_n}{n}} z_\kappa r_n\Big)$*

Proof. Note that for any $z, z' \in [\underline{p}, \overline{p}]$, we have

$$\begin{aligned}
\|P^r(z') - P^r(z)\| &= \|(0, z'-z, \max\{z'-b_1, 0\} - \max\{z-b_1, 0\}, \\
&\qquad \cdots, \max\{z'-b_{r-2}, 0\} - \max\{z-b_{r-2}, 0\})'\| \\
&= \|z'-z\| + \sum_{j=1}^{r-2} \max\{\|z'-z\|, \|z'-b_j\|, \|z-b_j\|\} \\
&\leq \|z'-z\| + (r-2) \sup_{z \in [\underline{p}, \overline{p}]} \|z\| \|z'-z\| \\
&= (1 + (r-2)\overline{p}) \|z'-z\|
\end{aligned}$$

Hence, there exists a finite constant $C \geq 1$, such that $\|P^r(\widehat{Z}_i) - P^r(Z_i)\| \leq Cr\|\widehat{Z}_i - Z_i\|$. Then from the result of Lemma 1.6, we get

$$E[\|P^r(\widehat{Z}_i) - P^r(Z_i)\|] \leq Cr_n E[\|\widehat{Z}_i - Z_i\|] = O\Big(\sqrt{\frac{\kappa_n}{n}} z_\kappa r_n\Big)$$

$\square$

Define matrices as follow: $M_r \equiv E[P^r(Z_i)P^r(Z_i)']$, $\overline{M}_r \equiv (1/n)\sum_{i=1}^n P^r(Z_i)P^r(Z_i)'$, and $\widehat{M}_r \equiv (1/n)\sum_{i=1}^n P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)'$. Note that M_r is bounded by

$$\|M_r\| \leq \sup_{z \in [0,1]} \|P^r(z)P^r(z)'\| = O(r_n).$$

Following results show that both of the stochastic sequences $\overline{M}_r$ and $\widehat{M}_r$ converge to M_r and bounded.

LEMMA A.2.2. *For $r = r_n$, we have the following results:*

(i) $E[\|\overline{M}_r - M_r\|] = O\Big(\sqrt{\frac{r_n^2 \log r_n}{n}}\Big)$

(ii) $E[\|\widehat{M}_r - M_r\|] = O\Big(\sqrt{\frac{\kappa_n}{n}} z_\kappa r_n^{\frac{3}{2}}\Big) + O\Big(\sqrt{\frac{r_n^2 \log r_n}{n}}\Big)$

Proof of Lemma A.2.2-(i). Note that Z_i for $i = 1, \cdots, n$ are series of independent random variables on a compact support $[\underline{p}, \overline{p}]$. Hence, $P^r(Z_i)P^r(Z_i)'$ is a sequence of independent, symmetric, non-negative random matrix of size $r \times r$ for $r \geq 2$. In addition, by the property of B-spline series, we have $\sup_{z \in [\underline{p}, \overline{p}]} \|P^r(z)\| = O(\sqrt{r})$ as the upper bound. Therefore, $\|E[P^r(Z_i)P^r(Z_i)']\| \leq \sup_{z \in [\underline{p}, \overline{p}]} \|P^r(z)P^r(z)'\| = \sup_{z \in [\underline{p}, \overline{p}]} \|P^r(z)\|^2 = O(r_n)$. Then the case satisfies the conditions in Lemma 6.2 of Belloni et al. (2015) which gives the following result:

$$E[\|\overline{M}_r - M_r\|] = O\Big(\sqrt{\frac{r_n^2 \log r_n}{n}}\Big)$$

□

Proof of Lemma A.2.2-(ii). From the triangle inequality, we get

$$\|\widehat{M}_r - M_r\| \leq \|\widehat{M}_r - \overline{M}_r\| + \|\overline{M}_r - M_r\|$$

Note that

$$\begin{aligned}\|\widehat{M}_r - \overline{M}_r\| &= \Big\|\frac{1}{n}\sum_{i=1}^{n} P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)' - \frac{1}{n}\sum_{i=1}^{n} P^r(Z_i)P^r(Z_i)'\Big\| \\ &\leq \frac{1}{n}\sum_{i=1}^{n} \|P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)' - P^r(Z_i)P^r(Z_i)'\|\end{aligned}$$

$$\leq \frac{2}{n}\sum_{i=1}^{n}\sup_{z\in[\underline{p},\overline{p}]}\|P^r(z)\|\|P^r(\widehat{Z}_i)-P^r(Z_i)\|$$
$$=\frac{2\sqrt{r_n}}{n}\sum_{i=1}^{n}\|P^r(\widehat{Z}_i)-P^r(Z_i)\|$$

Therefore, we have

$$E[\|\widehat{M}_r-\overline{M}_r\|]\leq 2r_nE[\|P^r(\widehat{Z}_i)-P^r(Z_i)\|]=O\Big(\sqrt{\frac{\kappa_n}{n}}z_\kappa r_n^{\frac{3}{2}}\Big)$$

Then the result is followed by Lemma A.2.2-(i). □

LEMMA A.2.3. *Let* $r=r_n$ *and suppose that* $r_n^2\log r_n/n=o(1)$ *and* $\sqrt{\frac{\kappa_n}{n}}z_\kappa r_n^{\frac{3}{2}}=o(1)$. *Then the following results hold:*

(i) $\|\overline{M}_r-M_r\|=o_p(1)$ *and* $\|\overline{M}_r\|=O_p(r_n)$

(ii) $\|\widehat{M}_r-\overline{M}_r\|=o_p(1)$ *and* $\|\widehat{M}_r-M_r\|=o_p(1)$

Proof of Lemma A.2.3-(i). Let C be a finite constant which satisfies $E[\|\overline{M}_r-M_r\|]\leq C\sqrt{\frac{\kappa_n}{n}}z_\kappa r_n^{\frac{3}{2}}$. For an arbitrary constant $\varepsilon>0$, use Markov inequality to have

$$\text{Prob}(\|\overline{M}_r-M_r\|\geq\varepsilon)\leq\frac{1}{\varepsilon}E[\|\overline{M}_r-M_r\|]\leq\frac{C}{\varepsilon}\sqrt{\frac{r_n^2\log r_n}{n}}=o(1)$$

and the last equality follows from the fact that $r_n^2\log r_n/n=o(1)$. In addition, use triangle inequality to show that $\|\overline{M}_r\|\leq\|M_r\|+\|\overline{M}_r-M_r\|=O_p(r_n)+o_p(1)$. □

Proof of Lemma A.2.3-(ii). From the proof of Lemma A.2.2-(ii), we may find a finite constant C such that $E[\|\widehat{M}_r-\overline{M}_r\|]\leq C\sqrt{\frac{\kappa_n}{n}}z_\kappa r_n^{\frac{3}{2}}$. Then for an arbitrary constant $\varepsilon>0$, Markov inequality yields the following relationship

$$\text{Prob}(\|\widehat{M}_r-\overline{M}_r\|\geq\varepsilon)\leq\frac{1}{\varepsilon}E[\|\overline{M}_r-M_r\|]\leq\frac{C}{\varepsilon}\sqrt{\frac{\kappa_n}{n}}z_\kappa r_n^{\frac{3}{2}}=o(1)$$

which claims the first result. Then the second result follows from the triangular inequality such as

$$\|\widehat{M}_r - M_r\| \le \|\widehat{M}_r - \overline{M}_r\| + \|\overline{M}_r - M_r\| = o_p(1) + o_p(1) = o_p(1)$$

$\square$

I present the proof for φ_1 in here which can be extended to the case of φ_0 as well. Note that

$$\begin{aligned}|\widehat{\varphi}_1(\omega|z) - \varphi_1(\omega|z)| &= \Big((\Re(\widehat{\varphi}_1(\omega|z)) - \Re(\varphi_1(\omega|z)))^2 + (\Im(\widehat{\varphi}_1(\omega|z)) - \Im(\varphi_1(\omega|z))^2\Big)^{\frac{1}{2}} \\ &\le |\Re(\widehat{\varphi}_1(\omega|z)) - \Re(\varphi_1(\omega|z))| \\ &\quad + |\Im(\widehat{\varphi}_1(\omega|z)) - \Im(\varphi_1(\omega|z)| \equiv \Gamma^{re}_{1,n} + \Gamma^{im}_{1,n}\end{aligned}$$

I show that both real and imaginary parts uniformly converge to zero at the same rate. Note that $\exp(\iota\omega Y_i) = \cos(\omega Y_i) + \iota\sin(\omega Y_i)$. Hence, for the real part, we have

$$\begin{aligned}\Gamma^{re}_{1,n} &= \Big\|P^r(z)'\Big(\sum_{i=1}^n P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)'\Big)^{-1}\Big(\sum_{i=1}^n P^r(\widehat{Z}_i)D_i\Re(\exp(\iota\omega Y_i))\Big) - \Re(\varphi_1(\omega|z))\Big\| \\ &\le \|P^r(z)\|\Big\|\Big(\sum_{i=1}^n P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)'\Big)^{-1}\Big(\sum_{i=1}^n P^r(\widehat{Z}_i)D_i\cos(\omega Y_i)\Big) \\ &\quad - E[P^r(Z_i)P^r(Z_i)']^{-1}E[P^r(Z_i)D_i\cos(\omega Y_i)]\Big\| \\ &\quad + \|P^r(z)'\beta^{re}_1(\omega) - \Re(\varphi_1(\omega|z))\|\end{aligned}$$

while the first term is decomposed as follows:

$$\begin{aligned}&\Big\|\Big(\sum_{i=1}^n P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)'\Big)^{-1}\Big(\sum_{i=1}^n P^r(\widehat{Z}_i)D_i\exp(\iota\omega Y_i)\Big) \\ &\qquad - E[P^r(Z_i)P^r(Z_i)']^{-1}E[P^r(Z_i)D_i\cos(\omega Y_i)]\Big\|\end{aligned}$$

$$
\begin{aligned}
&\leq \left\|\Big(\frac{1}{n}\sum_{i=1}^{n} P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)\Big)^{-1}\right\| \left\|\frac{1}{n}\sum_{i=1}^{n}(P^r(\widehat{Z}_i)-P^r(Z_i))D_i\cos(\omega Y_i)\right\| \\
&+\left\|\Big(\frac{1}{n}\sum_{i=1}^{n} P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)'\Big)^{-1} - E[P^r(Z_i)P^r(Z_i)']^{-1}\right\| \left\|\frac{1}{n}\sum_{i=1}^{n} P^r(Z_i)D_i\exp(\iota\omega Y_i)\right\| \\
&+\left\|E[P^r(Z_i)P^r(Z_i)]^{-1}\right\| \left\|\frac{1}{n}\sum_{i=1}^{n} P^r(Z_i)D_i\exp(\iota\omega Y_i) - E[P^r(Z_i)D_i\exp(\iota\omega Y_i)]\right\|
\end{aligned}
$$

Suppose that $\lambda_{\min}(\widehat{M}_r) \leq \lambda_{\min}(M_r)$. For sufficiently large n, we may find $\zeta < \infty$ such that $\lambda_{\min}(M_r) - \lambda_{\min}(\widehat{M}_r) = \zeta$ by Lemma A.2.3-(ii). Then,

$$
\left\|\Big(\frac{1}{n}\sum_{i=1}^{n} P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)\Big)^{-1}\right\| \leq 1/\lambda_{\min}(\widehat{M}_r) = 1/(\lambda_{\min}(M_r) + \zeta) < \infty.
$$

In addition,

$$
\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n}(P^r(\widehat{Z}_i)-P^r(Z_i))D_i\exp(\iota\omega Y_i)\right\| &\leq \frac{1}{n}\sum_{i=1}^{n}\|P^r(\widehat{Z}_i)-P^r(Z_i)\|\|D_i\exp(\iota\omega Y_i)\| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\|P^r(\widehat{Z}_i)-P^r(Z_i)\| \\
&= O_p\Big(\sqrt{\frac{\kappa_n}{n}} z_\kappa r_n\Big)
\end{aligned}
$$

given that $\|D_i\cos(\omega Y_i)\| \leq \|D_i\|\|\cos(\omega Y_i)\| \leq 1$. For the second term, note that $\left\|\Big(\frac{1}{n}\sum_{i=1}^{n} P^r(\widehat{Z}_i)P^r(\widehat{Z}_i)'\Big)^{-1} - E[P^r(Z_i)P^r(Z_i)']^{-1}\right\| = o_p(1)$ by Lemma A.2.2. Since $\sup_{z\in[0,1]}\|P^r(z)\| = O(\sqrt{r})$, we have

$$
\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n} P^r(Z_i)D_i\cos(\omega Y_i)\right\| &\leq \frac{1}{n}\sum_{i=1}^{n}\|P^r(Z_i)\|\|D_i\cos(\omega Y_i)\| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\|P^r(Z_i)\| \leq \sup_{Z_i}\|P^r(Z_i)\| = O(\sqrt{r_n})
\end{aligned}
$$

Given that $\lambda_{\min}(E[P^r(Z_i)P^r(Z_i)']) > 0$, we may bound $\left\|E[P^r(Z_i)P^r(Z_i)]^{-1}\right\|$ with a finite constant $1/\lambda_{\min}(\overline{M}_r)$. In addition,

$$
\|P^r(Z_i)D_i\cos(\omega Y_i)\| = \|P^r(Z_i)\|\|D_i\cos(\omega Y_i)\| \leq \|P^r(Z_i)\| = O(\sqrt{r})
$$

uniformly for all $\omega \in \mathbb{R}$. Then we may apply the uniform law of large numbers to have the following result:

$$\sup_{|\omega|\leq B_n} \left\| \frac{1}{n}\sum_{i=1}^{n} P^r(Z_i)D_i \cos(\omega Y_i) - E[P^r(Z_i)D_i\cos(\omega Y_i)] \right\| = o_p(1)$$

Therefore, we have

$$\begin{aligned}
&\sup_{|\omega|\leq B_n, z\in[\underline{p},\overline{p}]} \|\Re(\widehat{\varphi}_{1,n}(\omega|z)) - \Re(\varphi_1(\omega|z))\| \\
&\qquad \leq C\sqrt{r_n}\Big(O_p\Big(\sqrt{\frac{\kappa_n}{n}} z_\kappa r_n\Big) + O_p(\sqrt{r_n}) + o_p(1)\Big) + O(r_n^{-\alpha_1}B_n) \\
&\qquad = O_p\Big(\sqrt{\frac{\kappa_n}{n}} z_\kappa r_n^{\frac{3}{2}}\Big) + O(r_n^{-\alpha_1}B_n)
\end{aligned}$$

Similarly, it can be shown that the imaginary part of $\widehat{\varphi}_{1,n}(\omega|z)$ converges uniformly to that of $\varphi_1(\omega|z)$ at the same rate.

A.2.10 Proof of Theorem 1.2

LEMMA A.2.4. *Suppose that the Assumptions in Theorem 1.1 hold. Then,*

$$\sup_{|\omega|\leq B_n, z\in[\underline{p},\overline{p}]} \left| \frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)} - 1 \right| = O_p\Big(\frac{\kappa_n}{n} z_\kappa r_n^3\Big) + O_p\Big(r_n^{-2\alpha_0}B_n^2\Big)$$

Proof. Note that

$$\begin{aligned}
\left| \frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)} - 1 \right| &= \|\widehat{\varphi}_{0,n}(\omega|z)\|\|\varphi_0(\omega|z) - \widehat{\varphi}_{0,n}(\omega|z)\| \\
&= \|\widehat{\varphi}_{0,n}(\omega|z) - \varphi_0(\omega|z) + \varphi_0(\omega|z)\|\|\varphi_0(\omega|z) - \widehat{\varphi}_{0,n}(\omega|z)\| \\
&\leq \big(\|\widehat{\varphi}_{0,n}(\omega|z) - \varphi_0(\omega|z)\| + \|\varphi_0(\omega|z)\|\big)\|\widehat{\varphi}_{0,n}(\omega|z) - \varphi_0(\omega|z)\| \\
&\leq \big(\|\widehat{\varphi}_{0,n}(\omega|z) - \varphi_0(\omega|z)\| + 1\big)\|\widehat{\varphi}_{0,n}(\omega|z) - \varphi_0(\omega|z)\|
\end{aligned}$$

then from the result in Theorem 1.1,

$$\sup_{|\omega|\leq B_n, z\in[\underline{p},\overline{p}]} \left| \frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)} - 1 \right| = \Big(O_p\Big(n^{-\frac{1}{2}} z_\kappa \kappa_n^{\frac{1}{2}} r_n^{\frac{3}{2}}\Big) + O_p(r_n^{-\alpha_0}B_n) + 1\Big)$$

$$\times\Big(O_p\Big(n^{-\frac{1}{2}}z_\kappa\kappa_n^{\frac{1}{2}}r_n^{\frac{3}{2}}\Big)+O_p(r_n^{-\alpha_0}B_n)\Big)$$
$$=O_p(n^{-1}z_\kappa^2\kappa_n r_n^3)+O_p(r_n^{-2\alpha_0}B_n^2)$$

□

Note that

$$\begin{aligned}\sup_{\tau\in\mathcal{T}}\|\widehat{f}_{\Delta,n}(\tau|z)-f_\Delta(\tau|z)\| &\le \frac{1}{2\pi}\sup_{\tau\in\mathcal{T}}\int_{\mathbb{R}}\|\exp(-\iota\omega\tau)\|\left|\frac{\widehat{\varphi}_{1,n}(\omega|z)\varphi_\xi(h_n\omega)}{\widehat{\varphi}_{0,n}(\omega|z)}-\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\right|d\omega\\ &\le \frac{1}{2\pi}\int_{\mathbb{R}}\left|\frac{\widehat{\varphi}_{1,n}(\omega|z)\varphi_\xi(h_n\omega)}{\widehat{\varphi}_{0,n}(\omega|z)}-\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\right|d\omega\end{aligned}$$

since $\|\exp(-\iota\omega\tau)\|=1$ for all $\tau,\omega\in\mathbb{R}$. Using the triangular inequality, we may decompose the components as follows:

$$\begin{aligned}&\Big|\frac{\widehat{\varphi}_{1,n}(\omega|z)\varphi_\xi(h_n\omega)}{\widehat{\varphi}_{0,n}(\omega|z)}-\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\Big|\\ &\quad=\Big|\frac{\widehat{\varphi}_{1,n}(\omega|z)\varphi_\xi(h_n\omega)}{\widehat{\varphi}_{0,n}(\omega|z)}-\frac{\varphi_1(\omega|z)\varphi_\xi(h_n\omega)}{\widehat{\varphi}_{0,n}(\omega|z)}\\ &\qquad+\frac{\varphi_1(\omega|z)\varphi_\xi(h_n\omega)}{\widehat{\varphi}_{0,n}(\omega|z)}-\frac{\varphi_1(\omega|z)\varphi_\xi(h_n\omega)}{\varphi_0(\omega|z)}+\frac{\varphi_1(\omega|z)\varphi_\xi(h_n\omega)}{\varphi_0(\omega|z)}-\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\Big|\\ &\quad\le\Big|\frac{\varphi_\xi(h_n\omega)}{\widehat{\varphi}_{0,n}(\omega|z)}\Big|\|\widehat{\varphi}_{1,0}(\omega|z)-\varphi_1(\omega|z)\|\\ &\qquad+\|\varphi_1(\omega|z)\varphi_\xi(h_n\omega)\|\Big|\frac{1}{\widehat{\varphi}_{0,n}(\omega|z)}-\frac{1}{\varphi_0(\omega|z)}\Big|+\Big|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\Big|\|\varphi_\xi(h_n\omega)-1\|\\ &\quad=\|\varphi_\xi(h_n\omega)\|\Big|\frac{1}{\varphi_0(\omega|z)}\Big|\Big|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)}\Big|\|\widehat{\varphi}_{1,n}(\omega|z)-\varphi_1(\omega|z)\|\\ &\qquad+\|\varphi_\xi(h_n\omega)\|\Big|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\Big|\Big|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)}-1\Big|\\ &\qquad+\Big|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\Big|\|\varphi_\xi(h_n\omega)-1\|\\ &\quad\equiv\Phi_{1,n}(\omega|z)+\Phi_{2,n}(\omega|z)+\Phi_{3,n}(\omega|z)\end{aligned}$$

where

$$\Phi_{1,n}(\omega|z)=\|\varphi_\xi(h_n\omega)\|\left|\frac{1}{\varphi_0(\omega|z)}\right|\left|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)}\right||\widehat{\varphi}_{1,n}(\omega|z)-\varphi_1(\omega|z)| \tag{A.2.3}$$

$$\Phi_{2,n}(\omega|z) = \|\varphi_\xi(h_n\omega)\| \left|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\right| \left|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)} - 1\right| \tag{A.2.4}$$

$$\Phi_{3,n}(\omega|z) = \left|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\right| |\varphi_\xi(h_n\omega) - 1| \tag{A.2.5}$$

First, note that $\Phi_{1,n}(\omega|z) = 0$ for $|\omega| \geq 1/h_n$ by Condition 1.8. Then for all $z \in [\underline{p}, \overline{p}]$,

$$\begin{aligned}
\int_{\mathbb{R}} \Phi_{1,n}(\omega|z)d\omega &= \int_{|\omega|\leq 1/h_n} \left|\frac{1}{\varphi_0(\omega|z)}\right| \left|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)}\right| \|\widehat{\varphi}_{1,n}(\omega|z) - \varphi_1(\omega|z)\| d\omega \\
&\leq \sup_{|\omega|\leq 1/h_n} \left|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)}\right| \|\widehat{\varphi}_{1,n}(\omega|z) - \varphi_1(\omega|z)\| \int_{|\omega|\leq 1/h_n} \left|\frac{1}{\varphi_0(\omega|z)}\right| d\omega \\
&\leq \frac{C}{h_n}\Big(1 + O_p\Big(\frac{\kappa_n}{n} z_\kappa^2 r_n^3\Big) + O_p\Big(r_n^{-2\alpha_0} h_n^{-2}\Big)\Big) \\
&\quad \times \Big(O_p\Big(\sqrt{\frac{\kappa_n}{n}} z_\kappa r_n^{\frac{3}{2}}\Big) + O_p\Big(r_n^{-\alpha_1} h_n^{-1}\Big)\Big)\underline{\Upsilon}_0\Big(\frac{1}{h_n}\Big)^{-1} \\
&= \frac{1}{h_n}\Big(O_p\Big((\frac{\kappa_n}{n})^{\frac{3}{2}} z_\kappa^3 r_n^{\frac{9}{2}}\Big) + O_p\Big(r_n^{-(2\alpha_0+\alpha_1)} h_n^{-3}\Big)\Big)\underline{\Upsilon}_0\Big(\frac{1}{h_n}\Big)^{-1}
\end{aligned}$$

followed by Theorem 1.1 and Lemma A.2.4 given that $\kappa_n^{1/2} n^{-1/2} z_\kappa r_n^{(3-4\alpha_0)/2} h_n^{-3} = o(1)$ as implied by Condition 1.9. And the second term yields

$$\begin{aligned}
\int_{\mathbb{R}} \Phi_{2,n}(\omega|z)d\omega &= \int_{|\omega|\leq 1/h_n} \left|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\right| \left|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)} - 1\right| d\omega \\
&\leq \sup_{|\omega|\leq 1/h_n} \left|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)} - 1\right| \int_{|\omega|\leq 1/h_n} \left|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\right| d\omega \\
&\leq C \sup_{|\omega|\leq 1/h_n} \left|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)} - 1\right| \frac{1}{h_n} \frac{\overline{\Upsilon}_1(1/h_n)}{\underline{\Upsilon}_0(1/h_n)} \\
&= \Big(O_p\Big(\frac{\kappa_n}{n} z_\kappa^2 r_n^3\Big) + O_p\Big(r_n^{-2\alpha_0} h_n^{-2}\Big)\Big) \frac{1}{h_n} \frac{\overline{\Upsilon}_1(1/h_n)}{\underline{\Upsilon}_0(1/h_n)}
\end{aligned}$$

which is satisfied for all $z \in [\underline{p}, \overline{p}]$. Finally, note that Condition 1.8 implies $\|\varphi_\xi(h_n\omega) - 1\| = o(|\omega|^m)$ for all $|\omega| \leq 1/h_n$. Therefore,

$$\int_{\mathbb{R}} \Phi_{3,n}(\omega|z)d\omega = \int_{|\omega|\leq 1/h_n} \left|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\right| \|\varphi_\xi(h_n\omega) - 1\| d\omega + \underbrace{\int_{|\omega|>1/h_n} \left|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\right| d\omega}_{=o(1)}$$

$$\leq o(1)\int_{\mathbb{R}}\left|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\right||\omega|^m d\omega + o(1) = o(1)$$

while the result is followed by the result of Proposition 1.2. Therefore, we get $\sup_{\tau\in\mathcal{T}}\|\widehat{f}_{\Delta,n}(\tau|z) - f_\Delta(\tau|z)\| \leq C\Big((\frac{\kappa_n}{n})^{\frac{3}{2}} z_\kappa r_n^{\frac{9}{2}} h_n^{-1}\underline{\Upsilon}_0(1/h_n)^{-1} + r_n^{-(2\alpha_0+\alpha_1)}h_n^{-4}\Big)$ with probability approaching to one.

A.2.11 Proof of Theorem 1.3

Note that for any complex value $c \in \mathbb{C}$, $\Im(c) = (c - \overline{c})/2\iota$. Hence, we have

$$\begin{aligned}
\|\widehat{F}_{\Delta,n}(\tau|z) - F_\Delta(\tau|z)\| &= \Big\|\frac{1}{2} - \frac{1}{2\pi}\int_0^\infty (\iota\omega)^{-1}\Big(\exp(-\iota\omega\tau)\varphi_\xi(h_n\omega)\frac{\widehat{\varphi}_{1,n}(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)} \\
&\quad - \exp(\iota\omega\tau)\varphi_\xi(-h_n\omega)\frac{\widehat{\varphi}_{1,n}(-\omega|z)}{\widehat{\varphi}_{0,n}(-\omega|z)}\Big)d\omega - F_\Delta(\omega|z)\Big\| \\
&\leq \frac{1}{2\pi}\int_0^\infty |\iota\omega|^{-1}\Big\|\exp(-\iota\omega\tau)\Big(\varphi_\xi(h_n\omega)\frac{\widehat{\varphi}_{1,n}(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)} - \frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\Big) \\
&\quad - \exp(\iota\omega\tau)\Big(\varphi_\xi(-h_n\omega)\frac{\widehat{\varphi}_{1,n}(-\omega|z)}{\widehat{\varphi}_{0,n}(-\omega|z)} - \frac{\varphi_1(-\omega|z)}{\varphi_0(-\omega|z)}\Big)\Big\|d\omega \\
&\leq \frac{1}{2\pi}\int_0^\infty |\iota\omega|^{-1}\Big\|\exp(-\iota\omega\tau)\Big(\varphi_\xi(h_n\omega)\frac{\widehat{\varphi}_{1,n}(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)} - \frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\Big)\Big\| \\
&\quad + \Big\|\exp(\iota\omega\tau)\Big(\varphi_\xi(-h_n\omega)\frac{\widehat{\varphi}_{1,n}(-\omega|z)}{\widehat{\varphi}_{0,n}(-\omega|z)} - \frac{\varphi_1(-\omega|z)}{\varphi_0(-\omega|z)}\Big)\Big\|d\omega \\
&\equiv \frac{1}{2\pi}\int_0^\infty |\iota\omega|^{-1}(\Xi_n(\omega|z) + \Xi_n(-\omega|z))d\omega
\end{aligned}$$

where $\Xi_n(\omega|z) = \|\exp(\iota\omega\tau)\Big(\varphi_\xi(-h_n\omega)\frac{\widehat{\varphi}_{1,n}(-\omega|z)}{\widehat{\varphi}_{0,n}(-\omega|z)} - \frac{\varphi_1(-\omega|z)}{\varphi_0(-\omega|z)}\Big)\|$.

Notice that $\Xi_n(\omega|z)$ is decomposed into three terms such as $\Xi_n(\omega|z) \leq \Phi_{1,n}(\omega|z) + \Phi_{2,n}(\omega|z) + \Phi_{3,n}(\omega|z)$ where $\Phi_{1,n}, \Phi_{2,n}, \Phi_{3,n}$ are defined as (A.2.3), (A.2.4), (A.2.5) as shown in the proof of Theorem 1.3. First, using the fact that $\Phi_{1,n}(\omega|z) = 0$ for $|\omega| \geq 1/h_n$, we have

$$\int_0^\infty |\iota\omega|^{-1}\Phi_{1,n}(\omega|z)d\omega = \frac{1}{2}\int_{|\omega|<1/h_n} |\omega|^{-1}\Phi_{1,n}(\omega|z)d\omega$$

$$
\begin{aligned}
&\le \frac{1}{2}\sup_{|\omega|<1/h_n}\Big|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)}\Big||\widehat{\varphi}_{1,n}(\omega|z)-\varphi_1(\omega|z)|\int_{|\omega|<1/h_n}\Big|\frac{1}{\omega\varphi_0(\omega|z)}\Big|d\omega\\
&\qquad \le \frac{1}{h_n}\Big(O_p\Big((\frac{\kappa_n}{n})^{\frac{3}{2}}z_\kappa^3 r_n^{\frac{9}{2}}\Big)+O_p\Big(r_n^{-(2\alpha_0+\alpha_1)}h_n^{-3}\Big)\Big)(\ln h_n)^{-1}\underline{\Upsilon}_0\Big(\frac{1}{h_n}\Big)^{-1}
\end{aligned}
$$

while the result is followed by Theorem 1.1 and Conditions 1.9, 1.10. Second, the integral of the second part of decomposition is bounded as

$$
\begin{aligned}
\int_0^\infty |\iota\omega|^{-1}\Phi_{2,n}(\omega|z|d\omega &= \frac{1}{2}\int_{|\omega|<1/h_n}\Big|\frac{\varphi_1(\omega|z)}{\omega\varphi_0(\omega|z)}\Big|\Big|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)}-1\Big|d\omega\\
&\le \frac{1}{2}\sup_{|\omega|<1/h_n}\Big|\frac{\varphi_0(\omega|z)}{\widehat{\varphi}_{0,n}(\omega|z)}-1\Big|\int_{|\omega|<1/h_n}\Big|\frac{\varphi_1(\omega|z)}{\omega\varphi_0(\omega|z)}\Big|d\omega\\
&= \Big(O_p\Big(\frac{\kappa_n}{n}z_\kappa^2 r_n^3\Big)+O_p\Big(r_n^{-2\alpha_0}h_n^{-2}\Big)\Big)\frac{1}{h_n\ln h_n}\frac{\overline{\Upsilon}_1(1/h_n)}{\underline{\Upsilon}_0(1/h_n)}
\end{aligned}
$$

Lastly, we have

$$
\begin{aligned}
&\int_0^\infty |\omega|^{-1}\Phi_{3,n}(\omega|z)d\omega\\
&\quad = \frac{1}{2}\Big(\int_{|\omega|\le 1/h_n}\Big|\frac{\varphi_1(\omega|z)}{\omega\varphi_0(\omega|z)}\Big|\,\|\varphi_\xi(h_n\omega)-1\|d\omega+\int_{|\omega|>1/h_n}\Big|\frac{\varphi_1(\omega|z)}{\omega\varphi_0(\omega|z)}\Big|\,d\omega\Big)\\
&\quad \le o(1)\int_{\mathbb{R}}\Big|\frac{\varphi_1(\omega|z)}{\varphi_0(\omega|z)}\Big|\,|\omega|^{m-1}d\omega+\frac{1}{h_n}\frac{\overline{\Upsilon}_1(1/h_n)}{\underline{\Upsilon}_0(1/h_n)}=o(1)
\end{aligned}
$$

Therefore, we have

$$
\sup_{\tau\in\mathcal{T}}\|\widehat{F}_{\Delta,n}(\tau|z)-F_\Delta(\tau|z)\|\le C\Big((\frac{\kappa_n}{n})^{\frac{3}{2}}z_\kappa r_n^{\frac{9}{2}}(h_n\ln h_n)^{-1}\underline{\Upsilon}_0(1/h_n)^{-1}+r_n^{-(2\alpha_0+\alpha_1)}h_n^{-4}\Big)
$$

with probability approaching to one.

A.2.12 Proof of Corollary 1.3

Pick an arbitrary small number $\varepsilon>0$. Since $F_\Delta(\tau|z)$ is monotonically increasing and continuous over $\mathcal{T}$, it can be easily shown that $Q_\Delta(u|z)$ is monotonic and continuous as well. Therefore, we may find $\delta_1 > 0$ such that, for all $\delta \leq \delta_1$, $\|Q(u-\delta|z) - Q(u+\delta|z)\| < \varepsilon$ uniformly over $u \in [0,1]$. For $\delta_2 > 0$, there exists $\bar{n}$ such that for all $n \geq \bar{n}$,

$$\sup_{\tau\in\mathcal{T}} \|\widehat{F}_{\Delta,n}(\tau|z) - F_{\Delta}(\tau|z)\| < \delta_2 \tag{A.2.6}$$

with probability approaching to one by the result of Proposition 1.3. Note that, for arbitrary $u \in [0,1]$, and $n \geq \bar{n}$,

$$\begin{aligned}
\widehat{Q}_{\Delta,n}(u|z) &= \inf_{\tau\in\mathcal{T}}\{\tau : \widehat{F}_{\Delta,n}(\tau|z) \geq u\} \\
&\geq \inf_{\tau\in\mathcal{T}}\{\tau : F_{\Delta}(\tau|z) + \delta_2 \geq u\} = \inf_{\tau\in\mathcal{T}}\{\tau : F_{\Delta}(\tau|z) \geq u - \delta_2\} \\
&= Q_{\Delta}(u - \delta_2|z)
\end{aligned}$$

and similarly,

$$\begin{aligned}
\widehat{Q}_{\Delta,n}(u|z) &= \inf_{\tau\in\mathcal{T}}\{\tau : \widehat{F}_{\Delta,n}(\tau|z) \geq u\} \\
&\leq \inf_{\tau\in\mathcal{T}}\{\tau : F_{\Delta}(\tau|z) - \delta_2 \geq u\} = \inf_{\tau\in\mathcal{T}}\{\tau : F_{\Delta}(\tau|z) \geq u + \delta_2\} \\
&= Q_{\Delta}(u + \delta_2|z)
\end{aligned}$$

which implies that $\widehat{Q}_{\Delta,n}(u|z) \in [Q_{\Delta}(u-\delta_2|z), Q_{\Delta}(u+\delta_2|Z)]$. Let $\delta = \min\{\delta_1, \delta_2\}$. Then, $\|\widehat{Q}_{\Delta,n}(u|z) - Q_{\Delta}(u|z)\| \leq \|Q_{\Delta}(u-\delta|z) - Q_{\Delta}(u+\delta|z)\| < \varepsilon$ with probability approaching to one. Since the inequality holds for arbitrary $u \in [0,1]$, the result follows.

A.3 Inference via Bootstrap Method

In this section, I discuss the inferential theory on the estimated functional objects (1.16), (1.17), and (1.18). As the closed-form approximation for asymptotic variances of estimators are now available, I consider using bootstrapped samples to construct the confidence intervals for the estimators.

Recent study by Chernozhukov et al. (2016) present a simple algorithm based on repeated resampling procedure to approximate uniform confidence interval for estimators of the functional object. The algorithm is based on a resampling procedure and done by the following steps:

Step 1 fix B, compute $\widehat{F}_{\Delta}^{B}(\tau) = \{\widehat{F}_{\Delta,n}^{(b)}(\tau)\}_{b=1}^{B}$ with resampled data

Step 2 pointwise robust standard error is computed by

$$s_{\Delta}^{B}(\tau) = \frac{Q(\widehat{F}_{\Delta}^{B}(\tau), .75) - Q(\widehat{F}_{\Delta}^{B}(\tau), .25)}{\Phi^{-}(.75) - \Phi^{-}(.25)}$$

where $Q(\widehat{F}_{\Delta}^{B}(\tau), .)$ is empirical quantile of $\widehat{F}_{\Delta}^{B}(\tau)$ and $\Phi^{-}(\cdot)$ is the inverse of standard normal CDF

Step 3 pick a size α and find corresponding factor c_α by $1-\alpha$ quantile of $\{(\widehat{F}_{\Delta,n}^{(b)}(\tau) - \widehat{F}_{\Delta,n}(\tau))/s_{\Delta}^{B}(\tau)\}_{b=1}^{B}$

Step 4 upper and lower bounds of $100(1-\alpha)\%$ confidence interval are computed by $\tilde{F}_{\Delta,n}(\tau) \pm c_\alpha s(\tau)$

The confidence intervals for the quantile effects are simply constructed by left inverse of the lower and upper bounds of the intervals of $\widehat{F}_{\Delta,n}$.

The confidence interval constructed by the above algorithm is known to be valid in the following sense (Corollary 4, Chernozhukov et al. (2016)). Denote $L(\tau) = \tilde{F}_{\Delta,n}(\tau) - c_\alpha s(\tau)$ and $U(\tau) = \tilde{F}_{\Delta,n}(\tau) + c_\alpha s(\tau)$ be the lower and upper bound, respectively. Then for the interval $[L(\tau), U(\tau)]$ indexed by $\tau \in \mathcal{T}$, $F_\Delta(\tau) \in [L(\tau), U(\tau)]$ with probability at least $1-\alpha$.

A.4 Tables and Figures

Figure A.1: An illustrative example of identified distributions of treatment effect under rank invariance and conditional independence assumptions.

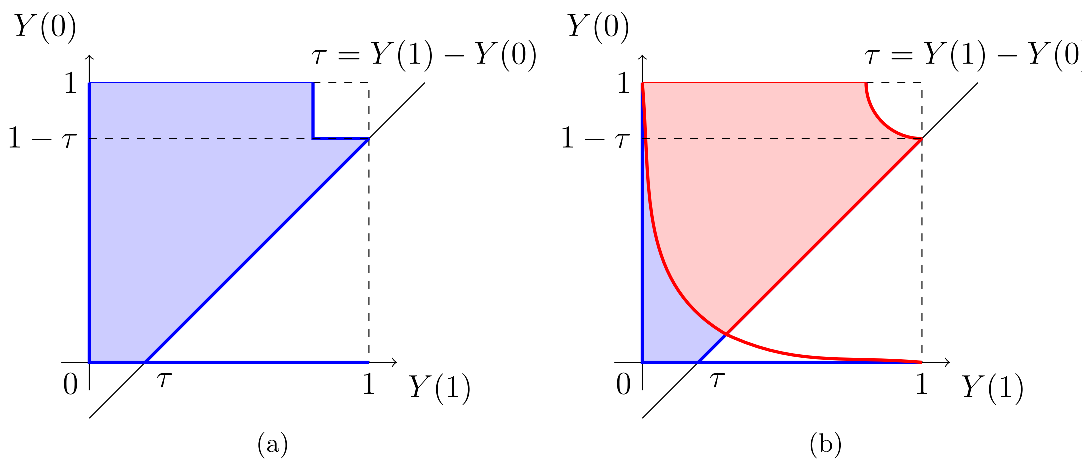


Table A.1: Finite-sample precision of the nonparametric density and distribution estimators measured by mean integrated squared errors. Each number represents the average of mean integrated squared errors among 1000 repetitions.

		$MISE(f)$			$MISE(F)$		
	Sample Size	$\alpha = 0.0$	$\alpha = 0.15$	$\alpha = 0.3$	$\alpha = 0.0$	$\alpha = 0.15$	$\alpha = 0.3$
Normal	$n = 300$	0.0030	0.0033	0.0042	0.0068	0.0061	0.0061
	$n = 500$	0.0029	0.0033	0.0041	0.0065	0.0061	0.0061
	$n = 1000$	0.0027	0.0032	0.0037	0.0062	0.0059	0.0058
	$n = 2000$	0.0026	0.0034	0.0446	0.0061	0.0058	0.0057
Laplace	$n = 300$	0.0012	0.0009	0.0010	0.0209	0.0207	0.0206
	$n = 500$	0.0010	0.0009	0.0009	0.0209	0.0207	0.0207
	$n = 1000$	0.0010	0.0009	0.0008	0.0203	0.0202	0.0204
	$n = 2000$	0.0010	0.0010	0.0199	0.0197	0.0198	0.0199
Exponential	$n = 300$	0.0029	0.0028	0.0028	0.0561	0.0696	0.1031
	$n = 500$	0.0029	0.0028	0.0028	0.0516	0.0676	0.0997
	$n = 1000$	0.0028	0.0027	0.0027	0.0430	0.0629	0.0887
	$n = 2000$	0.0026	0.0026	0.0026	0.0399	0.0610	0.0888

Figure A.2: Estimated distribution effect functions for heterogeneous causal effects when Δ is drawn from the normal distribution. Black solid lines indicate estimates of the distribution function of Δ for each of 1000 iterations. Blue solid line is the true distribution function of which mean is located by blue dashed line.

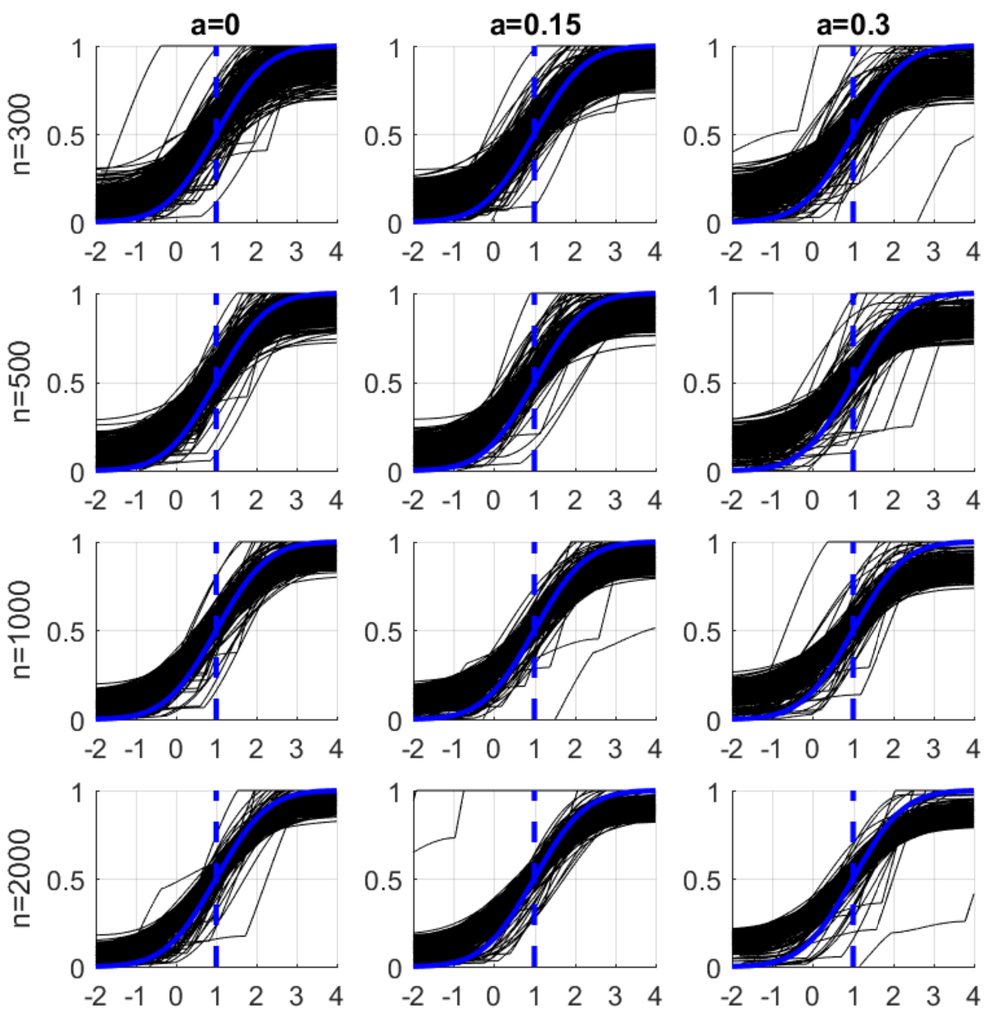


Figure A.3: Estimated quantile effect functions for heterogeneous causal effects when Δ is drawn from the normal distribution. Black solid line is the Monte Carlo average of the quantile effect estimates obtained with 1000 simulations. Red dashed lines show 95% range of the quantile effect estimates at each point. Blue solid line indicates the true quantile function.

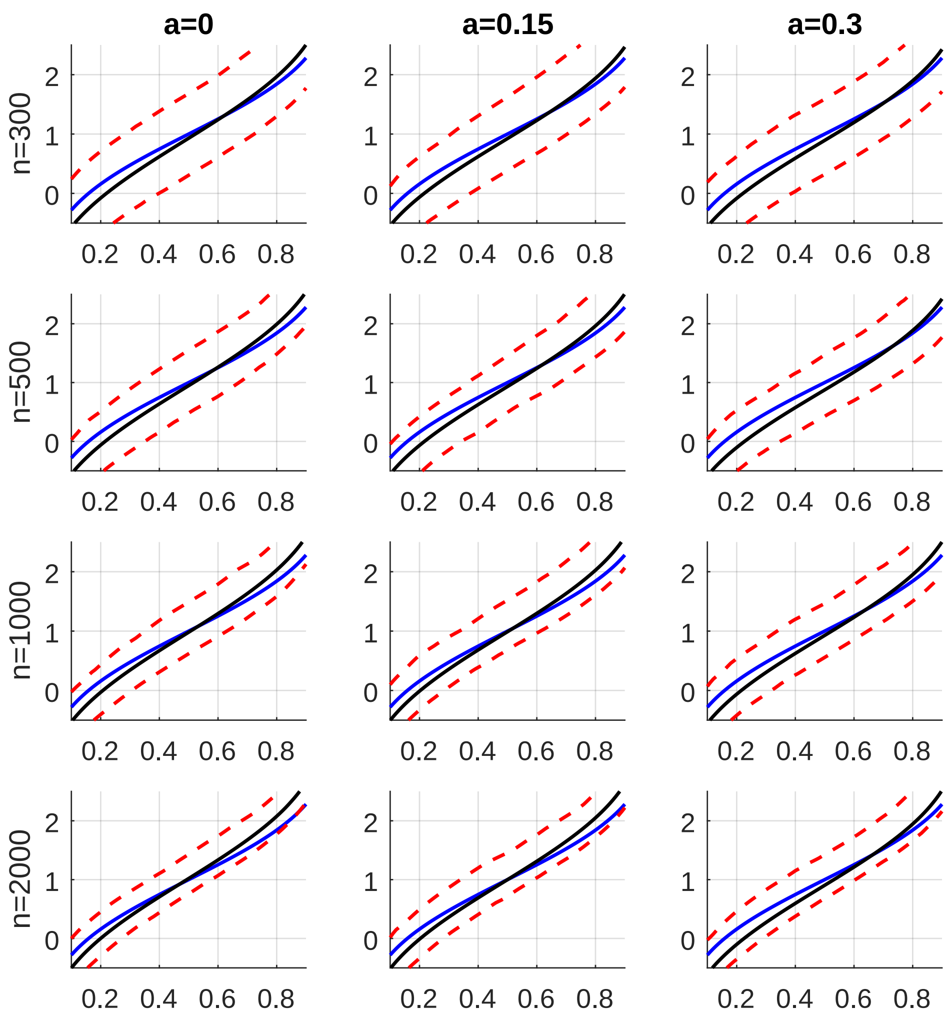


Figure A.4: Estimated quantile effect functions for heterogeneous causal effects when Δ is drawn from the Laplace distribution. Black solid line is the Monte Carlo average of the quantile effect estimates obtained with 1000 simulations. Red dashed lines show 95% range of the quantile effect estimates at each point. Blue solid line indicates the true quantile function.

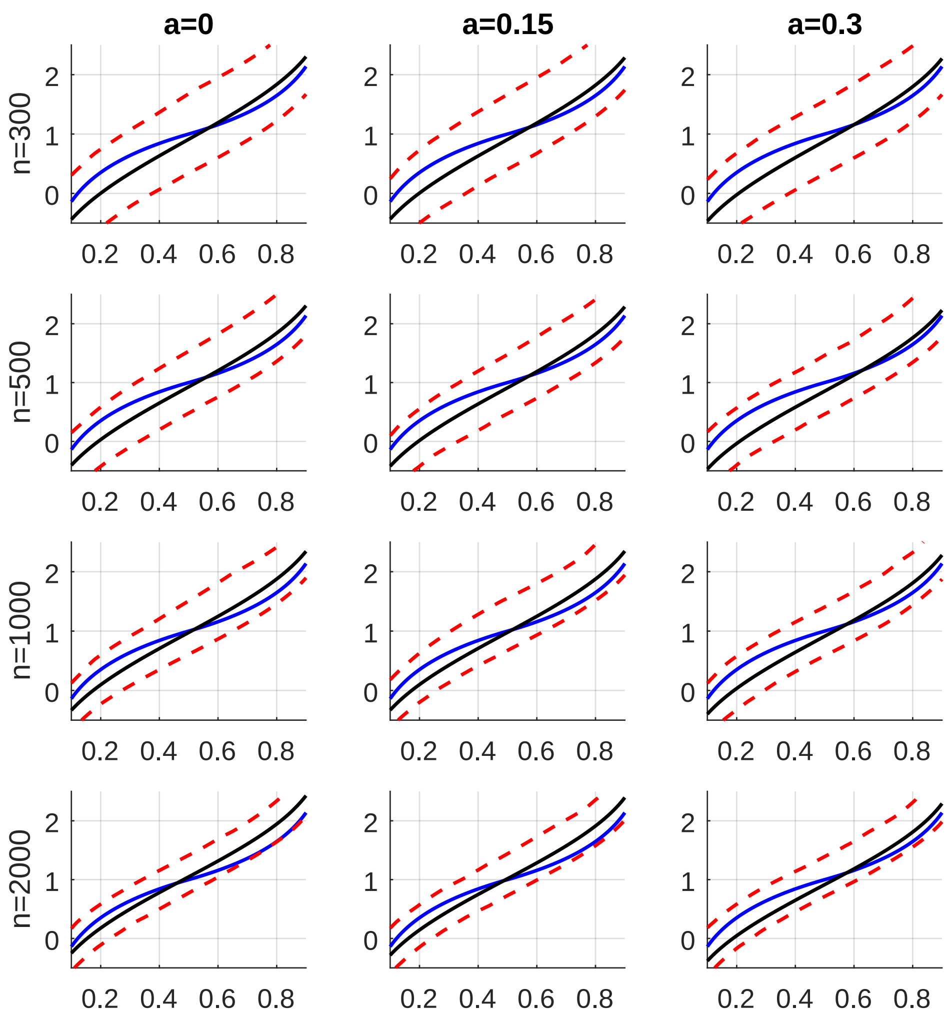
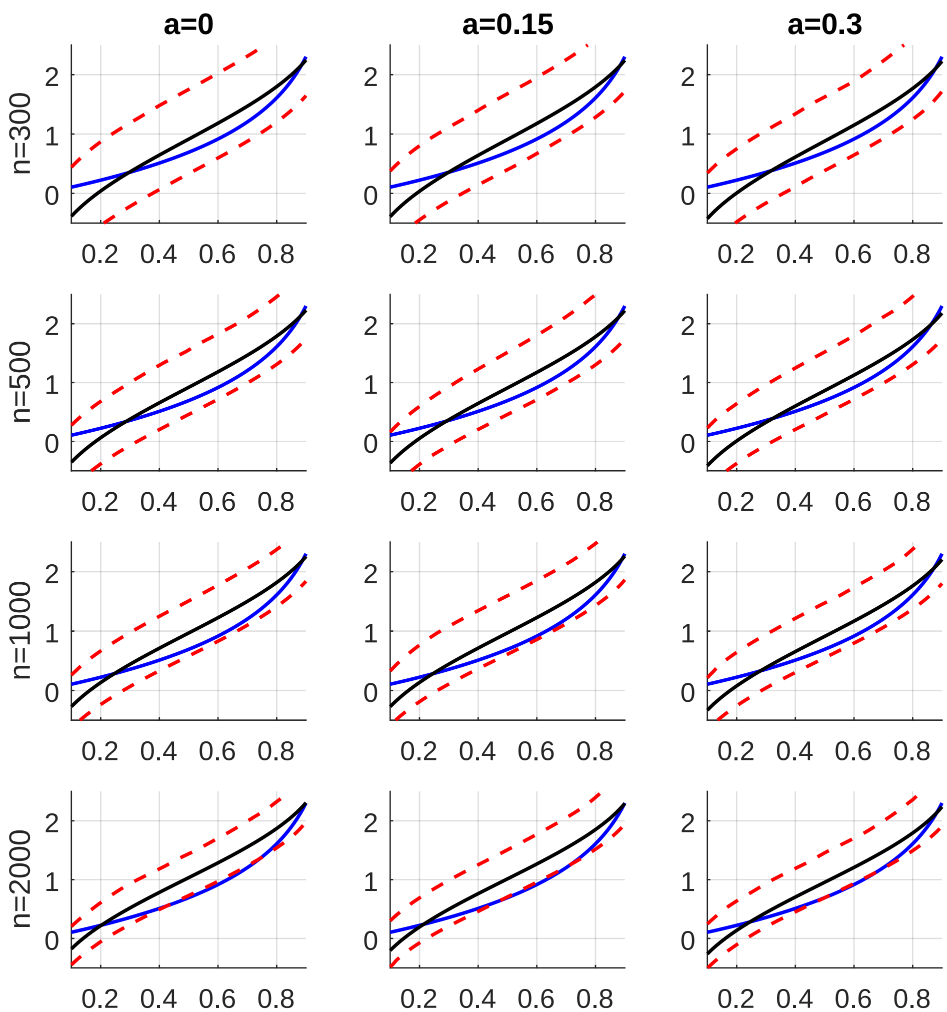


Figure A.5: Estimated quantile effect functions for heterogeneous causal effects when Δ is drawn from the exponential distribution. Black solid line is the Monte Carlo average of the quantile effect estimates obtained with 1000 simulations. Red dashed lines show 95% range of the quantile effect estimates at each point. Blue solid line indicates the true quantile function.



Chapter B: Appendix to Chapter 2

B.1 Proof of Proposition 2.1

I begin with showing that the characteristic functions of potential wages of displaced and non-displaced workers are identified as functions of observed wage, displacement status, and propensity scores. Consider the case of displaced workers ($D = 1$). Note that the observed wage of displaced workers given as (2.3) has deterministic component as a function of phi, a, s and stochastic components ρ, μ, ε. From conditions (C1) and (C3), we have $\ln w_0$ independent of D conditional on ϕ, a, s and log of previous wage denoted by $\ln w_{-1}$. This implies that

$$
\begin{aligned}
E[D\exp(\iota\omega\ln w)|\phi, a, s, \ln w_{-1}] &= E[\exp(\iota\omega\ln w)|D = 1, \phi, a, s, \ln w_{-1}]p(\phi, a, s, \ln w_{-1}) \\
&\quad + 0(1 - p(\phi, a, s, \ln w_{-1})) \\
&= E[\exp(\ln w_1)|D = 1, \phi, a, s, \ln w_{-1}]p(\phi, a, s, \ln w_{-1}) \\
&= E[\exp(\ln w_1)|\phi, a, s, \ln w_{-1}]p(\phi, a, s, \ln w_{-1})
\end{aligned}
$$

Given that the distribution of previous wage $\ln w_{-1}$ is identified via observed wage distribution, we have $\varphi_1(\omega|\phi, a, s)$ identified as conditional mean of

$$
E[D\exp(\iota\omega\ln w)|\phi, a, s, \ln w_{-1}]/p(\phi, a, s, \ln w_{-1})
$$

with respect to ϕ, a, s. Similarly, it can be easily shown that the characteristic function of wage distribution of non-displaced workers is identified via $E[(1 -$

$D) \exp(\iota\omega \ln w)|\phi, a, s, \ln w_{-1}]/(1 - p(\phi, a, s, \ln w_{-1}))$.

Next step is to show that the characteristic function of observed wages of displaced workers is decomposed as a product of the characteristic function of wages of non-displaced workers and Δ. Note that the only stochastic component in Δ is ρ^e as can be seen in (2.5). Then what condition (C2) implies is that Δ is independent to $\ln w_0$ conditional on ϕ, a, and e. Therefore,

$$\begin{aligned}\varphi_1(\omega|\phi, a, s) &= E[\exp(\iota\omega \ln w_1)|\phi, a, s] = E[\exp(\iota\omega(\ln w_0 + \Delta))|\phi, a, s] \\ &= E[\exp(\iota\omega \ln w_0)|\phi, a, s]E[\exp(\iota\omega\Delta)|\phi, a, s] \\ &= \varphi_0(\omega|\phi, a, s)\varphi_\Delta(\omega|\phi, a, s)\end{aligned}$$

where $\Delta = \ln w_1 - \ln w_0$ by definition. Hence, we have the characteristic function of Δ identified as $\varphi_\Delta(\omega|\phi, a, s) = \varphi_1(\omega|\phi, a, s)/\varphi_0(\omega|\phi, a, s)$ given that $\varphi_0(\omega|\phi, a, s) \neq 0$ for all $\omega \in \mathbb{R}$. Finally, the inversion formula shown by Gil-Pelaez (1951) proves that the conditional distribution of Δ is equal to

$$F_\Delta(\tau|\phi, a, s) = \frac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty} (\iota\omega)^{-1} \exp(-\iota\omega\tau)\varphi_\Delta(\omega|\phi, a, s)d\omega$$

for $\tau \in \mathbb{R}$.

B.2 Sensitivity Analysis: Linear Regression

In this section, I estimate the earnings losses via linear regression model. While the nonparametric method developed in this chapter does not require linearity assumption, results presented in this section has two important implications. One is to check robustness of the result discussed in section 2.4. In addition, heterogeneous mean effects estimated by regression models are compared to that of nonparametric estimates to emphasize the benefits of using flexible estimation method.

A simple linear model I consider in this section is written as follow:

$$\ln w_{it} = \alpha_t + \beta D_{it} + \gamma' X_{it} + \varepsilon_{it} \tag{B.2.1}$$

where Y_{it} is the log weekly wage of individual i at year t. The set of covariates include year fixed effect α_t, vector of demographic profiles X_{it}, and a dummy variable D_{it} indicating whether a worker has been displaced or not. The collection of regressors X_{it} includes gender, race, educational attainment. The displacement dummy D_{it} is defined as described in Section 2.4.

In addition, I estimate a more complex version of regression model to capture possible heterogeneity in the effects of displacement on earnings losses. The equation is written as follows:

$$\ln w_{it} = \alpha_t + \sum_{j \in \mathcal{J}} \beta_j \mathbf{1}(T_{it} = j) D_{it} + \gamma' X_{it} + \varepsilon_{it} \tag{B.2.2}$$

where j denotes the group of workers specified by various observable characteristics and $\mathcal{J}$ is the finite set of groups. Variable T_{it} indicates which group that a worker i at period t belongs to. I specifically look at two types of heterogeneity as in section 2.4. One is the educational attainment of workers which is classified into five groups– lower than high school, high school graduates, some college, college graduates, and advanced degree. Another dimension of heterogeneity is local unemployment rates. This is to investigate how earnings losses are different depending on how severe is the labor market faced by job seekers.

Table B.1: Estimates of the effect of displacement on earnings. Dependent variable is weekly wage measured in 2010 dollars. List of covariates are common across all models include gender, race, educational attainments, and year dummies.

	(1)	(2)	(3)	(4)
Displaced	-0.182***	-0.268***	-0.173***	-0.172***
	(0.011)	(0.011)	(0.029)	(0.029)
Tenure	0.025***	0.020***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Tenure^2	-0.071***	-0.049***	-0.007***	-0.006***
	(0.002)	(0.002)	(0.002)	(0.002)
Past Wage		0.906***	0.904***	0.902***
		(0.002)	(0.002)	(0.002)
Fixed Effects				
Region	X	X	O	O
Industry	X	X	O	O
Occupation	X	X	X	O
R^2	0.023	0.841	0.842	0.843
adjusted-R^2	0.022	0.841	0.842	0.842

I begin with presenting estimated coefficients for the simple model (B.2.1) in Table B.1. Estimates of β are shown in the row labeled as "Displaced" with their standard errors in parenthesis. As a benchmark, estimates in column (1) of Table B.1 are produced without controlling for the past wage. By comparing the estimate in column (1) with that of column (2), we may find that the mean effect of displacement is over-estimated in absolute if workers are not matched by their previous wage. However, it is important to point out that the implication from this finding is different from that of Jacobson et al. (1993). Given that the previous wage used in this regression is dated back to one year, it does not reflect the financial distress that the firm may have had before firing its employees. As suggested in Carrington (1993) and Neal (1995), adding industry fixed effect may capture the industry-specific wage trend and therefore, reduce downward bias in displacement effects. The results shown in columns (3) and (4) suggests that in my sample, industry- and occupation-specific effects does not alter the mean estimate of earnings losses significantly.

Besides the over-estimated effect in column (1), mean effect estimates shown in Table B.1 are largely consistent with the nonparametric estimate discussed in section 2.4. However, estimates of the heterogeneous effects through regression model B.2.2 suggest that we may find richer interpretation by using flexible nonparametric methods.

In Table B.2, I present estimates for heterogeneous earnings losses by displacement as a function of local unemployment rates. Three types of unemployment rates are considered. First is the local unemployment rate which is the fraction of unemployed within the same state and same year. Similarly, industry- and occupation-specific unemployment rates are computed for each industry and occupation, respectively. First two columns for all three cases are estimated by separating the whole sample into two groups. One is when the unemployment rate is above median in the sample period–which is about 5.5%–while the other group is the case when unemployment rates are below median. In general, estimated effect of displacement is larger in absolute value when the unemployment rates are higher. However, the difference is not significant.

B.3 Optimal Rate of Approximation via Cross-validation Method

In this section, I illustrate how to implement the cross-validation method to find the optimal degree of approximation regarding the bias-variance trade-off in the

Table B.2: Estimates of the effect of displacement on earnings by unemployment rates. Dependent variable is weekly wage measured in 2010 dollars. State, industry, and occupation specific unemployment rates are calculated using ASEC sample of each year. List of covariates include gender, race, educational attainments, and year dummies.

	Region			Industry			Occupation		
	(1) high	(2) low	(3)	(4) high	(5) low	(6)	(7) high	(8) low	(9)
Displaced	-0.177***	-0.170***	-0.124	-0.216***	-0.135***	-0.142**	-0.163***	-0.173***	-0.135***
	(0.044)	(0.039)	(0.081)	(0.046)	(0.037)	(0.055)	(0.044)	(0.037)	(0.050)
Displaced x Unemployment Rate			-0.003			-0.011			-0.011*
			(0.012)			(0.007)			(0.006)
Tenure	0.004***	0.005***	0.005***	0.002**	0.007***	0.004***	0.004***	0.004***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tenure^2	-0.010***	-0.011***	-0.010***	-0.004**	-0.015***	-0.008***	-0.010***	-0.008***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
Past Wage	0.909***	0.903***	0.906***	0.939***	0.834***	0.905***	0.928***	0.823***	0.905***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)	(0.006)	(0.002)
Constant	0.090***	0.003	0.010	0.003	0.004	0.026	0.018	0.006	0.034*
	(0.028)	(0.035)	(0.019)	(0.024)	(0.032)	(0.019)	(0.027)	(0.031)	(0.019)
Observations	76829	84120	160949	96310	64639	160949	104899	56050	160651
R^2	0.845	0.838	0.841	0.884	0.751	0.841	0.866	0.746	0.842
adjusted-R^2	0.845	0.838	0.841	0.884	0.751	0.841	0.866	0.746	0.841

nonparametrically estimated distribution of heterogeneous effects. In the context of the model provided in this chapter, the problem is reduced down to the choice of r, a non-negative integer representing the degree of approximation pre-determined to compute the empirical characteristic function (2.7). I find the optimal choice of r that minimizes the estimation bias by testing candidate values of r.

Cross-validation method is a statistical learning algorithm that has been widely used for model evaluation. Suppose that we have n number of observations. The method begins with separating the sample into two groups. One is called "training set" which is used to estimate the parameter of interest. In this chapter, the objective is the distribution of causal effects denoted by F_{Δ}. Another set of samples is called "validation set." These observations are used to compute a measurement of estimation error which is chosen by the researcher. A typical choice would be the squared sum of errors, however, I consider mean integrated squared error as it is a more intuitive measurement to evaluate the bias when the objective of interest is the distribution function rather than a scalar-valued parameter.

Training and validation sets are chosen by blocking method. The entire sample is randomly separated into K subgroups of equal size. Pick k-th subgroup among $1, \cdots, K$ as the validation set while the rest of the sample is used for training set. Let $\widehat{F}_{\Delta,r}^{train}$ be the estimated distribution of heterogeneous effects Δ_{it} using the training set which is computed under a pre-determined value r. This distribution is treated as if it is a population quantity. Meanwhile, estimate the distribution of Δ_{it} using validation set as well with the same value r fixed, and denote by $\widehat{F}_{\Delta,r}^{validation}$. Then the mean integrated squared error is computes by

$$MISE(r) = E[(\widehat{F}_{\Delta,r}^{validation}(\Delta_{it}) - \widehat{F}_{\Delta,r}^{train}(\Delta_{it}))^2]$$

while the expectation is approximated by sample mean over the density estimated using the training set.

Tuning parameter in this context is the degree of approximation r. I compute the mean integrated squared errors at non-zero integer values of r to find the value that minimizes estimation bias.

B.4 Tables and Figures

Figure B.1: Time series plot of average weekly earnings of workers by their displacement status. Blue dashed line indicates the series of average log wages of displaced workers within each year and red solid line indicates that of non-displaced workers. Individual wages are based on reported value in Monthly CPS, normalized to 2010 dollar values. Vertical lines at each point shows ± 1 standard deviation range of within-year averages.

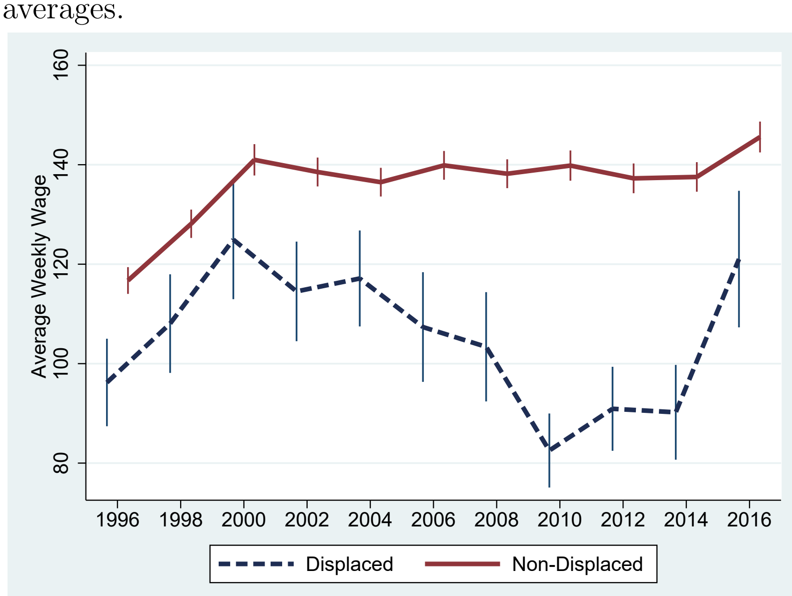


Figure B.2: Distribution of propensity score estimates by treatment groups. Red line indicates the distribution of estimated propensity scores of the group of displaced workers ($D = 1$) while blue line indicates that of the non-displaced workers ($D = 0$).

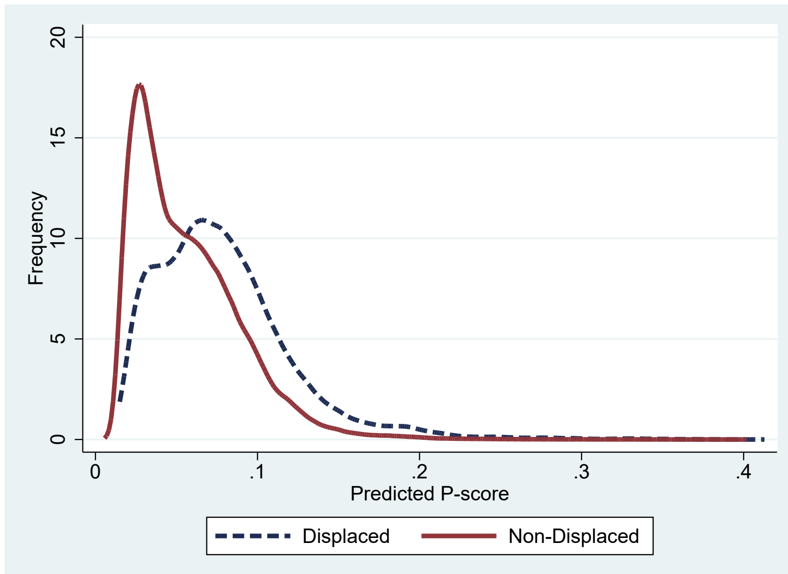


Figure B.3: Estimated distribution of the heterogeneous effect of displacement on earnings. Scale of the horizontal axis is the log difference in weekly earnings. Blue solid line represents the point estimates of the distribution function of causal effect while red dashed lines show 95% confidence interval of point estimates.

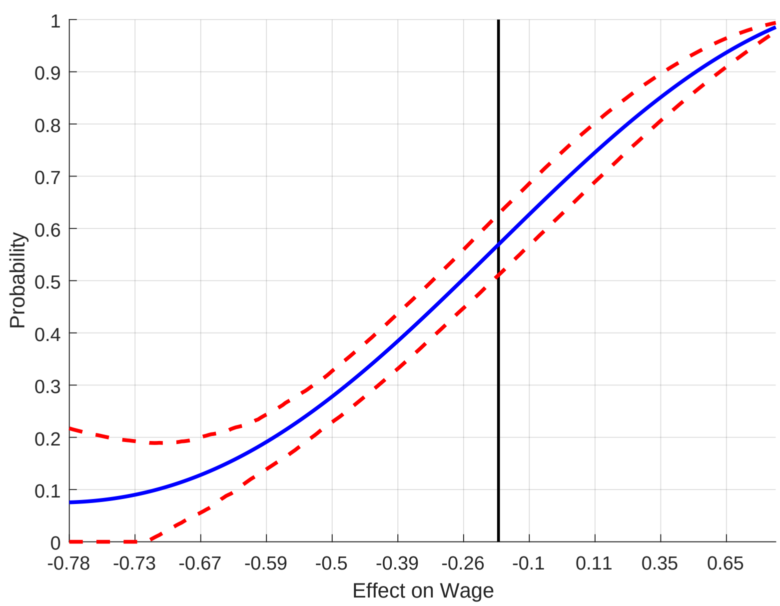


Table B.3: Summary of the labor market status of displaced workers by year within the sample period. Numbers in each column implies fraction of previously displaced workers who ended up being either employed, unemployed, or out of the labor force. Employed workers include both full-time and part-time positions. Column NILF includes workers who are currently not in the labor force for either being retired or other reasons.

Year	Employed	Unemployed	NILF
1996	71.31	16.84	11.85
1998	75.64	12.92	11.44
2000	74.61	13.05	12.34
2002	64.32	23.53	12.16
2004	66.67	21.36	11.98
2006	68.83	17.30	13.87
2008	68.56	19.71	11.73
2010	50.88	36.99	12.13
2012	58.45	28.21	13.34
2014	61.80	24.77	13.42
2016	66.90	18.78	14.32
Total	64.76	22.71	12.53

Table B.4: Summary of the labor market status of displaced workers by education level. Numbers in each column implies fraction of previously displaced workers who ended up being either employed, unemployed, or out of the labor force. Employed workers include both full-time and part-time positions. Column NILF includes workers who are currently not in the labor force for either being retired or other reasons.

Education	Employed	Unemployed	NILF
less than high school	51.76	31.66	16.58
high school	60.93	25.48	13.59
some college	64.82	21.59	13.60
college	71.81	18.34	9.85
advanced	75.08	16.32	8.60

Table B.5: Summary of the labor market status of displaced workers by age. Numbers in each column implies fraction of previously displaced workers who ended up being either employed, unemployed, or out of the labor force. Employed workers include both full-time and part-time positions. Column NILF includes workers who are currently not in the labor force for either being retired or other reasons.

Age	Employed	Unemployed	NILF
less than 25	63.34	23.76	12.90
25–35	68.77	21.03	10.20
35–45	68.31	21.96	9.73
45–55	64.19	24.15	11.66
more than 55	51.76	24.03	24.21

Table B.6: Demographic characteristics of workers in the sample for both displaced and non-displaced workers. Numbers shown are group-specific means while their standard errors are given in paranthesis below.

	Displaced	Non-displaced	Total		Displaced	Non-displaced	Total
Female	0.4326	0.5254	0.5200	Lower than High School	0.1116	0.1114	0.1114
	(0.0023)	(0.0006)	(0.0006)		(0.0015)	(0.0004)	(0.0004)
White	0.7164	0.7289	0.7282	High School	0.3251	0.3106	0.3114
	(0.0021)	(0.0005)	(0.0005)		(0.0022)	(0.0005)	(0.0005)
Black	0.1051	0.0987	0.0990	Some College	0.2131	0.1887	0.1901
	(0.0014)	(0.0003)	(0.0003)		(0.0019)	(0.0005)	(0.0004)
Hispanic	0.1241	0.1083	0.1092	College	0.2837	0.2905	0.2901
	(0.0015)	(0.0004)	(0.0004)		(0.0021)	(0.0005)	(0.0005)
Other	0.0543	0.0642	0.0636	Advanced	0.0665	0.0988	0.0969
	(0.0011)	(0.0003)	(0.0003)		(0.0012)	(0.0003)	(0.0003)
Age	40.7456	43.0089	42.8774	Weekly Wage	103.21	136.59	134.65
	(0.0549)	(0.0143)	(0.0139)		(1.52)	(0.45)	(0.44)
Tenure	7.26	23.86	22.89	Weekly Wage (past)	131.03	111.81	112.93
	(0.06)	(0.02)	(0.02)		(1.21)	(0.22)	(0.22)
Observations	45270	733784	779054				

Table B.7: Specification test for the propensity score to displacement. Numbers indicate the significance levels (p-values) of each variable tested against the displacement dummy after controlling for the estimated propensity score functions. Columns indicate different propensity score specifications from the smallest (1) to the largest (5).

	(1)	(2)	(3)	(4)	(5)
Worker Characteristics					
Female	0.893	0.000	0.000	0.774	0.893
High School	0.959	0.053	0.045	0.928	0.928
Some College	0.943	0.056	0.048	0.970	0.958
College	0.796	0.005	0.003	0.972	0.803
Advanced	0.727	0.002	0.001	0.996	0.728
Black	0.855	0.604	0.636	0.912	0.854
Hispanic	0.804	0.530	0.831	0.852	0.770
Other	0.908	0.450	0.475	0.929	0.912
Unemployment Rates					
by Region	0.960	0.000	0.651	0.963	0.973
by Industry	0.000	0.000	0.000	0.036	0.000
by Occupation	0.000	0.008	0.008	0.036	0.000
Other Variables					
Year Dummies	1.000	0.274	1.000	1.000	1.000
Tenure	0.995	0.000	0.000	0.989	0.994
Log Likelihood	4.53e+05	4.09e+05	4.09e+05	4.09e+05	4.09e+05
Observations	779054	779054	779054	779054	779054

Table B.8: Estimated quantile effects of the heterogeneous effect of displacement on earnings. Estimates are the mean and quantile effects of displacement on wage losses. Standard errors obtained by bootstrap with 1000 repetitions are given in parentheses.

	Quantile Effects				
Mean Effect	.10	.25	.50	.75	.90
-0.1938	-0.7086	-0.5321	-0.2635	0.1137	0.2201
(0.0083)	(0.0270)	(0.0056)	(0.0045)	(0.0050)	(0.0050)

Table B.9: Estimated quantile of the heterogeneous effect of displacement on earnings by groups, classified by pre-displacement tenure. Tenure for displaced workers is obtained from DWS supplement and tenure for non-displaced workers is from JT supplement and subtracted by 1 to get the tenure in previous year. Standard errors are in parentheses.

		Quantile Effects				
	Mean Effect	.10	.25	.50	.75	.90
$<$ 1 Year	-0.1589	-0.7475	-0.5236	-0.1755	0.3166	0.4439
	(0.0005)	(0.0023)	(0.0007)	(0.0007)	(0.0004)	(0.0004)
1-5 Years	-0.1843	-0.7575	-0.5516	-0.2250	0.2523	0.3789
	(0.0003)	(0.0010)	(0.0003)	(0.0003)	(0.0003)	(0.0002)
5-10 Years	-0.1602	-0.7371	-0.5286	-0.1847	0.3090	0.4372
	(0.0004)	(0.0008)	(0.0005)	(0.0004)	(0.0003)	(0.0003)
$>$ 10 Years	-0.0564	-0.6540	-0.3960	0.0235	0.5644	0.6951
	(0.0007)	(0.0012)	(0.0011)	(0.0009)	(0.0004)	(0.0004)

Table B.10: Estimated quantile of the heterogeneous effect of displacement on earnings by groups, classified by educational attainments. Standard errors are in parentheses.

		Quantile Effects				
	Mean Effect	.10	.25	.50	.75	.90
$<$ High School	-0.1700	-0.6945	-0.5223	-0.2346	0.1762	0.2908
	(0.0021)	(0.0046)	(0.0011)	(0.0008)	(0.0008)	(0.0008)
High School	-0.1832	-0.7011	-0.5282	-0.2516	0.1405	0.2507
	(0.0015)	(0.0033)	(0.0007)	(0.0007)	(0.0007)	(0.0008)
Some College	-0.1878	-0.7048	-0.5348	-0.2597	0.1321	0.2424
	(0.0014)	(0.0031)	(0.0008)	(0.0009)	(0.0008)	(0.0009)
College	-0.1859	-0.7040	-0.5291	-0.2536	0.1358	0.2453
	(0.0012)	(0.0035)	(0.0008)	(0.0006)	(0.0007)	(0.0007)
Advanced	-0.1826	-0.7056	-0.5277	-0.2464	0.1512	0.2626
	(0.0015)	(0.0037)	(0.0019)	(0.0012)	(0.0009)	(0.0010)

Table B.11: Estimated quantile of the heterogeneous effect of displacement on earnings by groups classified by year, local and industry-specific unemployment rates. State-level unemployment rates are obtained by Local Area Unemployment Statistics. Industyr- and occupation-specific unemployment rates are computed from ASEC sample of each year. Standard errors are in parentheses.

		Quantile Effects				
	Mean Effect	.10	.25	.50	.75	.90
by Year						
before 2008	-0.2088	-0.7499	-0.4947	-0.2278	-0.0008	0.1343
	(0.0075)	(0.0049)	(0.0034)	(0.0039)	(0.0055)	(0.0062)
after 2008	-0.2212	-0.7356	-0.4758	-0.2054	0.0273	0.1681
	(0.0072)	(0.0042)	(0.0031)	(0.0037)	(0.0050)	(0.0062)
by Region						
high unemployment rate	-0.2546	-0.7596	-0.4911	-0.2201	0.0109	0.1497
	(0.0076)	(0.0049)	(0.0029)	(0.0035)	(0.0046)	(0.0056)
low unemployment rate	-0.1786	-0.7353	-0.4822	-0.2148	0.0144	0.1518
	(0.0067)	(0.0051)	(0.0035)	(0.0041)	(0.0054)	(0.0063)
by Industry						
high unemployment rate	-0.3355	-0.8845	-0.5670	-0.2908	-0.0651	0.0657
	(0.0094)	(0.0087)	(0.0041)	(0.0037)	(0.0050)	(0.0059)
low unemployment rate	-0.1375	-0.6774	-0.4373	-0.1685	0.0689	0.2153
	(0.0058)	(0.0035)	(0.0032)	(0.0040)	(0.0057)	(0.0071)

Chapter C: Appendix to Chapter 3

C.1 Proofs

C.1.1 Definitions, Lemmas and Preliminary Results

For a matrix A, $\|A\|^2 = \operatorname{tr} AA'$ is the Euclidian matrix norm. When a is a vector $\|a\|^2 = a'a$. The matrix norm $\|A\|_2^2$ is given by $\|A\|_2^2 = \sup_{x\neq 0} x'A'Ax/x'x$. Define the norm $|f|_\infty = \sup_\chi |f(\chi)|$ for any function $f : \mathcal{X} \to \mathbb{R}$. Similarly define $\|f\|_\infty = \sup_\chi \|f(\chi)\|$ for any function $f : \mathcal{X} \to \mathbb{R}^d$.The following definitions are given in Rio (1993). For a nonincreasing function $h : \mathbb{R} \to \mathbb{R}$ define the inverse $h^{-1}(u) = \inf\{t : h(t) \leq u\}$. Let $\mathfrak{Q}_f(u)$ be the quantile function defined as the inverse of the tail probability $P(|f(\chi_t)| > t)$. Let $\lfloor t \rfloor$ be the largest integer smaller or equal to $t \in \mathbb{R}$ and define $\beta^{-1}(u) = \inf\left\{t : \beta_{\lfloor t \rfloor} \leq u\right\}$ where $\beta_{\lfloor t \rfloor}$ is the mixing coefficient defined in (3.11). Let $f(\chi_t)$ be a measurable function of χ_t and define

$$\|f\|_{2,\beta} = \sqrt{\int_0^1 \beta^{-1}(u) \left[\mathfrak{Q}_f(u)\right]^2 du} < \infty.$$

Lemma C.1.1. *Let $f(\chi_t)$ and $g(\chi_t)$ be measurable functions of χ_t where χ_t satisfies Condition 3.2. Let $F(\chi_t)$ be an envelope such that $|f(\chi_t)| \leq F(\chi_t)$ and $|g(\chi_t)| \leq F(\chi_t)$ and $E\left[|F(\chi_t)|^{2+\delta}\right] < \infty$. Then, $\sum_{s=-\infty}^{\infty} \operatorname{Cov}(f(\chi_t), g(\chi_t)) \leq 4\|F\|_{2,\beta}^2$.*

Proof. The proof follows the argument in Rio (1993). In particular,

$$|\operatorname{Cov}(f(\chi_t), g(\chi_{t+m}))| \leq 2\int_0^{\beta_m/2} \left(\left[\mathfrak{Q}_f(2u)\right]^2 + \left[\mathfrak{Q}_g(2u)\right]^2\right) du$$

by Rio (1993, p. 594), see also Doukhan, Massart, and Rio (1994, Proposition 1).

Now use that $\mathfrak{Q}_f(u) \leq \mathfrak{Q}_F(u)$ and $\mathfrak{Q}_g(u) \leq \mathfrak{Q}_F(u)$ such that

$$\sum_{m=-\infty}^{\infty} |\mathrm{Cov}(f(\chi_t), g(\chi_{t+m}))| \leq 4\int_0^1 \beta^{-1}(u)[\mathfrak{Q}_F(u)]^2 du = 4\|F\|_{2,\beta}^2 .$$

□

Lemma C.1.2. *For any f with $\|f\|_{2+\delta,P} < \infty$ and if $\sum_{n>0} n^{1/\delta}\beta_n < \infty$ it follows that $\|f\|_{2,\beta} \leq C\|f\|_{2+\delta,P}$ for some $C < \infty$.*

Proof. Use Lemma 2 of Doukhan et al. (1994) to bound $\|f\|_{2,\beta}$. In particular, set $\phi = x^{1+\delta/2}$ and define

$$\|f\|_{\phi,2} = \inf\left\{c > 0 : E\left(\phi\left((f/c)^2\right)\right) \leq 1\right\}$$
$$\Lambda_\phi(f) = \sup_{u\in]0,1]}\left(\left[\phi^{-1}(u)\right]^{-1/2}\mathfrak{Q}_f(u)\right)$$

such that from Doukhan, Massart, and Rio (1995, p. 402-404), it follows that $\Lambda_\phi(f) \leq \|f\|_{\phi,2} = \|f\|_{2+\delta,P}$ and from Doukhan et al. (1995, Lemma 2(a)) that

$$\|f\|_{2,\beta} \leq \|f\|_{2+\delta,P}\sqrt{1+\int_0^1 \phi^*\left(\beta^{-1}(u)\right)du} \tag{C.1.1}$$

where $\phi^*(y) = \sup_{x>0}[xy - \phi(x)]$ is the dual function of ϕ. The integral $\int_0^1 \phi^*\left(\beta^{-1}(u)\right)du$ is bounded if $\sum_{n>0} n^{1/\delta}\beta_n < \infty$ by Doukhan, Massart and Rio (1995, p. 403, S.1). □

Lemma C.1.3. *Assume Condition 3.4 holds. Let $\langle z\rangle = 1 + \|z\|^2$. Then,*
(i)

$$\inf_{\varphi_\kappa\in\mathbb{R}^\kappa}\sup_{z\in\mathbb{R}^d}\left\|\frac{\Gamma^{-1}(p(z,\varphi)) - \Psi_\kappa(z,\varphi_\kappa)}{\langle z\rangle^{s(2+\delta)/2}}\right\| = o\left(\kappa^{-\alpha}\right),$$

where $\alpha = ...$
(ii) if $E\left[\langle z\rangle^{s(2+\delta)}\right] < \infty$ then

$$\inf_{\varphi_\kappa\in\mathbb{R}^\kappa}\left\|\Gamma^{-1}(p(z,\varphi)) - \Psi_\kappa(z,\varphi_\kappa)\right\|_{2,P} = o\left(\kappa^{-\alpha}\right).$$

Proof. (i) Hansen (2015, Theorem 7) shows the result when $\dim(z) = 1$. Note that because $p^j > \underline{p}$ by Condition 3.3(i). This implies that $\Gamma^{-1}(p(z,\varphi)) < \bar{g}(z) = \Gamma^{-1}\left(1/\underline{p}\right) = K$ is bounded. Hansen uses $w(z) = \bar{g}(z)|z|^{s(2+\delta)}$. It follows that up to an irrelevant constant $\langle z\rangle^{s(2+\delta)/2} \geq w(z)$ and thus

$$\sup_{z\in\mathbb{R}^d}\left|\frac{\Gamma^{-1}(p(z,\varphi)) - \Psi_\kappa(z_t,\varphi_\kappa)}{\langle z\rangle^{s(2+\delta)/2}}\right| \leq \sup_{z\in\mathbb{R}^d}\left|\frac{\Gamma^{-1}(p(z,\varphi)) - \Psi_\kappa(z_t,\varphi_\kappa)}{w(z)}\right|$$

such that Hansen's proof applies for $\dim(z) = 1$. The result needs to be extended to $\dim(z) > 1$.

For (ii) note that

$$
\begin{aligned}
\left\|\Gamma^{-1}\left(p\left(z,\varphi\right)\right) - \Psi_{\kappa}\left(z_t,\varphi_{\kappa}\right)\right\|_{2,P}^{2} &= \int \left\|\Gamma^{-1}\left(p\left(z,\varphi\right)\right) - \Psi_{\kappa}\left(z_t,\varphi_{\kappa}\right)\right\|^{2} dP\left(z\right) \\
&\leq \int \left\|\frac{\Gamma^{-1}\left(p\left(z,\varphi\right)\right) - \Psi_{\kappa}\left(z_t,\varphi_{\kappa}\right)}{\left\langle z\right\rangle^{s(2+\delta)/2}}\right\|^{2} \left\langle z\right\rangle^{s(2+\delta)} dP\left(z\right) \\
&\leq \left(\sup_{z\in\mathbb{R}^d}\left\|\frac{\Gamma^{-1}\left(p\left(z,\varphi\right)\right) - \Psi_{\kappa}\left(z_t,\varphi_{\kappa}\right)}{\left\langle z\right\rangle^{s(2+\delta)/2}}\right\|\right)^{2} \int \left\langle z\right\rangle^{s(2+\delta)} dP\left(z\right)
\end{aligned}
$$

which implies that

$$
\begin{aligned}
&\inf_{\varphi_{\kappa}\in\mathbb{R}^{\kappa}} \left\|\Gamma^{-1}\left(p\left(z,\varphi\right)\right) - \Psi_{\kappa}\left(z,\varphi_{\kappa}\right)\right\|_{2,P} \\
&\qquad \leq K \inf_{\varphi_{\kappa}\in\mathbb{R}^{\kappa}} \sup_{z\in\mathbb{R}^d} \left\|\frac{\Gamma^{-1}\left(p\left(z,\varphi\right)\right) - \Psi_{\kappa}\left(z,\varphi_{\kappa}\right)}{\left\langle z\right\rangle^{s(2+\delta)/2}}\right\| = o\left(\kappa^{-\alpha}\right).
\end{aligned}
$$

$\square$

LEMMA C.1.4. *Define* $P_{\kappa}(z_t,\varphi) = \mathrm{diag}(p_{\kappa}(z_t,\varphi_{\kappa}))$ *and* $Q_{\kappa}(z_t,\varphi_{\kappa}) = (\mathrm{diag}(p_{\kappa}(z_t,\varphi_{\kappa})) - p_{\kappa}(z_t,\varphi_{\kappa})p_{\kappa}(z_t,\varphi_{\kappa})')$. *Similarly, let*

$$
P\left(z_t,\varphi\right) = \mathrm{diag}\left(p(z_t,\varphi)\right), \qquad Q(z_t,\varphi) = (\mathrm{diag}(p(z_t,\varphi_{\kappa})) - p(z_t,\varphi_{\kappa})p(z_t,\varphi)').
$$

Then, it follows that

(i)

$$
\begin{aligned}
Q_{\kappa}\left(z_t,\varphi_{\kappa}\right)^{-1} &= P_{\kappa}\left(z_t,\varphi_{\kappa}\right)^{-1} + \frac{\mathbf{1}\mathbf{1}'}{1-\sum_{j=1}^{J} p_{\kappa}^{j}\left(z_t,\varphi_{\kappa}\right)} \\
Q\left(z_t,\varphi\right)^{-1} &= P\left(z_t,\varphi\right)^{-1} + \frac{\mathbf{1}\mathbf{1}'}{1-\sum_{j=1}^{J} p^{j}\left(z_t,\varphi\right)},
\end{aligned}
$$

(ii) $\|Q_{\kappa}(z_t,\varphi_{\kappa})\| \leq 2J$, $\|Q(z_t,\varphi_{\kappa})\| \leq 2J$,

(iii) $\left\|Q\left(z_t,\varphi\right)^{-1}\right\| \leq 2\sqrt{J/\underline{p}}$,

(iv) $\|Q_{\kappa}(z_t,\varphi_{\kappa}) - Q\left(z_t,\varphi\right)\|_{2,P} = o\left(\kappa^{-\alpha}\right)$,

(v) $\left\|Q_{\kappa}(z_t,\varphi_{\kappa})^{-1} - Q\left(z_t,\varphi\right)^{-1}\right\|_{2,P} = o\left(\kappa^{-\alpha}\right)$,

(vi) $\left\|Q_{\kappa}(z_t,\varphi)^{-1/2}\right\| \leq \sqrt{2J/\underline{p}}$

Proof. For (i) use direct calculation to verify the result. For (ii) note that $|p_\kappa^j| \leq 1$ by construction such that

$$\begin{aligned} \|Q_\kappa(z_t,\varphi_\kappa)\| &\leq \|P_\kappa(z_t,\varphi_\kappa)\| + \left\|p_\kappa(z_t,\varphi_\kappa)p_\kappa(z_t,\varphi_\kappa)'\right\| \\ &\leq \sqrt{J} + J \leq 2J. \end{aligned} \tag{C.1.2}$$

where the same bound holds for $Q(z_t,\varphi_\kappa)$. For (iii) note that

$$\left\|Q(z_t,\varphi_\kappa)^{-1}\right\| \leq \left(\sum_{j=1}^{J}\left|\frac{1}{p^j(z_t,\varphi)}\right|^2\right)^{1/2} + \sqrt{J}\left|\frac{1}{p_0^0(z_t,\varphi)}\right| \leq 2\sqrt{J}\left|\frac{1}{\underline{p}}\right|. \tag{C.1.3}$$

For (iv) use

$$\begin{aligned} &\|Q_\kappa(z_t,\varphi_\kappa) - Q(z_t,\varphi)\| \\ &\qquad\leq \|P_\kappa(z_t,\varphi_k) - P(z_t,\varphi)\| + \left\|p_\kappa(z_t,\varphi_\kappa)p_\kappa(z_t,\varphi_\kappa)' - p(z_t,\varphi)p(z_t,\varphi)'\right\| \end{aligned}$$

and

$$\begin{aligned} E\left[\|P_\kappa(z_t,\varphi_k) - P(z_t,\varphi)\|^2\right] &= E\left[\sum_{j=1}^{J}\left(p_\kappa^j(z_t,\varphi_\kappa) - p^j(z_t,\varphi)\right)^2\right] \\ &= E\left[\|p_\kappa(z_t,\varphi_\kappa) - p(z_t,\varphi)\|^2\right] = o\left(\kappa^{-2a}\right) \end{aligned}$$

by Lemma C.1.3(ii). and

$$\begin{aligned} &\left\|p_\kappa(z_t,\varphi_\kappa)p_\kappa(z_t,\varphi_\kappa)' - p(z_t,\varphi)p(z_t,\varphi)'\right\| \\ &\leq \left\|p_\kappa(z_t,\varphi_\kappa)\left(p_\kappa(z_t,\varphi_\kappa) - p(z_t,\varphi)\right)'\right\| + \left\|\left(p_\kappa(z_t,\varphi_\kappa) - p(z_t,\varphi)\right)p(z_t,\varphi)'\right\| \\ &\leq 2\sqrt{J}\,\|p_\kappa(z_t,\varphi_\kappa) - p(z_t,\varphi)\| \end{aligned}$$

where $\left\|p_\kappa(z_t,\varphi_\kappa)\left(p_\kappa(z_t,\varphi_\kappa) - p(z_t,\varphi)\right)'\right\| \leq \|p_\kappa(z_t,\varphi_\kappa)\|\,\|(p_\kappa(z_t,\varphi_\kappa) - p(z_t,\varphi))\|$ and $\|p_\kappa(z_t,\varphi_\kappa)\| \leq \sqrt{J}$ was used. Jensen's inequality and Lemma C.1.3(ii) then give the result. For (v) use the inequality from Lewis and Reinsel (1985, p. 397),

$$\frac{\left\|Q_\kappa(z_t,\varphi_\kappa)^{-1} - Q(z_t,\varphi_\kappa)^{-1}\right\|}{2\sqrt{J/\underline{p}}\left(\left\|Q_\kappa(z_t,\varphi_\kappa)^{-1/2} - Q(z_t,\varphi_\kappa)^{-1/2}\right\| + 2\sqrt{J/\underline{p}}\right)} \leq \|Q_\kappa(z_t,\varphi_\kappa) - Q(z_t,\varphi_\kappa)\|$$

which implies that

$$\left\|Q_\kappa(z_t,\varphi_\kappa)^{-1} - Q(z_t,\varphi_\kappa)^{-1}\right\|_{2,P} = o\left(\kappa^{-\alpha}\right)$$

by using the result in (iv).

For (vi) use (i) to obtain

$$\left\|Q_\kappa(z_t,\varphi_\kappa)^{-1/2}\right\|^2 = \operatorname{tr} Q_\kappa(z_t,\varphi_\kappa)^{-1} = \sum_{j=1}^{J}\frac{1}{p_\kappa^j(z_t,\varphi_\kappa)} + J\frac{1}{p_\kappa^0(z_t,\varphi_\kappa)}.$$

Now take

$$\frac{1}{p_{\kappa}^{j}\left(z_{t},\varphi_{\kappa}\right)}=\frac{1}{p^{j}\left(z_{t},\varphi\right)}+\frac{p^{j}\left(z_{t},\varphi\right)-p_{\kappa}^{j}\left(z_{t},\varphi_{\kappa}\right)}{p_{\kappa}^{j}\left(z_{t},\varphi_{\kappa}\right)p^{j}\left(z_{t},\varphi\right)}$$
$$\leq\frac{1}{p^{j}\left(z_{t},\varphi\right)}+\frac{1}{\underline{p}}\left(\frac{1}{\underline{p}}+\left(\frac{1}{p_{\kappa}^{j}\left(z_{t},\varphi_{\kappa}\right)}-\frac{1}{p^{j}\left(z_{t},\varphi\right)}\right)\right)\left(p^{j}\left(z_{t},\varphi\right)-p_{\kappa}^{j}\left(z_{t},\varphi_{\kappa}\right)\right)$$

such that

$$\left(E\left[\frac{1}{p_{\kappa}^{j}\left(z_{t},\varphi_{\kappa}\right)}-\frac{1}{p^{j}\left(z_{t},\varphi\right)}\right]\right)^{2}$$
$$\leq\frac{1}{\underline{p}^{2}}E\left[\left(\frac{1}{\underline{p}}+\left(\frac{1}{p_{\kappa}^{j}\left(z_{t},\varphi_{\kappa}\right)}-\frac{1}{p^{j}\left(z_{t},\varphi\right)}\right)\right)^{2}\right]E\left[\left(p^{j}\left(z_{t},\varphi\right)-p_{\kappa}^{j}\left(z_{t},\varphi_{\kappa}\right)\right)^{2}\right]$$

and by Hölder's inequality

$$\left(\frac{1}{\underline{p}}+\left(\frac{1}{p_{\kappa}^{j}\left(z_{t},\varphi_{\kappa}\right)}-\frac{1}{p^{j}\left(z_{t},\varphi\right)}\right)\right)^{2}\leq\left(\frac{1}{\underline{p}}+\left|\frac{1}{p_{\kappa}^{j}\left(z_{t},\varphi_{\kappa}\right)}-\frac{1}{p^{j}\left(z_{t},\varphi\right)}\right|\right)^{2}$$
$$\leq2\left(\frac{1}{\underline{p}}^{2}+\left|\frac{1}{p_{\kappa}^{j}\left(z_{t},\varphi_{\kappa}\right)}-\frac{1}{p^{j}\left(z_{t},\varphi\right)}\right|^{2}\right)$$

□

LEMMA C.1.5. *Define* $P_{\kappa}\left(z_{t},\varphi\right)=\operatorname{diag}\left(p_{\kappa}(z_{t},\varphi_{\kappa})\right)$, $\mathcal{D}_{t}=\left(D_{t,1},...,D_{t,J}\right)'$, $Q\left(z\right)=\operatorname{diag}\left(p\left(z_{t},\varphi\right)\right)-p\left(z_{t},\varphi\right)p\left(z_{t},\varphi\right)'$ *and* $Q_{\kappa}(z_{t},\varphi_{\kappa})=\left(\operatorname{diag}\left(p_{\kappa}(z_{t},\varphi_{\kappa})\right)-p_{\kappa}(z_{t},\varphi_{\kappa})p_{\kappa}\left(z_{t},\varphi_{\kappa}\right)'\right)$. *Let*

$$H\left(\varphi_{\kappa}\right)=-E[\partial^{2}l_{\kappa}\left(D_{t},z_{t,}\varphi_{k}\right)/\partial\varphi_{k}\partial\varphi_{k}']$$
$$\Omega_{\varphi,\kappa}=E\left[\partial l_{\kappa}\left(D_{t},z_{t,}\varphi_{k}\right)/\partial\varphi_{k}\partial l_{\kappa}\left(D_{t},z_{t,}\varphi_{k}\right)/\partial\varphi_{k}'\right]$$
$$H_{T}\left(\varphi_{\kappa}\right)=T^{-1}\sum_{t=1}^{T}\partial^{2}l_{\kappa}\left(D_{t},z_{t,}\varphi_{k}\right)/\partial\varphi_{k}\partial\varphi_{k}'$$

If Conditions 3.2, 3.3(i) and 3.4 hold, then,

i)
$$\dot{p}_{\kappa}\left(z_{t},\varphi_{\kappa}^{*}\right)'=\frac{\partial p_{\kappa}\left(z_{t},\varphi_{\kappa}\right)}{\partial\varphi_{\kappa}'}=Q_{\kappa}\left(z_{t},\varphi_{\kappa}\right)\otimes\Psi^{\kappa}\left(z_{t}\right)' \tag{C.1.4}$$

ii)
$$\partial l_{\kappa}\left(D_{t},z_{t,}\varphi_{k}\right)/\partial\varphi_{k}'=\left(\mathcal{D}_{t}-p_{\kappa}(z_{t},\varphi_{\kappa})\right)\otimes\Psi^{\kappa}\left(z_{t}\right)' \tag{C.1.5}$$

iii)
$$\partial^{2}l_{\kappa}\left(D_{t},z_{t,}\varphi_{k}\right)/\left(\partial\varphi_{k}\partial\varphi_{k}'\right)=-Q_{\kappa}\left(z_{t},\varphi_{\kappa}\right)\otimes\Psi^{\kappa}\left(z_{t}\right)\Psi^{\kappa}\left(z_{t}\right)' \tag{C.1.6}$$

iv) $E\left[\left\|H_{T}\left(\varphi_{\kappa}\right)-H\left(\varphi_{\kappa}\right)\right\|^{2}\right]=O\left(\zeta\left(\kappa\right)^{2}/T\right)$,

v) $\|H(\varphi_\kappa) - \Omega_{\varphi,\kappa}\| = O\left(\kappa^{-\alpha/2}\zeta(\kappa)^{1/2}\right)$,
vi) for $\bar{\varphi}_\kappa$ *such that* $\|\bar{\varphi}_\kappa - \varphi^*_\kappa\| \leq \|\hat{\varphi}_\kappa - \varphi^*_\kappa\|$ *it follows that* $\|H(\varphi^*_\kappa) - H(\bar{\varphi}_\kappa)\| = O\left(\zeta(\kappa)^{2/(4+\delta)}\|\varphi^*_\kappa - \bar{\varphi}_\kappa\|\right)$,
vii) The smallest eigenvaule of $\Omega_{\varphi,\kappa}$ *is bounded away from zero uniformly in* κ *and* $\left\|H(\varphi^*_\kappa)^{-1}\right\|_2$ *is bounded for all* κ.

Proof. For (i), note that by direct calculation using (C.1.4) and the definition of $\Psi_{j,\kappa}(z_t,\varphi_\kappa)$ one obtains.

$$\begin{aligned}\frac{\partial p^j_\kappa(z_t,\varphi_\kappa)}{\partial\varphi'_{j\kappa}} &= \left(p^j_\kappa(z_t,\varphi_\kappa) - p^j_\kappa(z_t,\varphi_\kappa)^2\right)\Psi^\kappa(z_t)' \qquad (C.1.7)\\ \frac{\partial p^j_\kappa(z_t,\varphi_\kappa)}{\partial\varphi'_{l\kappa}} &= -p^l_\kappa(z_t,\varphi_\kappa)\,p^j_\kappa(z_t,\varphi_\kappa)\,\Psi^\kappa(z_t)'\end{aligned}$$

such that (i) follows immediately upon stacking the results in (C.1.7).

For (ii) note that it follows from (C.1.65) that

$$\partial l(D_t, z_{t,}\varphi_k)/\partial\varphi'_k = (\mathcal{D}_t - p_\kappa(z_t,\varphi_\kappa))'\,Q_\kappa(z_t,\varphi_\kappa)^{-1}\,\partial p_\kappa(z_t,\varphi_\kappa)/\partial\varphi'_\kappa. \qquad (C.1.8)$$

As in Cattaneo (2010), (ii) follows from combining (C.1.8) with (i). Finally, (iii) follows by differentiating (C.1.5) and applying (C.1.4).

For (iv) note that

$$\left\|Q_\kappa(z_{t+l},\varphi_\kappa)\otimes\Psi^\kappa(z_{t+l})\,\Psi^\kappa(z_{t+l})'\right\| \leq 2J\,\|\Psi^\kappa(z_{t+l})\|^2.$$

Then use Lemma C.1.1 to bound

$$\begin{aligned}&E\left[\|H_T(\varphi_\kappa) - H(\varphi_\kappa)\|^2\right] &(C.1.9)\\ &= T^{-1}\sum_{t=1-T}^{T-1}\left(1-\frac{|l|}{T}\right) &(C.1.10)\\ &\qquad\times \operatorname{trCov}\left(Q_\kappa(z_t,\varphi_\kappa)\otimes\Psi^\kappa(z_t)\,\Psi^\kappa(z_t)', Q_\kappa(z_{t+l},\varphi_\kappa)\otimes\Psi^\kappa(z_{t+l})\,\Psi^\kappa(z_{t+l})'\right)\\ &\leq T^{-1}8J\left\|\|\Psi^\kappa(z_t)\|^2\right\|^2_{2,\beta}. &(C.1.11)\end{aligned}$$

By Lemma C.1.2 it follows that

$$\left\|\|\Psi^\kappa(z_t)\|^2\right\|^2_{2,\beta} \leq C\,\|\Psi^\kappa(z_t)\|^2_{4+\delta,P} = \zeta(\kappa)^{2/(4+\delta)}$$

which establishes the result.

For (v) note that by (iii) it follows that

$$\Omega_{\varphi,\kappa} = -E\left[(\mathcal{D}_t - p_\kappa(z_t,\varphi_\kappa))\,(\mathcal{D}_t - p_\kappa(z_t,\varphi_\kappa))'\otimes\Psi^\kappa(z_t)\,\Psi^\kappa(z_t)'\right]$$

where

$$E\left[(\mathcal{D}_t - p_\kappa(z_t,\varphi_\kappa))\,(\mathcal{D}_t - p_\kappa(z_t,\varphi_\kappa))'\,|z_t\right] \qquad (C.1.12)$$

$$= E\left[\left(\mathcal{D}_t - p\left(z_t,\varphi\right)\right)\left(\mathcal{D}_t - p(z_t)\right)' \left|z_t\right.\right] + \left(p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa)\right)\left(p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa)\right)'$$
$$= Q\left(z_t\right) + \left(p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa)\right)\left(p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa)\right)'$$

and consider

$$\begin{aligned}
&\left\|H\left(\varphi_\kappa\right) - \Omega_{\varphi,\kappa}\right\| \\
&\leq E\left[\left\|\left(Q_\kappa\left(z_t,\varphi_\kappa\right) - Q\left(z_t\right)\right) \otimes \Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right\|\right] \\
&+ E\left[\left\|\left(p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa)\right)\left(p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa)\right)' \otimes \Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right\|\right].
\end{aligned} \tag{C.1.13}$$

Next use the fact that

$$\left\|\left(Q_\kappa\left(z_t,\varphi_\kappa\right) - Q\left(z_t\right)\right) \otimes \Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right\| = \left\|Q_\kappa\left(z_t,\varphi_\kappa\right) - Q\left(z_t\right)\right\| \left\|\Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right\| \tag{C.1.14}$$

where (C.1.14) implies that

$$\begin{aligned}
&E\left[\left\|\left(Q_\kappa\left(z_t,\varphi_\kappa\right) - Q\left(z_t\right)\right) \otimes \Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right\|\right] \\
&\leq E\left[\left\|\left(Q_\kappa\left(z_t,\varphi_\kappa\right) - Q\left(z_t\right)\right)\right\| \left\|\Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right\|\right]
\end{aligned} \tag{C.1.15}$$

where

$$\begin{aligned}
\left\|Q_\kappa\left(z_t,\varphi_\kappa\right) - Q\left(z_t\right)\right\| &= \left\|\operatorname{diag}\left(p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa)\right)\right\| \\
&\quad + \left\|p\left(z_t,\varphi\right)p\left(z_t,\varphi\right)' - p_\kappa(z_t,\varphi_\kappa)p_\kappa(z_t,\varphi_\kappa)'\right\| \qquad \text{(C.1.16)} \\
&\leq \left\|p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa)\right\| \\
&\quad + \left(\left\|p_\kappa(z_t,\varphi_\kappa)\right\| + \left\|p\left(z_t,\varphi\right)\right\|\right)\left\|p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa)\right\| \\
&\leq 3\left\|p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa)\right\|
\end{aligned}$$

Now use $p\left(z_t,\varphi\right) = \Gamma\left(\Gamma^{-1}\left(p\left(z_t,\varphi\right)\right)\right),\ p_\kappa(z_t,\varphi_\kappa) = \Gamma\left(\Psi_\kappa\left(z_t,\varphi_\kappa\right)\right),$ the mean value theorem leads to

$$p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa) = \frac{\partial\Gamma}{\partial\Psi}\left(\Gamma^{-1}\left(p\left(z_t,\varphi\right)\right) - \Psi_\kappa\left(z_t,\varphi_\kappa\right)\right) \tag{C.1.17}$$

and because the elements of $\frac{\partial\Gamma}{\partial\Psi}$ are bounded by 1/2 it follows as shown in Cattaneo (2010) that

$$\left\|p\left(z_t,\varphi\right) - p_\kappa(z_t,\varphi_\kappa)\right\| \leq C\left\|\Gamma^{-1}\left(p\left(z_t,\varphi\right)\right) - \Psi_\kappa\left(z_t,\varphi_\kappa\right)\right\| \tag{C.1.18}$$

for some constant. By Condition 3.4 and Lemma C.1.3(ii) it follows that

$$E\left[\left\|\Gamma^{-1}\left(p\left(z\right)\right) - \Psi_\kappa\left(z,\varphi_\kappa\right)\right\|^2\right] = O\left(\kappa^{-\alpha}\right)$$

such that the second line in (C.1.15) is bounded by

$$\begin{aligned}
&E\left[\left\|Q_\kappa\left(z_t,\varphi_\kappa\right) - Q\left(z_t\right) \otimes \Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right\|\right] \\
&\leq 3\left(E\left[\left\|\Gamma^{-1}\left(p\left(z\right)\right) - \Psi_\kappa\left(z,\varphi_\kappa\right)\right\|^2\right]\right)^{1/2}\left(E\left[\left\|\Psi^\kappa\left(z_t\right)\right\|^2\right]\right)^{1/2} = O\left(\kappa^{-\alpha/2}\zeta\left(\kappa\right)^{1/2}\right).
\end{aligned} \tag{C.1.19}$$

For the second line in (C.1.13) similarly use

$$E\left[\left\|\left(p\left(z_t,\varphi\right)-p_\kappa(z_t,\varphi_\kappa)\right)\left(p\left(z_t,\varphi\right)-p_\kappa(z_t,\varphi_\kappa)\right)'\otimes\Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right\|\right] \tag{C.1.20}$$
$$=2JE\left[\left\|p\left(z_t,\varphi\right)-p_\kappa(z_t,\varphi_\kappa)\right\|\left\|\Psi^\kappa\left(z_t\right)\right\|^2\right]$$
$$\leq CE\left[\left\|\Gamma^{-1}\left(p\left(z\right)\right)-\Psi_\kappa\left(z,\varphi_\kappa\right)\right\|^2\right]^{1/2}E\left[\left\|\Psi^\kappa\left(z_t\right)\right\|^4\right]^{1/2}=O\left(\kappa^{-\alpha/2}\zeta\left(\kappa\right)^{1/2}\right) \tag{C.1.21}$$

It now follows that

$$\left\|H\left(\varphi_\kappa\right)-\Omega_{\varphi,\kappa}\right\|=O\left(\kappa^{-\alpha/2}\zeta\left(\kappa\right)^{1/2}\right). \tag{C.1.22}$$

For (vi) use

$$\begin{aligned}\left\|H\left(\varphi_\kappa^*\right)-H\left(\bar{\varphi}_\kappa\right)\right\| &\leq E\left[\left\|\left(Q_\kappa\left(z_t,\varphi_\kappa^*\right)-Q_\kappa\left(z_t,\bar{\varphi}_\kappa\right)\right)\right\|\left\|\Psi^\kappa\left(z_t\right)\right\|^2\right]\\ &\leq E\left[\left\|\left(Q_\kappa\left(z_t,\varphi_\kappa^*\right)-Q_\kappa\left(z_t,\bar{\varphi}_\kappa\right)\right)\right\|^2\right]^{1/2}E\left[\left\|\Psi^\kappa\left(z_t\right)\right\|^4\right]^{1/2}\end{aligned}$$

where by (C.1.16) and (C.1.18) it follows that

$$\begin{aligned}E\left[\left\|\left(Q_\kappa\left(z_t,\varphi_\kappa^*\right)-Q_\kappa\left(z_t,\bar{\varphi}_\kappa\right)\right)\right\|^2\right] &\leq E\left[\left\|\Psi_\kappa\left(z,\varphi_\kappa^*\right)-\Psi_\kappa\left(z,\bar{\varphi}_\kappa\right)\right\|^2\right]\\ &\leq E\left[\left\|\Psi^\kappa\left(z_t\right)\right\|^2\right]\left\|\varphi_\kappa^*-\bar{\varphi}_\kappa\right\|^2\\ &=O\left(\zeta\left(\kappa\right)^{1/2}\left\|\varphi_\kappa^*-\bar{\varphi}_\kappa\right\|^2\right)\end{aligned}$$

which establishes the result.

For (vii) consider

$$\Omega_{\varphi,\kappa}=E\left[\left(Q\left(z_t\right)+\left(p\left(z_t,\varphi\right)-p_\kappa(z_t,\varphi_\kappa)\right)\left(p\left(z_t,\varphi\right)-p_\kappa(z_t,\varphi_\kappa)\right)'\right)\otimes\Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right]. \tag{C.1.23}$$

From the fact that $\left(p\left(z_t,\varphi\right)-p_\kappa(z_t,\varphi_\kappa)\right)\left(p\left(z_t,\varphi\right)-p_\kappa(z_t,\varphi_\kappa)\right)'\otimes\Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'$ is positive semi-definite for all values of z_t it follows that

$$\Omega_{\varphi,\kappa}\geq E\left[Q\left(z_t\right)\otimes\Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right] \tag{C.1.24}$$

where the inequality is in the sense of positive semi-definite matrices. Now using the fact that $Q\left(z_t\right)\geq\underline{p}^JI_J$ as implied by the results in Cattaneo (2010) and Condition 3.3-(i), it follows from Condition 3.4 and Magnus and Neudecker (1980) that

$$\begin{aligned}\lambda_{\min}\left(E\left[Q\left(z_t\right)\otimes\Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right]\right) &\geq\lambda_{\min}E\left[\underline{p}^J\left(I_J\otimes\Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right)\right] \qquad \text{(C.1.25)}\\ &\geq\underline{p}^J\lambda_{\min}E\left[\Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right]>\delta>0.\end{aligned}$$

Now use the inequality

$$\left\|H\left(\varphi_\kappa\right)^{-1}\right\|_2\leq\left\|\Omega_{\varphi,\kappa}^{-1}\right\|_2+\left\|H\left(\varphi_\kappa\right)^{-1}-\Omega_{\varphi,\kappa}^{-1}\right\|_2$$

where $\left\|\Omega_{\varphi,\kappa}^{-1}\right\|_2$ is bounded because $\Omega_{\varphi,\kappa}$ is real symmeric such that the largest eigenvalue of $\Omega_{\varphi,\kappa}^{-1}$ is the inverse of the smallest eigenvalue of $\Omega_{\varphi,\kappa}$ and therefore is bounded because of (C.1.25) (see Böttcher, 1996, p. 16). The second term tends to zero because of (v) and the fact that $\left\|\Omega_{\varphi,\kappa}^{-1}\right\|_2$ is bounded (see Lewis and Reinsel, 1985, p. 397). □

LEMMA C.1.6. *Let* $b(z)$, $b^0(z)$ *and* $B(z)$ *be as defined in Condition 3.3. Then*
i) for $\bar{\varphi}_\kappa$ *such that* $\|\varphi^*_\kappa - \bar{\varphi}_\kappa\| \leq \|\varphi^*_\kappa - \hat{\varphi}_\kappa\|$ *it follows that*

$$\left\|\widehat{\mathcal{K}}_\kappa(\bar{\varphi}_\kappa) - \mathcal{K}_\kappa(\varphi^*_\kappa)\right\|_2 = O\left(\frac{\zeta(\kappa)}{\sqrt{T}} + \|\varphi^*_\kappa - \bar{\varphi}_\kappa\|\right) = o_p(1)$$

ii)

$$\left(\widehat{\mathcal{K}}_\kappa(\bar{\varphi}_\kappa) - \mathcal{K}_\kappa(\varphi^*_\kappa)\right) T^{-1/2}\sum_{t=1}^{T\prime} \dot{p}_\kappa(z_t,\varphi^*_\kappa) Q_\kappa(z_t,\varphi^*_\kappa)^{-1}(\mathcal{D}_t - p(z_t,\varphi))$$
$$= O_p\left(\frac{\zeta(\kappa)^{3+\delta/2/(2+\delta/2)}}{\sqrt{T}} + \|\varphi^*_\kappa - \bar{\varphi}_\kappa\|\,\zeta(\kappa)^{1/(2+\delta/2)}\right) = o_p(1),$$

iii) $\widehat{\mathcal{K}}_\kappa(\varphi^*_\kappa) T^{-1/2}\sum_{t=1}^{T\prime} \dot{p}_\kappa(z_t,\varphi^*_\kappa) Q_\kappa(z_t,\varphi^*_\kappa)^{-1}(p(z_t,\varphi) - p(z_t,\varphi^*_\kappa)) = O_p\left(\zeta(\kappa)^{1/2}\kappa^{\alpha/2}\right) = o_p(1)$,
iv) $T^{-1/2}\sum_{t=1}^{T\prime}\left(\mathcal{K}_\kappa(\varphi^*_\kappa)\dot{p}_\kappa(z_t,\varphi^*_\kappa) Q_\kappa(z_t,\varphi^*_\kappa)^{-1} - B(z_t)\right)(\mathcal{D}_t - p(z_t,\varphi)) = o_p(1)$.
v) $\|B(z)\|^2 \leq J^2 E\left[\|Y_t\|^2 |z_t\right]\left(\frac{2(1+\underline{p})}{\underline{p}}\right)^2$ *and* $\int \|B(z)\|^2 dP_0(z) \leq J^2\left(\frac{2(1+\underline{p})}{\underline{p}}\right)^2 E\left[\|Y_t\|^2\right]$
vi) $\int B(z)(p_\kappa(z) - p_0(z))\,dP_0(z) = o(1)$
vii) $\int B(z)\,dP_0(z) = O(1)$

Proof. To show (i) define $H(\varphi^*_\kappa) = -E\left[\partial^2 l(D_t, z_{t,}\varphi^*_k)/\partial\varphi_k\partial\varphi_k'\right]$ by Lemma C.1.5(vii) it follows that

$$\left\|H(\varphi^*_\kappa)^{-1}\right\|_2 \leq C_1 < \infty \tag{C.1.26}$$

for all κ. Let $B_{q,i,j}(z)$ be the i,j-th element of the matrix $B(z)Q_\kappa(z_t,\varphi_\kappa)$. Now consider

$$\begin{aligned}\|\mathcal{K}_\kappa(\varphi^*_\kappa)\|^2 &\leq \left\|\int B(z)\,\dot{p}_\kappa(z,\varphi^*_\kappa)'\,dP_0\right\|^2 \left\|H(\varphi^*_\kappa)^{-1}\right\|_2^2 \\ &\leq C\left\|\int B(z)\left(Q_\kappa(z_t,\varphi^*_\kappa)\otimes\Psi^\kappa(z_t)'\right)dP_0\right\| \\ &\leq C_1\left(\sum_{j=l}^{\kappa}\sum_{i,j=1}^{J}\left(\int B_{q,ij}(z)\,\psi_{l\kappa}(z_t)'\,dP_0\right)^2\right)^{1/2} \leq C < \infty,\end{aligned} \tag{C.1.27}$$

where the first inequality uses Lewis and Reinsel (1985, Equation 2.6) and the second inequality uses (C.1.26). The last inequality follows by Condition 3.4-(iv) if it can be established that $B_{q,ij}(z) \in L_2(P_0)$. By (v) below, Lemma C.1.4-(ii), and the Cauchy-Schwarz inequality $\left\|\int B(z)Q_\kappa(z_t,\varphi_\kappa)\,dP_0\right\|^2 \leq \int\|B(z)\|^2 dP_0 \int \|Q_\kappa(z_t,\varphi_\kappa)\|^2 dP_0 \leq 4J^2\int\|B(z)\|^2 dP_0 < \infty$ which establishes that $B_{ij}(z) \in L_2(P_0)$ as required by Condition 3.4(iv).

Next note that by arguments given in Lewis and Reinsel (1985, p. 397), it follows that

$$\left\|H_T(\bar{\varphi}_\kappa)^{-1} - H(\bar{\varphi}_\kappa)^{-1}\right\|_2 = O_p\left(\|H_T(\bar{\varphi}_\kappa) - H(\bar{\varphi}_\kappa)\|_2\right)$$

and consequently $\left\|H_T(\bar{\varphi}_\kappa)^{-1}\right\|_2 = O_p(1)$ because $\left\|H(\bar{\varphi}_\kappa)^{-1}\right\|_2$ is bounded by Lemma C.1.5(vii). Then, one obtains

$$\begin{aligned}\left\|\int B(z)\,\dot{p}_\kappa(z,\varphi_\kappa^*)'\,dP_0 H_T(\bar{\varphi}_\kappa)^{-1} - \mathcal{K}_\kappa(\varphi_\kappa^*)\right\| \\ = O_p\left(\|\mathcal{K}_\kappa(\varphi_\kappa^*)\|_2\right) O_p\left(\left\|H_T(\bar{\varphi}_\kappa)^{-1}\right\|_2\right) O_p\left(\|H_T(\bar{\varphi}_\kappa) - H(\varphi_\kappa^*)\|_2\right)\end{aligned} \quad \text{(C.1.28)}$$

where by Lemma C.1.5(iv) and the fact that $\|.\|_2 \leq \|.\|$ one has

$$\|H_T(\bar{\varphi}_\kappa) - H(\bar{\varphi}_\kappa)\|_2 = O_p\left(\zeta(\kappa)/\sqrt{T}\right) \quad \text{(C.1.29)}$$

and by Lemma C.1.5(vi) that

$$\|H_T(\bar{\varphi}_\kappa) - H_T(\varphi_\kappa^*)\|_2 = O\left(\zeta(\kappa)^{1/2}\|\varphi_\kappa^* - \bar{\varphi}_\kappa\|\right). \quad \text{(C.1.30)}$$

By (C.1.27) it follows that $\|\mathcal{K}_\kappa(\varphi_\kappa^*)\| = O_p(1)$. Next, consider

$$\begin{aligned}\left\|\hat{\mathcal{K}}_\kappa(\bar{\varphi}_\kappa) - \int B(z)\,\dot{p}_\kappa(z,\varphi_\kappa^*)'\,dP_0 H_T(\bar{\varphi}_\kappa)^{-1}\right\| \\ \leq \left\|\int B(z)\left(\dot{p}_\kappa(z,\bar{\varphi}_\kappa)' - \dot{p}_\kappa(z,\varphi_\kappa^*)'\right)dP_0\right\| \left\|H_T(\bar{\varphi}_\kappa)^{-1}\right\|_2.\end{aligned} \quad \text{(C.1.31)}$$

Note that

$$\frac{\partial^2 p_\kappa^j(z_t,\varphi_\kappa)}{\partial\varphi_{j\kappa}\partial\varphi_{j\kappa}'} = \left(p_\kappa^j(z_t,\varphi_\kappa) - p_\kappa^j(z_t,\varphi_\kappa)^2\right)\left(1 - 2p_\kappa^j(z_t,\varphi_\kappa)^2\right)\Psi^\kappa(z_t)\,\Psi^\kappa(z_t)' \quad \text{(C.1.32)}$$

$$\frac{\partial p_\kappa^j(z_t,\varphi_\kappa)}{\partial\varphi_{l\kappa}'} = -p_\kappa^l(z_t,\varphi_\kappa)\,p_\kappa^j(z_t,\varphi_\kappa)\left(1 - 2p_\kappa^j(z_t,\varphi_\kappa)^2\right)\Psi^\kappa(z_t)\,\Psi^\kappa(z_t)'$$

such that

$$\left\|\frac{\partial^2 p_\kappa^j(z_t,\varphi_\kappa)}{\partial\varphi_{j\kappa}\partial\varphi_{j\kappa}'}\right\| \leq 3\left\|\Psi^\kappa(z_t)\,\Psi^\kappa(z_t)'\right\| \quad \text{(C.1.33)}$$

with the same bound holding for the cross-derivatives. Now bounding (C.1.30) one obtains from (C.1.33) that

$$\begin{aligned}\left\|\int B(z)\left(\dot{p}_\kappa(z,\bar{\varphi}_\kappa)' - \dot{p}_\kappa(z,\varphi_\kappa^*)'\right)dP_0\right\| \\ \leq 3J^2\int\|B(z)\|\,\|\Psi^\kappa(z_t)\|^2\,dP_0\,\|\varphi_\kappa^* - \bar{\varphi}_\kappa\|\end{aligned} \quad \text{(C.1.34)}$$

$$\leq 3J^2 \left(\int \|B(z)\|^2 \, dP_0 \right)^{1/2} \left(\int \|\Psi^\kappa(z_t)\|^4 \, dP_0 \right)^{1/2} \|\varphi_\kappa^* - \bar{\varphi}_\kappa\|$$
$$= O\left(\zeta(\kappa)^{1/2} \|\varphi_\kappa^* - \bar{\varphi}_\kappa\| \right)$$

and (C.1.28), (C.1.29), (C.1.30), (C.1.31) and (C.1.34) imply that

$$\left\| \widehat{\mathcal{K}}_\kappa(\bar{\varphi}_\kappa) - \mathcal{K}_\kappa(\varphi_\kappa^*) \right\| = O_p\left(\frac{\zeta(\kappa)}{\sqrt{T}} + \zeta(\kappa)^{1/2} \|\varphi_\kappa^* - \bar{\varphi}_\kappa\| \right).$$

For (ii) use the submultiplicative inequality for Euclidean matrix norms (see Horn and Johnson, 1985, p. 291), the first term (C.1.70) is bounded by

$$\left\| \widehat{\mathcal{K}}_\kappa(\bar{\varphi}_\kappa) - \mathcal{K}_\kappa(\varphi_\kappa^*) \right\|_2 \left\| T^{-1/2} \sum_{t=1}^{T\prime} \dot{p}_\kappa(z_t, \varphi_\kappa^*) \, Q_\kappa(z_t, \varphi_\kappa^*)^{-1} (\mathcal{D}_t - p(z_t, \varphi)) \right\|. \tag{C.1.35}$$

The second term in (C.1.35) is

$$E\left[\left\| T^{-1/2} \sum\nolimits_{t=1}^{T} \dot{p}_\kappa(z_t, \varphi_\kappa^*) \, Q_\kappa(z_t, \varphi_\kappa^*)^{-1} (\mathcal{D}_t - p(z_t)) \right\|^2 \right] \tag{C.1.36}$$
$$= T^{-1} \operatorname{tr} \left(\sum_{t=1}^{T} E\left[\dot{p}_\kappa(z_t, \varphi_\kappa^*) \, Q_\kappa(z_t, \varphi_\kappa^*)^{-1} Q(z_t, \varphi) \, Q_\kappa(z_t, \varphi_\kappa^*)^{-1} \dot{p}_k(z_t, \varphi_\kappa^*)' \right] \right)$$
$$= \operatorname{tr} E\left[Q(z_1, \varphi) \otimes \Psi^\kappa(z_1) \, \Psi^\kappa(z_1)' \right] = E\left[\left\| Q(z_1, \varphi)^{1/2} \right\|^2 \|\Psi^\kappa(z_1)\|^2 \right]$$
$$\leq 2J \left(E\left[\|\Psi^\kappa(z_1)\|^{4+\delta} \right] \right)^{1/(2+\delta/2)} = 2J\zeta(\kappa)^{1/(2+\delta/2)}$$

such that upon combining (C.1.35), (C.1.36), (C.1.29) and (C.1.30) it follows that (ii) is

$$O\left(\frac{\zeta(\kappa)^{3+\delta/2/(2+\delta/2)}}{\sqrt{T}} + \|\varphi_\kappa^* - \bar{\varphi}_\kappa\| \, \zeta(\kappa)^{1/2+1/(2+\delta/2)} \right) = o_p(1).$$

For (iii) recall the definition of $\varphi_\kappa^* = \arg\max L(\varphi_\kappa)$ with $L(\varphi_\kappa) = E[l_\kappa(D_t, z_t, \varphi_\kappa)]$ which implies that

$$0 = E\left[\frac{\partial l_\kappa(D_t, z_t, \varphi_\kappa^*)}{\partial \varphi_\kappa} \right] = E\left[\dot{p}_\kappa(z_t, \varphi_\kappa^*)' \, Q_\kappa(z_t, \varphi_\kappa^*)^{-1} (\mathcal{D}_t - p_\kappa(z_t, \varphi_\kappa^*)) \right]$$

and consequently, because of $E[D_t|z_t] = p(z_t, \varphi)$, it follows that

$$E\left[\dot{p}_\kappa(z_t, \varphi_\kappa^*)' \, Q_\kappa(z_t, \varphi_\kappa^*)^{-1} (p(z_t, \varphi) - p_\kappa(z_t, \varphi_\kappa^*)) \right] = 0.$$

This shows that for $\Upsilon_\kappa(z_t, \varphi_\kappa^*) = (p(z_t, \varphi) - p_\kappa(z_t, \varphi_\kappa^*)) \otimes \Psi^\kappa(z_t)$

$$E \left\| T^{-1/2} \sum_{t=1}^{T} (p(z_t, \varphi) - p_\kappa(z_t, \varphi_\kappa^*)) \otimes \Psi^\kappa(z_t) \right\|^2 \tag{C.1.37}$$

$$= \operatorname{tr} \sum_{l=-T+1}^{T-1} \left(1 - \frac{|l|}{T}\right) \operatorname{Cov}\left(\Upsilon_\kappa\left(z_t, \varphi_\kappa^*\right), \Upsilon_\kappa\left(z_{t+l}, \varphi_\kappa^*\right)'\right).$$

Noting that

$$\left|\psi_{l,\kappa}\left(z_t\right)\left(p^j\left(z_t\right) - p_\kappa^j(z_t, \varphi_\kappa^*)\right)\right| \leq \sup_z \left|\frac{p^j\left(z\right) - p_\kappa^j(z, \varphi_\kappa^*)}{\langle z \rangle}\right| \langle z_t \rangle \left|\psi_{l,\kappa}\left(z_t\right)\right|$$

It now follows from Lemma C.1.1 that

$$\operatorname{tr} \sum_{l=-T+1}^{T-1} \operatorname{Cov}\left(\Upsilon_\kappa\left(z_t, \varphi_\kappa^*\right), \Upsilon_\kappa\left(z_{t+l}, \varphi_\kappa^*\right)'\right)$$
$$\leq 4 \sum_{j=1}^{J} \sup_z \left|\frac{p^j\left(z\right) - p_\kappa^j(z, \varphi_\kappa^*)}{\langle z \rangle}\right| \sum_{l=1}^{J\kappa} \left\|\psi_{l,\kappa}\left(z_t\right) \langle z_t \rangle\right\|_{2,\beta}^2. \quad \text{(C.1.38)}$$

Using (C.1.1) the RHS of (C.1.38) can be bounded by

$$\begin{aligned}
&4 \sum_{j=1}^{J} \sup_z \left|\frac{p^j\left(z\right) - p_\kappa^j(z, \varphi_\kappa^*)}{\langle z \rangle}\right| \sum_{l=1}^{J\kappa} \left\|\psi_{l,\kappa}\left(z_t\right) \langle z_t \rangle\right\|_{2+\delta,P} \sqrt{1 + \int_0^1 \phi^*\left(\beta^{-1}\left(u\right)\right) du} \\
&\leq C \sum_{j=1}^{J} \sup_z \left|\frac{p^j\left(z\right) - p_\kappa^j(z, \varphi_\kappa^*)}{\langle z \rangle}\right| \sum_{l=1}^{J\kappa} \left\|\psi_{l,\kappa}\left(z_t\right)\right\|_{4+\delta,P} \left\|\langle z_t \rangle\right\|_{4+2\delta,P} \\
&= O\left(\zeta\left(\kappa\right) \kappa^{\alpha}\right)
\end{aligned}$$

where Lemma C.1.2 and Conditions 3.4(i) and (ii) were used, and the first inequality uses the Cauchy-Schwarz inequality. This shows that the bound for (C.1.71) is given by

$$\begin{aligned}
&\left\|\widehat{\mathcal{K}}_\kappa\left(\varphi_\kappa^*\right) T^{-1/2} \sum_{t=1}^{T} \dot{p}_\kappa\left(z_t, \varphi_\kappa^*\right) Q_\kappa\left(z_t, \varphi_\kappa^*\right)^{-1}\left(p\left(z_t, \varphi\right) - p_\kappa(z_t, \varphi_\kappa^*)\right)\right\| \\
&\leq \left\|\widehat{\mathcal{K}}_\kappa\left(\varphi_\kappa^*\right)\right\|_2 \left\|T^{-1/2} \sum_{t=1}^{T} \dot{p}_\kappa\left(z_t, \varphi_\kappa^*\right) Q_\kappa\left(z_t, \varphi_\kappa^*\right)^{-1/2}\left(p\left(z_t, \varphi\right) - p_\kappa(z_t, \varphi_\kappa^*)\right)\right\| \\
&= O_p\left(1\right) O_p\left(\zeta\left(\kappa\right)^{1/2} \kappa^{\alpha/2}\right).
\end{aligned}$$

where $\left\|\widehat{\mathcal{K}}_\kappa\left(\varphi_\kappa^*\right)\right\|_2 = O_p\left(1\right)$ by (i) and (C.1.27).
For (iv) consider

$$\begin{aligned}
&E\left\|T^{-1/2} \sum_{t=1}^{T} \left(\mathcal{K}_\kappa\left(\varphi_\kappa^*\right) \dot{p}_\kappa\left(z_t, \varphi_\kappa^*\right) Q_\kappa\left(z_t, \varphi_\kappa^*\right)^{-1} - B\left(z_t\right)\right)\left(\mathcal{D}_t - p\left(z_t, \varphi\right)\right)\right\|^2 \\
&\leq T^{-1} \sum_{t=1}^{T} E\left[\left\|\left(\mathcal{K}_\kappa\left(\varphi_\kappa^*\right) \dot{p}_\kappa\left(z_t, \varphi_\kappa^*\right) Q_\kappa\left(z_t, \varphi_\kappa^*\right)^{-1} - B\left(z_t\right)\right) Q\left(z_t\right)\right\|^2\right]
\end{aligned}$$

$$\leq E\left[\left\|\mathcal{K}_\kappa\left(\varphi_\kappa^*\right)\dot{p}_\kappa\left(z_1,\varphi_\kappa^*\right)Q_\kappa\left(z_1,\varphi_\kappa^*\right)^{-1/2}-B\left(z_t\right)Q_\kappa\left(z_1,\varphi_\kappa^*\right)^{1/2}\right\|^2\right.$$
$$\left.\times\left\|Q_\kappa\left(z_1,\varphi_\kappa^*\right)^{-1/2}\right\|^2\left\|Q\left(z_t\right)\right\|^2\right]$$
$$\leq 2JE\left[\left\|\left(\mathcal{K}_\kappa\left(\varphi_\kappa^*\right)\dot{p}_\kappa\left(z_1,\varphi_\kappa^*\right)Q_\kappa\left(z_1,\varphi_\kappa^*\right)^{-1}-B\left(z_t\right)\right)Q_\kappa\left(z_1,\varphi_\kappa^*\right)^{1/2}\right\|^2\right]=o\left(1\right) \tag{C.1.39}$$

where $\left\|Q\left(z_1\right)\right\|^2\leq 2J$ by Lemma C.1.4(ii) Then, using (C.1.4),

$$\mathcal{K}_\kappa\left(\varphi_\kappa^*\right)\dot{p}_\kappa\left(z_1,\varphi_\kappa^*\right)Q_\kappa\left(z_1,\varphi_\kappa^*\right)^{-1}=\mathcal{K}_\kappa\left(\varphi_\kappa^*\right)\left(I_J\otimes\Psi^\kappa\left(z_t\right)\right)$$

and

$$\begin{aligned}\mathcal{K}_\kappa\left(\varphi_\kappa^*\right)&=\int B\left(z\right)\dot{p}_\kappa\left(z,\varphi_\kappa^*\right)'dP_0\Omega_{\varphi,\kappa}^{-1}\\&=\int B\left(z\right)\left(Q_\kappa\left(z_1,\varphi_\kappa^*\right)\otimes\Psi^\kappa\left(z_1\right)\right)'dP_0E\left[Q_\kappa\left(z_1,\varphi_\kappa^*\right)\otimes\Psi^\kappa\left(z_1\right)\Psi^\kappa\left(z_1\right)'\right]^{-1}\end{aligned}$$

shows that $\mathcal{K}_\kappa\left(\varphi_\kappa^*\right)$ solves the population projection problem

$$\begin{aligned}&\min_\Pi E\left[\left\|\left(\Pi\left(I_J\otimes\Psi^\kappa\left(z_1\right)\right)-B\left(z_1\right)\right)Q_\kappa\left(z_1,\varphi_\kappa^*\right)^{1/2}\right\|^2\right]\\&\quad\leq\min_\Pi E\left[\left\|\left(\Pi\left(I_J\otimes\Psi^\kappa\left(z_1\right)\right)-B\left(z_1\right)\right)\right\|^2\left\|Q_\kappa\left(z_1,\varphi_\kappa^*\right)^{1/2}\right\|^2\right]\\&\quad\quad\leq 2J\min_\Pi E\left[\left\|\left(\Pi\left(I_J\otimes\Psi^\kappa\left(z_1\right)\right)-B\left(z_1\right)\right)\right\|^2\right]=o\left(\kappa^{-\alpha}\right)\end{aligned}$$

where Lemma C.1.4(ii) was used in the second inequality and the last equality follows from Condition 3.3(iii). This establishes the $o\left(1\right)$ result in (C.1.39).

For (iv) note that

$$\left\|b^{0,j}\left(z_t\right)\right\|\leq E\left[\left\|Y_t\right\|\left|z_t\right.\right]/\underline{p},$$

and

$$\left\|b^j\left(z_t\right)\right\|\leq E\left[\left\|Y_t\right\|\left|z_t\right.\right]/\underline{p}+2E\left[\left\|Y_t\right\|\left|z_t\right.\right]=E\left[\left\|Y_t\right\|\left|z_t\right.\right]\left(\frac{1+2\underline{p}}{\underline{p}}\right).$$

We then have

$$\left\|B\left(z\right)\right\|^2\leq\left(\left\|b\left(z\right)\right\|+\left\|b^0\left(z\right)\mathbf{1}'\right\|\right)^2$$

with

$$\left\|b\left(z\right)\right\|\leq JE\left[\left\|Y_t\right\|\left|z_t\right.\right]\left(\frac{1+2\underline{p}}{\underline{p}}\right)$$

and

$$\left\|b^0\left(z\right)\mathbf{1}'\right\|\leq JE\left[\left\|Y_t\right\|\left|z_t\right.\right]/\underline{p}$$

leading to

$$\|B(z)\|^2 \le J^2 \left(E\left[\|Y_t\| \,|z_t\right]\right)^2 \left(\frac{2\left(1+\underline{p}\right)}{\underline{p}}\right)^2 \le J^2 E\left[\|Y_t\|^2 \,|z_t\right] \left(\frac{2\left(1+\underline{p}\right)}{\underline{p}}\right)^2$$

where the second inequality uses a conditional version of Jensen's inequality. This implies that

$$\begin{aligned}
&\int \|B(z)\|^2 \, dP_0(z) \\
&\qquad \le J^2 \left(\frac{2\left(1+\underline{p}\right)}{\underline{p}}\right)^2 \int E\left[\|Y_t\|^2 \,|z_t\right] dP_0(z) = J^2 \left(\frac{2\left(1+\underline{p}\right)}{\underline{p}}\right)^2 E\left[\|Y_t\|^2\right].
\end{aligned}$$

For (v) use the Cauchy-Schwarz inequality

$$\begin{aligned}
&\left\|\int B(z)\left(p_\kappa(z) - p_0(z)\right) dP_0(z)\right\| \\
&\qquad \le \int \|B(z)\left(p_\kappa(z) - p_0(z)\right)\| \, dP_0(z) \\
&\qquad \le \int \|B(z)\| \, \|\left(p_\kappa(z) - p_0(z)\right)\| \, dP_0(z) \\
&\qquad \le \left(\int \|B(z)\|^2 \, dP_0(z)\right)^{1/2} \left(\int \|\left(p_\kappa(z) - p_0(z)\right)\|^2 \, dP_0(z)\right)^{1/2}
\end{aligned}$$

where $\int \|\left(p_\kappa(z) - p_0(z)\right)\|^2 dP_0(z) = o(1)$ by Lemma C.1.3(ii) and using (C.1.18) and where $\int \|B(z)\|^2 dP_0(z) = O(1)$ by Lemma C.1.6(iv).

For (vi) use

$$\left\|\int B(z)\, dP_0(z)\right\| \le \int \|B(z)\| \, dP_0(z) \le \left(\int \|B(z)\|^2 \, dP_0(z)\right)^{1/2} < \infty$$

where the second inequality uses a conditional version of Jensen's inequality and the last integral is bounded by (iv). □

Consider a stochastic process $\{\omega_{t,\kappa}\}_{t=1}^T$ where $\omega_{t,\kappa} \in \mathbb{R}^\kappa$ for each t. Denote $\widehat{\Omega}_T \equiv T^{-1}\sum_{t=1}^T \omega_{t,\kappa}\omega_{t,\kappa}'$ and $\overline{\Omega}_T \equiv T^{-1}\sum_{t=1}^T E[\omega_{t,\kappa}\omega_{t,\kappa}']$. Following lemma extends the result of Rudelson (1999) and Belloni et al. (2015) by generalizing the case to weakly dependent processes.

Lemma C.1.7. *For $\nu > 4$, ζ_κ, μ_κ, following conditions are satisfied: i) $E\|\omega_{t,\kappa}\|^\nu \le \zeta_\kappa$, ii) $T^{-1}\sum_{t=1}^T \|E[w_{t,\kappa}w_{t,\kappa}']\| \le \mu_\kappa$, and iii) $\omega_{t,\kappa}$ is β-mixing and satisfies Condition 3.2. Then, we have $E\|\widehat{\Omega}_T - \overline{\Omega}_T\| = O\left(\sqrt{\mu_\kappa \zeta_\kappa^{2/(\nu-2)} \log \kappa / T}\right)$.*

Proof. Set $c_\kappa = \zeta_\kappa^{\nu/(\nu-2)}$. Let

$$\widehat{\Omega}_T^- = \frac{1}{T}\sum_{t=1}^T \omega_{t,\kappa}\omega_{t,\kappa}' \mathbf{1}(\|\omega_{t,\kappa}\| \le c_\kappa)$$
$$\widehat{\Omega}_T^+ = \frac{1}{T}\sum_{t=1}^T \omega_{t,\kappa}\omega_{t,\kappa}' \mathbf{1}(\|\omega_{t,\kappa}\| > c_\kappa).$$

By triangular inequality, the deviation can be decomposed as

$$E\|\widehat{\Omega}_T - \overline{\Omega}_T\| = \underbrace{E\|\widehat{\Omega}_T^- - E[\widehat{\Omega}_T^-]\|}_{(1)} + \underbrace{E\|\widehat{\Omega}_T^+ - E[\widehat{\Omega}_T^+]\|}_{(2)},$$

For (1), note that $\|\omega_{t,\kappa}\omega_{t,\kappa}'\mathbf{1}(\|\omega_{t,\kappa}\| \le c_\kappa)\| = \|\omega_{t,\kappa}\mathbf{1}(\|\omega_{t,\kappa}\| \le c_\kappa)\|^2 \le c_\kappa^2$. And also,

$$\begin{aligned}
\|E[\widehat{\Omega}_T^-]\| &\le \frac{1}{T}\sum_{t=1}^T \|E[\omega_{t,\kappa}\omega_{t,\kappa}'\mathbf{1}(\|\omega_{t,\kappa}\| \le c_\kappa)]\| \\
&= \frac{1}{T}\sum_{t=1}^T \max_{\|\lambda\|=1} E\left[(\lambda'\omega_{t,\kappa})^2\mathbf{1}(\|\omega_{t,\kappa}\| \le c_\kappa)\right] \\
&\le \frac{1}{T}\sum_{t=1}^T \max_{\|\lambda\|=1} E\left[(\lambda'\omega_{t,\kappa})^2\right] \\
&= \frac{1}{T}\sum_{t=1}^T \|E[\omega_{t,\kappa}\omega_{t,\kappa}']\| \le \mu_\kappa.
\end{aligned}$$

The results collectively satisfies the conditions for Lemma 6.2 of Belloni et al. (2015) Hence, we conclude

$$E\|\widehat{\Omega}_T^- - E[\widehat{\Omega}_T^-]\| = O\left(\sqrt{\frac{\mu_\kappa c_\kappa^2 \log\kappa}{T}}\right) = O\left(\sqrt{\frac{\mu_\kappa \zeta^{\frac{2}{\nu-2}}\log\kappa}{T}}\right)$$

Secondly, we show that (2) is also bounded by the same rate. For convenience, let $W_t \equiv \omega_{t,\kappa}\omega_{t,\kappa}'\mathbf{1}(\|\omega_{t,\kappa}\| > c_\kappa) - E[\omega_{t,\kappa}\omega_{t,\kappa}'\mathbf{1}(\|\omega_{t,\kappa}\| > c_\kappa)]$. Then,

$$\begin{aligned}
E\|\widehat{\Omega}_T^+ - E[\widehat{\Omega}_T^+]\|^2 &= E\left[\mathrm{tr}\left((\widehat{\Omega}_T^+ - E[\widehat{\Omega}_T^+])(\widehat{\Omega}_T^+ - E[\widehat{\Omega}_T^+)')\right)\right] \\
&= \frac{1}{T^2}\sum_{t=1}^T E\left[\mathrm{tr}(W_tW_t')\right] + \frac{1}{T^2}\sum_{t\neq s} E\left[\mathrm{tr}(W_tW_s')\right] \\
&\le \frac{1}{T^2}\sum_{t=1}^T E\left[\|\omega_{t,\kappa}\|^4\mathbf{1}(\|\omega_{t,\kappa}\| > c_\kappa)\right] + o(1) && \text{(C.1.40)} \\
&\le \frac{1}{T^2}\sum_{t=1}^T \frac{E\|\omega_{t,\kappa}\|^\nu}{c_\kappa^{\nu-4}} && \text{(C.1.41)}
\end{aligned}$$

$$\leq \frac{1}{T}\frac{\zeta_\kappa}{c_\kappa^{\nu-4}}, \tag{C.1.42}$$

where each of the inequalities are justified by the following arguments: (C.1.40) is followed by the fact that $E[\text{tr}(W_t, W_s)] = \text{tr}(E[W_t W_s']) = \text{tr}(Cov(\omega_{t,\kappa}\omega_{t,\kappa}', \omega_{s,\kappa}\omega_{s,\kappa}'))$ and the result of Lemma 8 (in page 18), of which conditions are verified by (iii). (C.1.41) is straightforward by the Lemma 1 of Hansen (2015). And (C.1.42) is from the condition i). Then from the Liapunov's inequality, we have

$$E\|\widehat{\Omega}_T^+ - E[\widehat{\Omega}_T^+]\| \leq \left(E\|\widehat{\Omega}_T^+ - E[\widehat{\Omega}_T^+]\|^2\right)^{1/2} \leq \sqrt{\frac{1}{T}\frac{\zeta_\kappa}{c_\kappa^{\nu-4}}} = \sqrt{\frac{\zeta_\kappa^{\frac{2}{\nu-2}}}{T}}.$$

Therefore, the result follows. □

Lemma C.1.8. *Let* $\widehat{\Omega} = T^{-1}\sum_{t=1}^T \Psi^\kappa(z_t)\Psi^\kappa(z_t)'$, *then* $E\|\widehat{\Omega} - I_\kappa\| = O\left(\sqrt{\zeta_\kappa^{2/(\nu-2)}\log\kappa/T}\right)$.

Proof. The result is followed by Lemma C.1.7 with $\omega_{t,\kappa} = \Psi^\kappa(z_t)$ where $\Psi^\kappa(\cdot)$ is the $\kappa \times 1$ vector of basis functions satisfying $E\|\Psi^\kappa(z_t)\|^\nu = O(\zeta_\kappa)$ for $\nu > 4$ and $E[\Psi^\kappa(z_t)\Psi^\kappa(z_t)'] = I_\kappa$. □

Following lemmas help establish high-level assumptions to guarantee consistency result for the variance estimator defined as (3.18) and (3.19). Let $v_t = v_t(\chi_t, \alpha_0, p_0)$ and $\widehat{v}_t(\varphi) = \widehat{v}_t(\chi_t, \widehat{\alpha}, \varphi)$.

Lemma C.1.9. *i)* $\sup_t E\|v_t\|^{4+\delta} < \infty$ *for some* $\delta > 0$, *there exists a positive sequence* $\mu_{\kappa,T} \to 0$ *and* $\mu_{\kappa,T}^* \to 0$ *such that ii)* $E\|\widehat{v}_t(\varphi^*)\widehat{v}_{t-h}(\varphi^*)' - v_t v_{t-h}'\| = O(\mu_{\kappa,T})$ *and iii)* $\|\widehat{v}_t(\widehat{\varphi})\widehat{v}_{t-h}(\widehat{\varphi})' - \widehat{v}(\varphi^*)\widehat{v}_{t-h}(\varphi^*)'\| = O_p(\mu_{\kappa,T}^*\|\widehat{\varphi} - \varphi^*\|)$

C.1.2 Proofs

Proof of Theorem 3.1. The proof proceeds by verifying that the high level conditions in Newey (1994) are satisfied. In particular we show that the following conditions hold:
i) for p_0 denoting the true propensity score, there exists a functional $D(z,p)$ that is linear in p and a function $b(\xi_t)$ such that for $\|p - p_0\|_{2,P}$ small enough

$$\|h_t(p) - h_t(p_0) - D(\xi_t, p - p_0)\|_{2,P}^2 \leq b(\xi_t)\|p - p_0\|_{2,P}^2, \tag{C.1.43}$$

and

$$E[b(\xi_t)]\sqrt{T}\|p - p_0\|_{2,P}^2 \to 0, \tag{C.1.44}$$

ii) (Stochastic Equicontinuity): let P_0 be the true marginal distribution of ξ_t. Then it follows that

$$T^{-1/2}\sum_{t=1}^{T}\left(D\left(\xi_t,\hat{p}-p_0\right)-\int D\left(\xi_t,\hat{p}-p_0\right)dP_0\right)\rightarrow^p 0, \tag{C.1.45}$$

iii) There is a function $\gamma\left(\xi_t\right)$ sucht that

$$E\left[\gamma\left(\xi_t\right)\right]=0, \tag{C.1.46}$$

$$E\left[\left\|\gamma\left(\xi_t\right)\right\|^2\right]<\infty, \tag{C.1.47}$$

$$T^{1/2}\left(T^{-1}\sum_{t=1}^{T}\left(\gamma\left(\xi_t\right)-\int D\left(\xi,\hat{p}-p_0\right)dP_0\right)\right)\rightarrow^p 0. \tag{C.1.48}$$

It should be noted that (C.1.48) combines Newey (1994, Condition 5.3 (ii)) with the last part of Newey (1994, Condition 5.3 (i)). If (C.1.43)-(C.1.48) hold then it follows from Newey (1994, Lemma 5.1) that

$$T^{-1/2}\sum\nolimits_{t=1}^{T}\left(\hat{h}_t\left(\hat{p}\right)-\theta\left(\alpha_0\right)\right)=T^{-1/2}\sum\nolimits_{t=1}^{T}\left(m\left(\xi_t,\alpha,p\right)+\gamma\left(\xi_t\right)\right)+o_p\left(1\right). \tag{C.1.49}$$

We thus in turn show that our regularity conditions imply (C.1.43)-(C.1.48). For (C.1.43), note that from the derivation of D^j and suppressing the argument z_t in $g^j\left(z_t\right)$ and $p^j\left(z_t\right)$ for notational convenience we have for $g^j=p^j$ that

$$\begin{aligned}D^j\left(\xi_t,p-p_0\right)=\Bigg[&Y_t\left(\frac{D_{t,0}}{\left(p_0^0\right)^2}p_0^j\left(p^0-p_0^0\right)-\frac{D_{t,j}}{\left(p_0^j\right)^2}p_0^j\left(p^j-p_0^j\right)\right)\\&+Y_t\left(\frac{D_{t,j}}{p_0^j}-\frac{D_{t,0}}{p_0^0}\right)\left(p^j-p_0^j\right)\Bigg]\end{aligned} \tag{C.1.50}$$

and for $g^j=1$ that

$$D^j\left(\xi_t,p-p_0\right)=\left[Y_t\left(\frac{D_{t,0}}{\left(p_0^0\right)^2}\left(p^0-p_0^0\right)-\frac{D_{t,j}}{\left(p_0^j\right)^2}\left(p^j-p_0^j\right)\right)\right]. \tag{C.1.51}$$

Recall that $h_{t,j}\left(p^j,p^0\right)=Y_t\left(\frac{D_{t,j}}{p^j}-\frac{D_{t,0}}{p^0}\right)g^j$ such that it can be shown[1] for the case $g^j=p^j$ that

$$h_{t,j}\left(p^j,p^0\right)-h_{t,j}\left(p_0^j,p_0^0\right)-D^j\left(\xi_t,p-p_0\right)=-Y_tD_{t,0}\frac{p^j\left(p^0-p_0^0\right)^2}{\left(p_0^0\right)^2p^0} \tag{C.1.52}$$

such that

$$\left\|h_t\left(p\right)-h_t\left(p_0\right)-D\left(\xi_t,p-p_0\right)\right\|\leq\frac{\left\|Y_t\right\|}{\left(p_0^0\right)^2p^0}\left(p^0-p_0^0\right)^2. \tag{C.1.53}$$

[1] See Appendix C.2.

By the triangular inequality for $\mathbb{R}$ one obtains

$$\left|p_0^0\right| \leq \left|p^0\right| + \left|p^0 - p_0^0\right|$$

such that

$$\left|p_0^0\right| - \left|p^0 - p_0^0\right| \leq \left|p^0\right|$$

and using $|p_0^0| > \underline{p}$ and as long as $|p^0 - p_0^0| < \underline{p}$ one obtains

$$\frac{1}{|p^0|} \leq \frac{1}{\underline{p}} + \frac{|p^0 - p_0^0|}{\underline{p}\left(\underline{p} - |p^0 - p_0^0|\right)} \leq \frac{1}{\underline{p}} + \frac{|p^0 - p_0^0|}{\underline{p}^2} \leq \frac{1}{\underline{p}}\left(1 + \frac{2}{\underline{p}}\right)$$

and substituting in (C.1.53) leads to

$$\left\|h_t(p) - h_t(p_0) - D(\xi_t, p - p_0)\right\| \leq \frac{\|Y_t\|\left(\underline{p} + 2\right)}{\underline{p}^4}\left(p^0 - p_0^0\right)^2 \leq C\,\|Y_t\|\left|p^0 - p_0^0\right|$$

where $|p^0 - p_0^0| \leq 2$ was used and C is a generic constant. Squaring both sides of the inequalty and integrating with respect to $dP(z)$ leads to

$$\left\|h_t(p) - h_t(p_0) - D(\xi_t, p - p_0)\right\|_{2,P}^2 \leq \|Y_t\|^2 \int \left|p^0 - p_0^0\right|^2 dP(z) \leq \|Y_t\|^2 \left\|p^0 - p_0^0\right\|_{2,P}^2$$

where

$$\left\|p^0 - p_0^0\right\|_{2,P} = \left(\int \left|p^0 - p_0^0\right|^2 dP(z)\right)^{1/2} \leq \left(\sum_{j=0}^{J} \left|p^j - p_0^j\right|^2 dP(z)\right)^{1/2} = \|p - p_0\|_{2,P}\,. \tag{C.1.54}$$

When $g^j = 1$ the above calculations are

$$\begin{aligned} h_{t,j}\left(p^j, p^0\right) - h_{t,j}\left(p_0^j, p_0^0\right) - D^j\left(\xi_t, p - p_0\right) \\ = Y_t\left(\frac{D_{t,j}}{\left(p_0^j\right)^2 p^j}\left(p^j - p_0^j\right)^2 - \frac{D_{t,0}}{\left(p_0^0\right)^2 p^0}\left(p^0 - p_0^0\right)^2\right) \end{aligned}$$

such that

$$\left\|h_t(p) - h_t(p_0) - D(\xi_t, p - p_0)\right\|_{2,P}^2 \leq \frac{4\,\|Y_t\|^2\left(\underline{p} + 2\right)^2}{\underline{p}^8}\,\|p - p_0\|_{2,P}^2$$

by the same argument as before. Next note that $E\left[\|Y_t\|^2\right] \leq E\left[\|\xi_t\|^2\right] < \infty$ by Condition 3.2. This, together with the assumptions in Condition 3.3(ii) that $\sqrt{T}\,\|p - p_0\|_{2,P}^2 = o_p(1)$ establishes (C.1.43) and (C.1.44).

To show that (C.1.45) holds consider the case $g^j = p^j$. Let $D(\xi_t, p) = f(\xi_t, z_t)$ where the j-th component of f is

$$f^j(\xi_t, z_t) = Y_t \frac{D_{t,0}}{\left(p_0^0\right)^2} g_0^j p^0 + Y_t\left(-\frac{D_{t,0}}{p_0^0}\right) p^j \tag{C.1.55}$$

and where $f(\xi_t)$ is a class of functions indexed by $p \in \mathcal{P}$. Similarly, define $D(\xi_t, p') = g(\xi_t, z_t)$ for some $p' \neq p$. Then, because of linearity $D(\xi_t, p - p') = f(\xi_t, z_t) - g(\xi_t, z_t)$. For generic $\xi = (x, y, D)$ and noting that $|p^j|$, $|D_{t,j}|$ and $|g^j|$ are all bounded by 1,it follows that

$$\left\| f^j(\xi, z) \right\| \leq 2\frac{\|Y\|}{\underline{p}} = F(\xi) \tag{C.1.56}$$

where such that $F(\xi)$ defined in (C.1.56) is the envelope of $f(\xi, z)$. By Condition 3.2 it follows that $E\left[F(\xi_t)^{2+\delta}\right] < \infty$. To show stochastic equicontinuity of $f^j(\xi_t)$ it is enough to check that $f^j(\xi_t)$ satisfies the conditions of Theorem 3 and Corollary 10 in Kuersteiner (2016), which build on Doukhan et al. (1995). The theorem implies that f^j is Donsker and thus establishes asymptotic stochastic equicontinuity. To check the conditions of Corollary 10, note that the β-mixing sequence satisfies the summability requirement $\sum_{m=0}^{\infty} m^{1/(r-1)}\beta_m < \infty$ with $r > p/2$ by Condition 3.2. Further, $f(\xi_t) \in C^s\left(\mathbb{R}^d, \vartheta\right)$ with $\vartheta < -1$ by Condition 3.4(i), $\|\chi_t\|_{p(s-\vartheta),P} < \infty$ by Condition 3.4(ii). It now follows from Theorem 3 and Corollary 10 in Kuersteiner (2016) that for T large enough, any $\eta > 0$ and any $\delta > 0$ there exists an $\epsilon > 0$ such that

$$\Pr\left(\sup_{\|f-g\|_{2,\beta}\leq\epsilon}\left\|T^{-1/2}\sum_{t=1}^{T}\left(f(\xi_t) - g(\xi_t) - \int (f(\xi_t) - g(\xi_t))\,dP_0\right)\right\| > \delta\right) \leq \delta. \tag{C.1.57}$$

Using the basic inequality

$$\begin{aligned}
&\Pr\left(\left\|T^{-1/2}\sum_{t=1}^{T}\left(D(\xi_t, \hat{p} - p_0) - \int D(\xi_t, \hat{p} - p_0)\,dP_0\right)\right\| > \delta\right) \qquad \text{(C.1.58)}\\
&\leq \Pr\left(\sup_{\|f-g\|_{2,\beta}\leq\epsilon}\left\|T^{-1/2}\sum_{t=1}^{T}\left(f(\xi_t) - g(\xi_t) - \int (f(\xi_t) - g(\xi_t))\,dP_0\right)\right\| > \delta\right)\\
&+ \Pr\left(\hat{p} \in \left\{p : \|D(\xi_t, p - p_0)\|_{2,\beta} > \epsilon\right\}\right)
\end{aligned}$$

then establishes (C.1.45) if we can show that the second term is small. Note that from (C.1.55) it follows that

$$D^j(\xi_t, \hat{p} - p_0) = Y_t\frac{D_{t,0}}{(p_0^0)^2}g_0^j\left(\hat{p}^0 - p_0^0\right) + Y_t\left(-\frac{D_{t,0}}{p_0^0}\right)\left(\hat{p}^j - p_0^j\right).$$

By Lemma C.1.2 it follows that

$$\|D(\xi_t, \hat{p} - p_0)\|_{2,\beta} \leq C\,\|D(\xi_t, \hat{p} - p_0)\|_{2+\delta,P}$$

where $C = \sqrt{1 + \int_0^1 \phi^*\left(\beta^{-1}(u)\right)du}$ which is bounded as long as $\sum_{n>0} n^{1/\delta}\beta_n < \infty$. Then,

$$\|D(\xi_t, \hat{p} - p_0)\|_{2,\beta} \tag{C.1.59}$$

$$\leq C \left\| D\left(\xi_t, \hat{p} - p_0\right) \right\|_{2+\delta,P}$$
$$\leq 2C \frac{\left\| Y_t \right\|_{4+2\delta,P} \left\| \hat{p} - p_0 \right\|_{4+2\delta,P}}{\underline{p}}$$

where $\left\| \hat{p} - p_0 \right\|_{4+2\delta,P} = o(1)$ Condition 3.3(ii). Since the RHS of the last inequality in (C.1.59) tends to zero it follows that

$$\Pr\left(\hat{p} \in \left\{p : \left\| D\left(\xi_t, p - p_0\right) \right\|_{2,\beta} > \epsilon\right\}\right) \to 0.$$

It then follows that (C.1.45) holds by combining (C.1.58) and (C.1.59).

For (C.1.46) note that

$$\begin{aligned}\gamma^j\left(\xi_t, \alpha, p\right) = &-\frac{E\left[Y_t\left(d_j\right)|z_t\right]}{p^j\left(z_t\right)} g^j\left(z_t\right)\left(D_{t,j} - p^j\left(z_t\right)\right) \\ &+ E\left[Y_t\left(d_j\right) - Y_t\left(d_0\right)|z_t\right]\left(D_{t,j} - p^j\left(z_t\right)\right) \\ &+ \frac{E\left[Y_t\left(d_0\right)|z_t\right]}{p^0\left(z_t\right)} g^j\left(z_t\right)\left(D_{t,0} - p^0\left(z_t\right)\right).\end{aligned}$$

such that $E\left(\gamma^j\left(\xi_t, \alpha, p\right)|z_t\right) = 0$ follows immediately because $E\left[D_{t,j}|z_t\right] = p^j\left(z_t\right)$ by definition.

For (C.1.47) note that

$$\left\|\gamma^j\left(\xi_t, \alpha, p\right)\right\| \leq 4\left(\left\|\frac{E\left[Y_t\left(d_j\right)|z_t\right]}{p^j\left(z_t\right)}\right\| + \left\|\frac{E\left[Y_t\left(d_0\right)|z_t\right]}{p^j\left(z_t\right)}\right\|\right) \leq 8\frac{E\left[\left\|Y_t\right\||z_t\right]}{p^j\left(z_t\right)} \quad \text{(C.1.60)}$$

where

$$\left\|E\left[Y_t\left(d_j\right)|z_t\right]\right\| = \left\|E\left[Y_t D_{t,j}|z_t\right]\right\| \leq E\left[\left\|Y_t\right\||z_t\right]$$

was used. Then, using the conditional Jensen's inequality and (C.1.60) we obtain

$$E\left[\left\|\gamma^j\left(\xi_t, \alpha, p\right)\right\|^2\right] \leq 16 E\left[\frac{\left(E\left[\left\|Y_t\right\||z_t\right]\right)^2}{p^j\left(z_t\right)^2}\right] \leq 16\frac{E\left[\left\|Y_t\right\|^2\right]}{\underline{p}^2}$$

and $\left\|\gamma\left(\xi_t, \alpha, p\right)\right\|^2 = \sum_{j=1}^J \left\|\gamma^j\left(\xi_t, \alpha, p\right)\right\|^2$ such that using Condition 3.2

$$E\left[\left\|\gamma\left(\xi_t, \alpha, p\right)\right\|^2\right] \leq 64J\frac{E\left[\left\|Y_t\right\|^2\right]}{\underline{p}^2} < \infty$$

as had to be shown.

Finally, for (C.1.48) consider (C.1.50) where

$$\begin{aligned}E\left[D^j\left(\xi_t, \hat{p} - p_0\right)|z_t\right] = &\frac{E\left[Y_t\left(d_0\right)|z_t\right] g_0^j}{p_0^0}\left(\hat{p}^0 - p_0^0\right) - \frac{E\left[Y_t\left(d_j\right)|z_t\right] g_0^j}{p_0^j}\left(\hat{p}^j - p_0^j\right) \\ &+ \left(E\left[Y_t\left(d_j\right)|z_t\right] - E\left[Y_t\left(d_0\right)|z_t\right]\right)\left(\hat{p}^j - p_0^j\right).\end{aligned}$$

Stack $D(\xi_t, \hat{p}-p_0) = \left(D^1(\xi_t, \hat{p}-p_0), ..., D^J(\xi_t, \hat{p}-p_0)\right)'$ such that for $b(z_t), b^0(z_t)$ and $\mathbf{e}_1$ defined in Lemma C.1.6 one can write

$$D(\xi_t, \hat{p}-p_0) = \left(b(z_t) - b^0(z_t)\mathbf{1}'\right)(\hat{p}-p_0).$$

In the same way let $\gamma(\xi_t) = \left(\gamma^1(\xi_t), ..., \gamma^J(\xi_t)\right)$ and recall that for $\mathcal{D}_t = (D_{t,1}, \cdots, D_{t,J})$, $\gamma(\xi_t) = (b(z_t) - b^0(z_t)\mathbf{1}')(\mathcal{D}_t - p_0(z_t))$. It then follows that for $B(z) = b(z) - b^0(z)\mathbf{1}'$, we can write

$$\begin{aligned} &T^{1/2}\left(T^{-1}\sum_{t=1}^{T}\gamma(\xi_t) - \int D(\xi, \hat{p}-p_0)\,dP_0\right) && \text{(C.1.61)} \\ &= T^{1/2}\left(T^{-1}\sum_{t=1}^{T}B(z_t)(\mathcal{D}_t - p_0(z_t)) - \int B(z)(\hat{p}(z) - p_0(z))\,dP_0(z)\right). \end{aligned}$$

Use the mean value expansion

$$\hat{p}^j - p^j_\kappa = p^j_\kappa(z_t, \hat{\varphi}_\kappa) - p^j_\kappa(z_t, \varphi^*_\kappa) = \frac{\partial p^j_\kappa(z_t, \bar{\varphi}_\kappa)}{\partial \varphi'_{j,\kappa}}\left(\hat{\varphi}_{j,\kappa} - \varphi^*_{j,\kappa}\right)$$

where $\varphi_\kappa = \left(\varphi'_{1,\kappa}, ..., \varphi'_{J,\kappa}\right)'$ and $p^j_\kappa(z_t, \varphi_\kappa)$ only depends on $\varphi_{j,\kappa}$ and where $\left\|\bar{\varphi}_{j,\kappa} - \varphi^*_{j,\kappa}\right\| \leq \left\|\hat{\varphi}_{j,\kappa} - \varphi^*_{j,\kappa}\right\|$. Defining

$$\dot{p}_\kappa(z_t, \varphi_\kappa)' = \frac{\partial p_\kappa(z_t, \varphi_\kappa)}{\partial \varphi'_\kappa} \tag{C.1.62}$$

and denoting the remainder term by $r(\hat{\varphi}_\kappa - \varphi^*_\kappa)$, a $J \times 1$ vector, one obtains

$$\hat{p} - p_\kappa = \dot{p}_\kappa(z_t, \bar{\varphi}_\kappa)'(\hat{\varphi}_\kappa - \varphi^*_\kappa). \tag{C.1.63}$$

Next, letting

$$\frac{1}{T}\frac{\partial(L_{T,\kappa}(\varphi_\kappa)/T)}{\partial \varphi_\kappa} = s_T(\varphi_\kappa),\ \frac{1}{T}\frac{\partial^2(L_{T,\kappa}(\varphi_\kappa))}{\partial \varphi_\kappa \partial \varphi'_\kappa} = H_T(\varphi_\kappa)$$

and use a mean value expansion of the likelihood to obtain

$$(\hat{\varphi}_\kappa - \varphi^*_\kappa) = H_T(\bar{\varphi}_\kappa)^{-1} s_T(\varphi^*_\kappa) \tag{C.1.64}$$

where $\|\bar{\varphi}_\kappa - \varphi^*_\kappa\| \leq \|\hat{\varphi}_\kappa - \varphi^*_\kappa\|$. Let $P_\kappa(z_t, \varphi)$ and $Q_\kappa(z_t, \varphi_\kappa)$ be as defined in Lemma C.1.5. Simple algebra then shows that

$$\dot{p}_\kappa(z_t, \varphi_\kappa) Q_\kappa(z_t, \varphi_\kappa)^{-1}(\mathcal{D}_t - p_\kappa(z_t, \varphi_\kappa)) = \partial l(D_t, z_{t,}\varphi_k)/\partial \varphi_k. \tag{C.1.65}$$

Now consider

$$\begin{aligned} &\int B(z)(\hat{p}(z) - p_0(z))\,dP_0(z) \\ &= \int B(z)(p_\kappa(z) - p_0(z))\,dP_0(z) + \int B(z)\,\dot{p}_\kappa(z, \bar{\varphi}_\kappa)'\,dP_0(\hat{\varphi}_\kappa - \varphi^*_\kappa) && \text{(C.1.66)} \end{aligned}$$

such that substituting from (C.1.63), (C.1.64) and (C.1.65) it follows that

$$\int B(z)\,\dot{p}_\kappa(z,\bar{\varphi}_\kappa)'\,dP_0(\hat{\varphi}_\kappa-\varphi_\kappa^*)=\int B(z)\,\dot{p}_\kappa(z,\bar{\varphi}_\kappa)'\,dP_0H_T(\bar{\varphi}_\kappa)^{-1}\,s_T(\varphi_\kappa^*)\,. \tag{C.1.67}$$

Now considering (C.1.67), substituting from (C.1.65) one has

$$\int B(z)\,\dot{p}_\kappa(z,\bar{\varphi}_\kappa)'\,dP_0H_T(\bar{\varphi}_\kappa)^{-1}\,s_T(\varphi_\kappa^*) \tag{C.1.68}$$

$$=T^{-1}\sum_{t=1}^{T}\int B(z)\,\dot{p}_\kappa(z,\bar{\varphi}_\kappa)'\,dP_0H_T(\bar{\varphi}_\kappa)^{-1}\,\dot{p}_\kappa(z_t,\varphi_\kappa^*)\,Q_\kappa(z,\varphi_\kappa^*)^{-1}(\mathcal{D}_t-p(z_t,\varphi_\kappa^*))$$

$$=T^{-1}\sum_{t=1}^{T\prime}\widehat{\mathcal{K}}_\kappa(\bar{\varphi}_\kappa)\,\dot{p}_\kappa(z_t,\varphi_\kappa^*)\,Q_\kappa(z_t,\varphi_\kappa^*)^{-1}(\mathcal{D}_t-p(z_t,\varphi_\kappa^*))$$

where $\widehat{\mathcal{K}}_\kappa(\bar{\varphi}_\kappa)=\int B(z)\,\dot{p}_\kappa(z,\bar{\varphi}_\kappa)'\,dP_0H_T(\bar{\varphi}_\kappa)^{-1}$. Let

$$\mathcal{K}_\kappa(\varphi_\kappa^*)=\int B(z)\,\dot{p}_\kappa(z,\varphi_\kappa^*)'\,dP_0H(\varphi_\kappa^*)^{-1}\,.$$

To bound (C.1.68), first consider

$$T^{-1/2}\sum_{t=1}^{T}\widehat{\mathcal{K}}_\kappa(\bar{\varphi}_\kappa)\,\dot{p}_\kappa(z_t,\varphi_\kappa^*)\,Q_\kappa(z_t,\varphi_\kappa^*)^{-1}(\mathcal{D}_t-p(z_t,\varphi_\kappa^*)) \tag{C.1.69}$$

$$=\left(\widehat{\mathcal{K}}_\kappa(\bar{\varphi}_\kappa)-\mathcal{K}_\kappa(\varphi_\kappa^*)\right)T^{-1/2}\sum_{t=1}^{T}\dot{p}_\kappa(z_t,\varphi_\kappa^*)\,Q_\kappa(z_t,\varphi_\kappa^*)^{-1}(\mathcal{D}_t-p(z_t,\varphi)) \tag{C.1.70}$$

$$+\widehat{\mathcal{K}}_\kappa(\bar{\varphi}_\kappa)\,T^{-1/2}\sum_{t=1}^{T}\dot{p}_\kappa(z_t,\varphi_\kappa^*)\,Q_\kappa(z_t,\varphi_\kappa^*)^{-1}(p(z_t,\varphi)-p_\kappa(z_t,\varphi_\kappa^*)) \tag{C.1.71}$$

$$\begin{aligned}&+T^{-1/2}\sum_{t=1}^{T}\left(\mathcal{K}_\kappa(\varphi_\kappa^*)\,\dot{p}_\kappa(z_t,\varphi_\kappa^*)\,Q_\kappa(z_t,\varphi_\kappa^*)^{-1/2}-B(z_t)\,Q_\kappa(z_t,\varphi_\kappa^*)^{1/2}\right)\\&\qquad\times Q_\kappa(z_t,\varphi_\kappa^*)^{-1/2}(\mathcal{D}_t-p(z_t,\varphi))\end{aligned} \tag{C.1.72}$$

$$+T^{-1/2}\sum_{t=1}^{T}B(z_t)\,(\mathcal{D}_t-p(z_t,\varphi))\,. \tag{C.1.73}$$

By Lemma C.1.6(i)-(iv) it follows that (C.1.69)-(C.1.72) are $o_p(1)$. It follows from (C.1.61), (C.1.66), (C.1.67) and (C.1.68) that

$$\begin{aligned}&T^{1/2}\left(T^{-1}\sum_{t=1}^{T}\gamma(\xi_t)-\int D(\xi,\hat{p}-p_0)\,dP_0\right)\\&\qquad=\int B(z)\,(p_\kappa(z)-p_0(z))\,dP_0(z)+o_p(1)=o_p(1)\end{aligned}$$

where the last line follows from Lemma C.1.6(v)-(vii). This establishes (C.1.48). □

Proof of Theorem 3.2. The proof follows Hirano et al. (2003) and Cattaneo (2010, Theorem B-1) with the necessary adjustments. First note that by (C.1.18)

$$
\begin{aligned}
&E\left(\|p\left(z_t,\varphi\right)-p_\kappa(z_t,\varphi_\kappa)\|\right)\\
&\qquad\le C_0 E\left(\left\|\Gamma^{-1}\left(p\left(z_t,\varphi\right)\right)-\Psi_\kappa\left(z_t,\varphi_\kappa\right)\right\|^2\right)^{1/2}=O\left(\kappa^{-\alpha}\right). \quad \text{(C.1.74)}
\end{aligned}
$$

As shown in Cattaneo (2010), $Q_\kappa\left(z_t,\varphi_k\right)\ge\prod_{j=1}^{J}p_\kappa^j(z_t,\varphi_\kappa)I_J$ where the inequality is in the sense of positive semidefiniteness. If e_j is the j-th unit vector in $\mathbb{R}^J$ then $e_j'Q_\kappa\left(z_t,\varphi_k\right)e_j\ge\prod_{j=1}^{J}p_\kappa^j(z_t,\varphi_\kappa)$ which implies $\operatorname{tr}Q_\kappa\left(z_t,\varphi_k\right)\ge J\cdot\prod_{j=1}^{J}p_\kappa^j(z_t,\varphi_\kappa)>0$. It is easy to see that $\operatorname{tr}Q_\kappa\left(z_t,\varphi_k\right)<J$. One also obtains

$$
\prod_{j=1}^{J}p^j(z_t)-\left|\prod_{j=1}^{J}p^j(z_t)-\prod_{j=1}^{J}p_\kappa^j(z_t,\varphi_\kappa)\right|\le\prod_{j=1}^{J}p_\kappa^j(z_t,\varphi_\kappa) \quad \text{(C.1.75)}
$$

such that

$$
\begin{aligned}
E\left[Q_\kappa\left(z_t,\varphi_k\right)\right]&\ge E\left[\prod_{j=1}^{J}p_\kappa^j(z_t,\varphi_\kappa)\right]I_J\\
&\ge E\left[\prod_{j=1}^{J}p^j(z_t)\right]I_J-E\left[\left|\prod_{j=1}^{J}p^j(z_t)-\prod_{j=1}^{J}p_\kappa^j(z_t,\varphi_\kappa)\right|\right]I_J \quad \text{(C.1.76)}\\
&\ge E\left[p(z_t)^J\right]I_J-E\left[\left|\prod_{j=1}^{J}p^j(z_t)-\prod_{j=1}^{J}p_\kappa^j(z_t,\varphi_\kappa)\right|\right]I_J\\
&\ge E\left[p(z_t)^J\right]I_J-o\left(\kappa^{-\alpha}\right)\ge\underline{p}^J I_J-o\left(\kappa^{-\alpha}\right)
\end{aligned}
$$

By Lemma C.1.8 it follows that for $\check{\Omega}=T^{-1}\sum_{t=1}^{T}\check{\Psi}^\kappa\left(z_t\right)\check{\Psi}^\kappa\left(z_t\right)'$, $\left\|\check{\Omega}-I\right\|=O\left(\kappa/\sqrt{T}\right)$ which in turn implies

$$
\lambda_{\min}\check{\Omega}>1/2 \text{ with probability approaching one (wpa1).} \quad \text{(C.1.77)}
$$

Now

$$
\begin{aligned}
E\left[\left\|T^{-1}\sum_{t=1}^{T}\partial l\left(D_t,z_{t,}\varphi_k^*\right)/\partial\varphi_k\right\|\right]&=E\left[\left\|T^{-1}\sum_{t=1}^{T}\left(\mathcal{D}_t-p(z_t)\right)'\otimes\Psi^\kappa\left(z_t\right)\right\|^2\right]^{1/2} \quad \text{(C.1.78)}\\
&\quad+T^{-1}\sum_{t=1}^{T}E\left[\left\|\left(p(z_t)-p_\kappa\left(z_t,\varphi_\kappa^*\right)\right)'\right\|\left\|\Psi^\kappa\left(z_t\right)\right\|\right]
\end{aligned}
$$

where

$$
\begin{aligned}
&E\left[\left\|T^{-1}\sum_{t=1}^{T}\left(\mathcal{D}_t-p(z_t)\right)'\otimes\Psi^\kappa\left(z_t\right)\right\|^2\right] \quad \text{(C.1.79)}\\
&=T^{-2}\sum_{t=1}^{T}\operatorname{tr}E\left[Q\left(z_t\right)\otimes\Psi^\kappa\left(z_t\right)\Psi^\kappa\left(z_t\right)'\right]\\
&=T^{-2}\sum_{t=1}^{T}E\left[\left(\operatorname{tr}Q\left(z_t\right)\right)\left\|\Psi^\kappa\left(z_t\right)\right\|^2\right] \quad \text{(C.1.80)}
\end{aligned}
$$

$$\le JT^{-2}\sum_{t=1}^{T}E\left[\left\|\Psi^{\kappa}\left(z_t\right)\right\|^2\right]$$
$$=O\left(T^{-1}\zeta\left(\kappa\right)^2\right)$$

where $0<\operatorname{tr}Q_{\kappa}\left(z_t,\varphi_k\right)<J$ was used. For the second term note that

$$E\left[\left\|\left(p(z_t)-p_{\kappa}\left(z_t,\varphi_{\kappa}\right)\right)'\right\|\left\|\Psi^{\kappa}\left(z_t\right)\right\|\right]$$
$$\le\left(E\left[\left\|\left(p(z_t)-p_{\kappa}\left(z_t,\varphi_{\kappa}\right)\right)'\right\|^2\right]E\left[\left\|\Psi^{\kappa}\left(z_t\right)\right\|^2\right]\right)^{1/2}=O\left(\kappa^{-\alpha}\zeta\left(\kappa\right)\right)\quad\text{(C.1.81)}$$

It follows that

$$T^{-1}\textstyle\sum_{t=1}^{T}\partial l\left(D_t,z_{t,}\varphi_k^*\right)/\partial\varphi_k=O_p\left(T^{-1/2}\zeta\left(\kappa\right)+\kappa^{-\alpha}\zeta\left(\kappa\right)\right).\qquad\text{(C.1.82)}$$

For the δ in (C.1.25) it follows from (C.1.82) that for any $\varepsilon>0$ there is a $C<\infty$ such that for T large enough

$$\Pr\left(\left\|T^{-1}\textstyle\sum_{t=1}^{T}\partial l\left(D_t,z_{t,}\varphi_k^*\right)/\partial\varphi_k\right\|<\frac{\delta}{2}C\left(T^{-1/2}\zeta\left(\kappa\right)+\kappa^{-\alpha}\zeta\left(\kappa\right)\right)\right)\le1-\frac{\varepsilon}{2}\qquad\text{(C.1.83)}$$

Next, keeping C and ε fixed as in (C.1.83) and adapting the argument in Hirano et al. (2003, p. 1180) use a mean value expansion around φ_{κ}^* to obtain

$$\sup_{\left\|\varphi_{\kappa}-\varphi_{\kappa}^*\right\|\le C\left(\frac{\zeta(\kappa)}{\sqrt{T}}+\kappa^{-\alpha}\zeta(\kappa)\right)}E\left[\left\|p_{\kappa}\left(z_t,\varphi_{\kappa}\right)-p_{\kappa}\left(z_t,\varphi_{\kappa}^*\right)\right\|\right]$$
$$\le\sup_{\left\|\varphi_{\kappa}-\varphi_{\kappa}^*\right\|\le C\left(\frac{\zeta(\kappa)}{\sqrt{T}}+\kappa^{-\alpha}\zeta(\kappa)\right)}C_0E\left[\left\|\Psi_{\kappa}\left(z_t,\varphi_{\kappa}\right)-\Psi_{\kappa}\left(z_t,\varphi_{\kappa}^*\right)\right\|\right]$$
$$\le E\left[\left\|\Psi^{\kappa}\left(z_t\right)\right\|\right]C_0C\left(\frac{\zeta\left(\kappa\right)}{\sqrt{T}}+\kappa^{-\alpha}\zeta\left(\kappa\right)\right)$$
$$\le2JC_0C\zeta\left(\kappa\right)\left(\frac{\zeta\left(\kappa\right)}{\sqrt{T}}+\kappa^{-\alpha}\zeta\left(\kappa\right)\right)$$

where the second inequality uses $E\left[\left\|Q_{\kappa}\left(z_t,\bar{\varphi}_{\kappa}\right)\otimes\Psi^{\kappa}\left(z_t\right)'\right\|\right]\le E\left[\left\|Q_{\kappa}\left(z_t,\bar{\varphi}_{\kappa}\right)\right\|\left\|\Psi^{\kappa}\left(z_t\right)\right\|\right]$, $\left\|Q_{\kappa}\left(z_t,\bar{\varphi}_{\kappa}\right)\right\|\le2J$ by (C.1.2) and Condition 3.4. Using the second order expansion

$$T^{-1}\left(L_{T,\kappa}\left(\varphi_{\kappa}\right)-L_{T,\kappa}\left(\varphi_{\kappa}^*\right)\right)=s_{T,\kappa}\left(\varphi_{\kappa}^*\right)\left(\varphi_{\kappa}-\varphi_{\kappa}^*\right)+\frac{1}{2}\left(\varphi_{\kappa}-\varphi_{\kappa}^*\right)'H_{T,\kappa}\left(\bar{\varphi}_{\kappa}\right)\left(\varphi_{\kappa}-\varphi_{\kappa}^*\right)\qquad\text{(C.1.84)}$$

with $\left\|\bar{\varphi}_{\kappa}-\varphi_{\kappa}^*\right\|\le\left\|\varphi_{\kappa}-\varphi_{\kappa}^*\right\|$. By (C.1.6) and the same argument as in Hirano et al. (2003, p. 1181), we have

$$\frac{1}{2}H_{T,\kappa}\left(\bar{\varphi}_{\kappa}\right)=-\left(2T\right)^{-1}\sum_{t=1}^{T}Q_{\kappa}\left(z_t,\varphi_{\kappa}\right)\otimes\Psi^{\kappa}\left(z_t\right)\Psi^{\kappa}\left(z_t\right)'$$
$$\ge-2\delta\left(I\otimes T^{-1}\sum_{t=1}^{T}\Psi^{\kappa}\left(z_t\right)\Psi^{\kappa}\left(z_t\right)'\right).$$

such that the eigenvalues of $(1/2)H_{T,\kappa}(\bar{\varphi}_\kappa)$ are bounded away from zero in absolute value by δ because of (C.1.77). Then, again using the argument in Hirano et al. (2003, p. 1181), with probability greater than $1-\varepsilon$ and for $\|\varphi_\kappa - \varphi_\kappa^*\| \leq \frac{\zeta(\kappa)}{\sqrt{T}} + \kappa^{-\alpha}\zeta(\kappa)$ it follows after rearranging (C.1.84) that

$$\begin{aligned}
T^{-1}\left(L_{T,\kappa}(\varphi_\kappa) - L_{T,\kappa}(\varphi_\kappa^*)\right) &\leq s_{T,\kappa}(\varphi_\kappa^*)(\varphi_\kappa - \varphi_\kappa^*) - \delta\frac{1}{2}\|(\varphi_\kappa - \varphi_\kappa^*)\|^2 \\
&\leq \left(\|s_{T,\kappa}(\varphi_\kappa^*)\| - \frac{\delta}{2}C\left(\frac{\zeta(\kappa)}{\sqrt{T}} + \kappa^{-\alpha}\zeta(\kappa)\right)\right)\|\varphi_\kappa - \varphi_\kappa^*\| \\
&< 0 \text{ w.p. } 1-\varepsilon/2
\end{aligned}$$

such that the result in (3.16) follows from the argument in Hirano et al. (2003, p. 1181).

To establish (3.17) note that $p(z_t,\varphi) - p_\kappa(z_t,\varphi_\kappa) = \frac{\partial\Gamma}{\partial\Psi}\left(\Gamma^{-1}(p(z_t,\varphi)) - \Psi_\kappa(z_t,\varphi_\kappa)\right)$. For $p \leq 4+\delta$, consider

$$\begin{aligned}
&\int \|p(z) - p_\kappa(z,\hat{\varphi}_\kappa)\|^p \, dP_0(z) \\
&\qquad \leq 2^{p-1}\int \left(\|p(z) - p_\kappa(z,\varphi_\kappa^*)\|^p + \|p_\kappa(z,\varphi_\kappa^*) - p_\kappa(z,\hat{\varphi}_\kappa)\|^p\right) dP_0(z) \quad \text{(C.1.85)}
\end{aligned}$$

where for the first term in (C.1.85) we note that

$$\begin{aligned}
&\int \|p(z) - p_\kappa(z,\varphi_\kappa^*)\|^p \, dP_0(z) \\
&\leq \left\|\frac{\partial\Gamma}{\partial\Psi}\right\|^p \left(\sup_z \left\|\frac{\Gamma^{-1}(p(z_t,\varphi)) - \Psi_\kappa(z_t,\varphi_\kappa)}{\langle z\rangle^{s(2+\delta)/2}}\right\|\right)^p \int \langle z\rangle^{s(2+\delta)/2} \, dP_0(z) = o\left(\kappa^{-\alpha p}\right)
\end{aligned}$$

by Lemma C.1.3. For the second term in (C.1.85) use

$$\int \|p_\kappa(z,\varphi_\kappa^*) - p_\kappa(z,\hat{\varphi}_\kappa)\|^p \, dP_0(z) \leq \int \left\|\frac{\partial p_\kappa(z,\tilde{\varphi}_\kappa)}{\partial\varphi_\kappa}\right\|^p dP_0(z)\,\|(\hat{\varphi}_\kappa - \varphi_\kappa^*)\|^p$$

with

$$\left\|\frac{\partial p_\kappa(z,\tilde{\varphi}_\kappa)}{\partial\varphi_\kappa}\right\| \leq \|Q_\kappa(z_t,\varphi_\kappa)\|\,\|\Psi^\kappa(z_t)\|$$

such that from $\|Q_\kappa(z_t,\varphi_\kappa)\| \leq 2J$ it follows that

$$\begin{aligned}
\int \left\|\frac{\partial p_\kappa(z,\tilde{\varphi}_\kappa)}{\partial\varphi_\kappa}\right\|^p dP_0(z)\,\|\hat{\varphi}_\kappa - \varphi_\kappa^*\|^p &\leq (2J)^p E\left[\|\Psi^\kappa(z_t)\|^p\right] O_p\left(\left(\frac{\zeta(\kappa)}{\sqrt{T}} + \kappa^{-\alpha}\zeta(\kappa)\right)^p\right) \\
&= \zeta(\kappa)\, O_p\left(\left(\frac{\zeta(\kappa)}{\sqrt{T}} + \kappa^{-\alpha}\zeta(\kappa)\right)^p\right)
\end{aligned}$$

by Condition 3.4(ii) as long as $p \leq 4+\delta$. It now follows that

$$\int \|p(z) - p_\kappa(z,\hat{\varphi}_\kappa)\|^p \, dP_0(z) = \zeta(\kappa)\, O_p\left(\left(\frac{\zeta(\kappa)}{\sqrt{T}} + \kappa^{-\alpha}\zeta(\kappa)\right)^p\right) + O\left(\kappa^{-\alpha p}\right)$$

$$= O_p\left(\frac{\zeta(\kappa)^{p+1}}{T^{p/2}} + \kappa^{-\alpha p}\zeta(\kappa)^{p+1}\right)$$

where $\kappa^{-\alpha}\zeta(\kappa)/\sqrt{T} = o\left(\zeta(\kappa)^2/T\right)$ because $\zeta(\kappa)^2\kappa/T \to 0$ such that $\zeta(\kappa)^2/T = o(\kappa^{-1})$. $\square$

Proof of Theorem 3.3. Throughout the proof, $\overline{C}$ implies a generic constant. Without the loss of generality, consider the case $E[v_t(\chi_t, \alpha_0, p_0)] = 0$. By triangular inequality,

$$\|\widehat{V} - V\| \leq \sum_{h=-B}^{B}\left|K\left(\frac{h}{B}\right)\right|\left\|\widehat{\Omega}_h - \frac{1}{T}\sum_{t=h+1}^{T}\widehat{v}_t(\varphi^*)\widehat{v}_{t-h}(\varphi^*)'\right\| \tag{C.1.86}$$

$$+ \sum_{h=-B}^{B}\left|K\left(\frac{h}{B}\right)\right|\left\|\frac{1}{T}\sum_{t=h+1}^{T}\widehat{v}_t(\varphi^*)\widehat{v}_{t-h}(\varphi^*)' - E[\widehat{v}_t(\varphi^*)\widehat{v}_{t-h}(\varphi^*)']\right\| \tag{C.1.87}$$

$$+ \sum_{h=-B}^{B}\left\|K\left(\frac{h}{B}\right)E[\widehat{v}_t(\varphi^*)\widehat{v}_{t-h}(\varphi^*)'] - E[v_t v_{t-h}']\right\| \tag{C.1.88}$$

$$+ \sum_{|h|>B}\|E[v_t v_{t-h}']\| \tag{C.1.89}$$

Consider term (C.1.89). By Condition 3.5 and inequality bound for mixing process, $\|E[v_t v_{t-h}']\| \leq \overline{C}\sup_t(E\|v_t\|^p)^{1/p}\beta_h^{1-2/p} \leq \overline{C}\beta_h^{1-2/p}$. Then for sufficiently large B, we get $\sum_{|h|>B}\|E[v_t v_{t-h}']\| \leq \overline{C}\sum_{h=B}^{\infty}\beta_h^{1-2/p} = o(1)$.

For the term (C.1.88), note that

$$\sum_{h=-B}^{B}\left\|K\left(\frac{h}{B}\right)E[\widehat{v}_t(\varphi^*)\widehat{v}_{t-h}(\varphi^*)'] - E[v_t v_{t-h}']\right\|$$

$$\leq \sum_{h=-B}^{B}\left|K\left(\frac{h}{B}\right)\right|E\|\widehat{v}_t(\varphi^*)\widehat{v}_{t-h}(\varphi^*)' - v_t v_{t-h}'\|$$

$$+ \sum_{h=-B}^{B}\left|K\left(\frac{h}{B}\right) - 1\right|\|E[v_t v_{t-h}']\| \tag{C.1.90}$$

For the second term in (C.1.90), we have

$$\sum_{h=-B}^{B}\left|K\left(\frac{h}{B}\right) - 1\right|\|E[v_t v_{t-h}']\| \leq 2\sup_t(E\|v_t\|^p)^{1/p}\sum_{h=0}^{B}\left|K\left(\frac{h}{B}\right) - 1\right|\beta_h^{1-2/p}$$

$$\leq 2\overline{C}\sum_{h=0}^{B}\left|K\left(\frac{h}{B}\right) - 1\right|\beta_h^{1-2/p} = o(1)$$

while the first step followed by mixing inequality and second from Condition 3.5, Condition 3.6 i), ii). For the first term,

$$\sum_{h=-B}^{B}\left|K\left(\frac{h}{B}\right)\right| E\|\widehat{v}_t(\varphi^*)\widehat{v}_{t-h}(\varphi^*)'-v_tv_{t-h}'\| \leq O(B_T\mu_{\kappa,T})\sum_{h=-B_T}^{B_T}\left|K\left(\frac{h}{B_T}\right)\right|\frac{1}{B_T}$$
$$\leq O(B_T\mu_{\kappa,T})\int|K(u)|du = o(1)$$

which is followed by Condition 3.6 and the result ii) of Lemma C.1.9.

Finally, consider the term (C.1.86). Notice that for each h,

$$\left\|\widehat{\Omega}_h - \frac{1}{T}\sum_{t=h+1}^{T}\widehat{v}_t(\varphi^*)\widehat{v}_{t-h}(\varphi^*)'\right\| \leq \frac{1}{T}\sum_{t=h+1}^{T}\|\widehat{v}_t(\widehat{\varphi})\widehat{v}_{t-h}(\widehat{\varphi})'-\widehat{v}_t(\varphi^*)\widehat{v}_{t-h}(\varphi^*)'\|$$
$$= O_p(\mu_{\kappa,T}^*\|\widehat{\varphi}-\varphi^*\|)$$

followed by the result iii) of Lemma C.1.9. Therefore,

$$\sum_{h=-B}^{B}\left|K\left(\frac{h}{B}\right)\right|\left\|\widehat{\Omega}_h - \frac{1}{T}\sum_{t=h+1}^{T}\widehat{v}_t(\varphi^*)\widehat{v}_{t-h}(\varphi^*)'\right\|$$
$$\leq O_p(B_T\mu_{\kappa,T}^*\|\widehat{\varphi}-\varphi^*\|)\sum_{h=-B_T}^{B_T}\left|K\left(\frac{h}{B_T}\right)\right|\frac{1}{B_T}$$
$$\leq O_p(B_T\mu_{\kappa,T}^*\|\widehat{\varphi}-\varphi^*\|)\int|K(u)|du = o_p(1)$$

which completes the proof of $\|\widehat{V}-V\| = o_p(1)$.

To show the second part of the Theorem 3.3, it is sufficient to show $\|\widehat{V}-\tilde{V}\| = o_p(1)$. For each h,

$$\widehat{\Omega}_h - \tilde{\Omega}_h = \frac{1}{T}\sum_{t=h+1}^{T}\left(\widehat{v}_t\widehat{v}_{t-h}' - (\widehat{v}_t-\overline{v}_T)(\widehat{v}_{t-h}-\overline{v}_T)'\right)$$
$$= \left(\frac{1}{T}\sum_{t=h+1}^{T}\widehat{v}_t\right)\overline{v}_T' + \overline{v}_T\left(\frac{1}{T}\sum_{t=1}^{T-h}\widehat{v}_t\right)' - \overline{v}_T\overline{v}_T'$$

Note that by Theorem 3.1, $\sum_{t=1}^{T}\widehat{v}_t = O_p(\sqrt{T})$. This implies that $\|\widehat{\Omega}_h-\tilde{\Omega}_h\| = O_p(T^{-1})$. Therefore,

$$\|\widehat{V}-\tilde{V}\| \leq \sum_{h=-B_T}^{B_T}\left|K\left(\frac{h}{B_T}\right)\right|\|\widehat{\Omega}_h-\tilde{\Omega}_h\|$$
$$\leq O_p(B_TT^{-1})\sum_{h=-B_T}^{B_T}\left|K\left(\frac{h}{B_T}\right)\right|\frac{1}{B_T}$$

$$\leq O_p(B_T T^{-1}) \int |K(u)| du = o_p(1)$$

For $\tilde{V}$, it suffices to show that $\|\tilde{V} - \widehat{V}\| = o_p(1)$. Note that for each h,

$$\tilde{\Omega}_h = \widehat{\Omega}_h - \left(\frac{1}{T}\sum_{t=1}^{T}\widehat{v}_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{v}_t\right)' + \left(\frac{1}{T}\sum_{t=1}^{T}\widehat{v}_t\right)\left(\frac{1}{T}\sum_{t=T-h+1}^{T}\widehat{v}_t\right)' + \left(\frac{1}{T}\sum_{t=1}^{h}\widehat{v}_t\right)\left(\frac{1}{T}\sum_{t}^{T}\widehat{v}_t\right)'$$

Therefore,

$$\begin{aligned}
\|\tilde{V} - \widehat{V}\| &= \left\|\sum_{h=-B}^{B} K\left(\frac{h}{B}\right)(\tilde{\Omega}_h - \widehat{\Omega}_h)\right\| \\
&\leq \sum_{h=-B}^{B} K\left(\frac{h}{B}\right)\left\|\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{v}_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{v}_t\right)'\right\| \\
&\quad + \sum_{h=-B}^{B} K\left(\frac{h}{B}\right)\left\|\frac{1}{T}\left(\sum_{t=1} h\widehat{v}_T + \sum_{t=T-h+1}^{T}\widehat{v}_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{v}_t\right)'\right\| = o_p(1)
\end{aligned}$$

which completes the proof. □

Proof of Proposition 3.1. Proof is straightforward by the following decomposition:

$$\begin{aligned}
\sqrt{T}\widehat{\Omega}^{-1/2}(\widehat{\alpha} - \alpha_0) &= \underbrace{(\widehat{\Omega}^{-1/2} - \Omega^{-1/2})}_{o_p(1)}\underbrace{\sqrt{T}(\widehat{\alpha} - \alpha_0)}_{O_p(1)} + \Omega^{-1/2}\underbrace{\sqrt{T}(\widehat{\alpha} - \alpha_0)}_{\to^d N(0,\Omega)} \\
&\to^d N(0, I_{d_\alpha})
\end{aligned}$$

where $\sqrt{T}(\widehat{\alpha} - \alpha_0) \to^d N(0, \Omega)$ by Theorem 3.1 and $\|\widehat{\Omega}^{-1/2} - \Omega^{-1/2}\| = o_p(1)$ by Theorem 3.3 and continuous mapping theorem. □

C.2 Derivation of Equations (C.1.52) and (C.1.53)

Recall that $h_{t,j}(p^j, p^0) = Y_t\left(\frac{D_{t,j}}{p^j} - \frac{D_{t,0}}{p^0}\right) g^j$ and for $g^j = p^j$

$$\begin{aligned}
&D^j(\xi_t, p - p_0) \\
&= Y_t\left[\left(\frac{D_{t,0}}{(p_0^0)^2} p_0^j (p^0 - p_0^0) - \frac{D_{t,j}}{(p_0^j)^2} p_0^j (p^j - p_0^j)\right) + \left(\frac{D_{t,j}}{p_0^j} - \frac{D_{t,0}}{p_0^0}\right)(p^j - p_0^j)\right]
\end{aligned}$$

while for $g^j = 1$ we have

$$D^j(\xi_t, p - p_0) = Y_t \left(\frac{D_{t,0}}{(p_0^0)^2} \left(p^0 - p_0^0\right) - \frac{D_{t,j}}{\left(p_0^j\right)^2} \left(p^j - p_0^j\right) \right).$$

Simple algebra for the case $g^j = p^j$ then shows that

$$\begin{aligned}
&h_{t,j}\left(p^j, p^0\right) - h_{t,j}\left(p_0^j, p_0^0\right) - D^j(\xi_t, p - p_0) \\
&= Y_t \left(\frac{D_{t,j}}{p^j} - \frac{D_{t,0}}{p^0} \right) p^j - Y_t \left(\frac{D_{t,j}}{p_0^j} - \frac{D_{t,0}}{p_0^0} \right) p_0^j \\
&\quad - Y_t \left[\left(\frac{D_{t,0}}{(p_0^0)^2} p_0^j \left(p^0 - p_0^0\right) - \frac{D_{t,j}}{\left(p_0^j\right)^2} p_0^j \left(p^j - p_0^j\right) \right) + \left(\frac{D_{t,j}}{p_0^j} - \frac{D_{t,0}}{p_0^0} \right) \left(p^j - p_0^j\right) \right] \\
&= Y_t \Bigg(\left(\frac{D_{t,j}}{p^j} - \frac{D_{t,0}}{p^0} \right) p^j - \left(\frac{D_{t,j}}{p_0^j} - \frac{D_{t,0}}{p_0^0} \right) p_0^j \\
&\quad - \left(-\frac{D_{t,j}}{\left(p_0^j\right)^2} p_0^j + \frac{D_{t,j}}{p_0^j} - \frac{D_{t,0}}{p_0^0} \right) \left(p^j - p_0^j\right) - \frac{D_{t,0}}{(p_0^0)^2} p_0^j \left(p^0 - p_0^0\right) \Bigg) \\
&= Y_t \Bigg(\left(\frac{D_{t,j}}{p^j} - \frac{D_{t,0}}{p^0} \right) p^j - \left(-\frac{D_{t,j}}{\left(p_0^j\right)^2} p_0^j \left(p^j - p_0^j\right) \right) \\
&\quad - \left(\frac{D_{t,j}}{p_0^j} - \frac{D_{t,0}}{p_0^0} \right) p^j - \frac{D_{t,0}}{(p_0^0)^2} p_0^j \left(p^0 - p_0^0\right) \Bigg) \\
&= Y_t \left(\left(\frac{-D_{t,j}}{p_0^j p^j} \left(p^j - p_0^j\right) - \frac{D_{t,0}}{p_0^0 p^0} \left(p^0 - p_0^0\right) \right) p^j + \frac{D_{t,j}}{p_0^j} \left(p^j - p_0^j\right) - \frac{D_{t,0}}{(p_0^0)^2} p_0^j \left(p^0 - p_0^0\right) \right) \\
&= Y_t D_{t,0} \left(\frac{p^j}{p_0^0 p^0} - \frac{p^j}{(p_0^0)^2} \right) \left(p^0 - p_0^0\right) \\
&= Y_t \left(D_{t,0} \left(\frac{p_0^0 - p^0}{(p_0^0)^2 p^0} \right) p^j \left(p^0 - p_0^0\right) \right)
\end{aligned}$$

while for $g^j = 1$ one has

$$\begin{aligned}
&h_{t,j}\left(p^j, p^0\right) - h_{t,j}\left(p_0^j, p_0^0\right) - D^j(\xi_t, p - p_0) \\
&= Y_t \left(\frac{D_{t,j}}{p^j} - \frac{D_{t,0}}{p^0} \right) - Y_t \left(\frac{D_{t,j}}{p_0^j} - \frac{D_{t,0}}{p_0^0} \right) - Y_t \left(\frac{D_{t,0}}{(p_0^0)^2} \left(p^0 - p_0^0\right) - \frac{D_{t,j}}{\left(p_0^j\right)^2} \left(p^j - p_0^j\right) \right)
\end{aligned}$$

$$
\begin{aligned}
&= Y_t \left(\left(\frac{D_{t,j}}{p^j} - \frac{D_{t,0}}{p^0} \right) - \left(\frac{D_{t,j}}{p_0^j} - \frac{D_{t,0}}{p_0^0} \right) + \frac{D_{t,j}}{\left(p_0^j\right)^2} \left(p^j - p_0^j\right) - + \frac{D_{t,0}}{\left(p_0^0\right)^2} \left(p^0 - p_0^0\right) \right) \\
&= Y_t \left(\left(D_{t,j} \frac{p_0^j - p^j}{p_0^j p^j} - D_{t,0} \frac{p_0^0 - p^0}{p_0^0 p^0} \right) + \frac{D_{t,j}}{\left(p_0^j\right)^2} \left(p^j - p_0^j\right) - \frac{D_{t,0}}{\left(p_0^0\right)^2} \left(p^0 - p_0^0\right) \right) \\
&= Y_t \left(\frac{D_{t,j}}{\left(p_0^j\right)^2 p^j} \left(p^j - p_0^j\right)^2 - \frac{D_{t,0}}{\left(p_0^0\right)^2 p^0} \left(p^0 - p_0^0\right)^2 \right).
\end{aligned}
$$

C.3 Tables and Figures

Table C.1: Monte Carlo averages of point estimates of the impulse-response functions at persistence level $\rho = 0.70$. Each column indicate the following estimation methods: (P1) parametric estimates, (P2) local projection, (OP) ordered probit, (MN) multinomial logit, (Pl) sieve approximation with polynomial basis, (TP) sieve approximation with trigonometric basis, (HM) sieve approximation with Hermite polynomial basis, and (Wav) sieve approximation with wavelet basis. Numbers in parenthesis indicate Monte Carlo standard deviations.

		Parametric		Semi-parametric					
Periods	Actual	(P1)	(P2)	(OP)	(MN)	(Pl)	(TP)	(HM)	(Wav)
0	0.330	0.330	0.329	0.326	0.326	0.326	0.326	0.326	0.322
		(0.020)	(0.032)	(0.041)	(0.041)	(0.041)	(0.042)	(0.041)	(0.043)
1	0.231	0.230	0.230	0.226	0.226	0.226	0.226	0.226	0.223
		(0.012)	(0.028)	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)	(0.051)
4	0.079	0.078	0.078	0.076	0.076	0.076	0.076	0.076	0.074
		(0.008)	(0.028)	(0.049)	(0.049)	(0.049)	(0.050)	(0.049)	(0.050)
8	0.019	0.019	0.019	0.016	0.016	0.015	0.016	0.015	0.015
		(0.004)	(0.026)	(0.045)	(0.045)	(0.045)	(0.046)	(0.045)	(0.046)
12	0.005	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.003
		(0.002)	(0.026)	(0.047)	(0.047)	(0.047)	(0.048)	(0.047)	(0.048)
16	0.001	0.001	0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002
		(0.001)	(0.026)	(0.047)	(0.047)	(0.047)	(0.048)	(0.047)	(0.048)
20	0.000	0.000	0.000	-0.004	-0.004	-0.004	-0.004	-0.004	-0.005
		(0.000)	(0.026)	(0.045)	(0.045)	(0.046)	(0.046)	(0.046)	(0.046)
24	0.000	0.000	0.000	-0.004	-0.004	-0.004	-0.004	-0.004	-0.005
		(0.000)	(0.026)	(0.046)	(0.046)	(0.046)	(0.047)	(0.046)	(0.047)
28	0.000	0.000	0.000	-0.004	-0.004	-0.004	-0.004	-0.004	-0.005
		(0.000)	(0.027)	(0.049)	(0.049)	(0.049)	(0.050)	(0.049)	(0.049)
32	0.000	0.000	0.000	-0.002	-0.002	-0.002	-0.002	-0.002	-0.003
		(0.000)	(0.026)	(0.044)	(0.044)	(0.044)	(0.045)	(0.044)	(0.045)
36	0.000	0.000	-0.001	-0.004	-0.004	-0.004	-0.004	-0.004	-0.005
		(0.000)	(0.026)	(0.044)	(0.044)	(0.044)	(0.045)	(0.045)	(0.046)

Table C.2: Monte Carlo averages of point estimates of the impulse-response functions at persistence level $\rho = 0.85$. Each column indicate the following estimation methods: (P1) parametric estimates, (P2) local projection, (OP) ordered probit, (MN) multinomial logit, (Pl) sieve approximation with polynomial basis, (TP) sieve approximation with trigonometric basis, (HM) sieve approximation with Hermite polynomial basis, and (Wav) sieve approximation with wavelet basis. Numbers in parenthesis indicate Monte Carlo standard deviations.

		Parametric		Semi-parametric					
Periods	Actual	(P1)	(P2)	(OP)	(MN)	(Pl)	(TP)	(HM)	(Wav)
0	0.330	0.332	0.332	0.323	0.323	0.322	0.322	0.322	0.317
		(0.023)	(0.039)	(0.060)	(0.060)	(0.060)	(0.061)	(0.060)	(0.062)
1	0.281	0.282	0.280	0.272	0.272	0.271	0.271	0.271	0.267
		(0.017)	(0.037)	(0.066)	(0.066)	(0.066)	(0.067)	(0.066)	(0.068)
4	0.172	0.172	0.170	0.165	0.165	0.164	0.165	0.165	0.162
		(0.010)	(0.037)	(0.067)	(0.068)	(0.067)	(0.068)	(0.067)	(0.070)
8	0.090	0.089	0.087	0.080	0.080	0.080	0.080	0.080	0.077
		(0.009)	(0.039)	(0.066)	(0.066)	(0.066)	(0.067)	(0.066)	(0.068)
12	0.047	0.046	0.045	0.036	0.036	0.036	0.036	0.036	0.034
		(0.007)	(0.039)	(0.062)	(0.063)	(0.063)	(0.064)	(0.062)	(0.065)
16	0.025	0.024	0.022	0.015	0.015	0.015	0.015	0.015	0.013
		(0.005)	(0.038)	(0.063)	(0.063)	(0.063)	(0.064)	(0.063)	(0.065)
20	0.013	0.013	0.013	0.003	0.003	0.002	0.003	0.002	0.000
		(0.003)	(0.038)	(0.062)	(0.062)	(0.062)	(0.062)	(0.062)	(0.064)
24	0.007	0.007	0.006	-0.001	-0.001	-0.002	-0.001	-0.001	-0.003
		(0.002)	(0.037)	(0.060)	(0.060)	(0.060)	(0.061)	(0.060)	(0.064)
28	0.003	0.004	0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.006
		(0.001)	(0.037)	(0.061)	(0.061)	(0.061)	(0.062)	(0.061)	(0.065)
32	0.002	0.002	0.002	-0.008	-0.008	-0.009	-0.008	-0.008	-0.011
		(0.001)	(0.038)	(0.061)	(0.061)	(0.061)	(0.062)	(0.061)	(0.064)
36	0.001	0.001	0.000	-0.006	-0.006	-0.006	-0.006	-0.006	-0.008
		(0.001)	(0.035)	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.065)

Table C.3: Monte Carlo averages of point estimates of the impulse-response functions at persistence level $\rho = 0.98$. Each column indicate the following estimation methods: (P1) parametric estimates, (P2) local projection, (OP) ordered probit, (MN) multinomial logit, (Pl) sieve approximation with polynomial basis, (TP) sieve approximation with trigonometric basis, (HM) sieve approximation with Hermite polynomial basis, and (Wav) sieve approximation with wavelet basis. Numbers in parenthesis indicate Monte Carlo standard deviations.

		Parametric		Semi-parametric					
Periods	Actual	(P1)	(P2)	(OP)	(MN)	(Pl)	(TP)	(HM)	(Wav)
0	0.330	0.332	0.301	0.281	0.281	0.277	0.280	0.278	0.262
		(0.023)	(0.098)	(0.141)	(0.143)	(0.145)	(0.151)	(0.145)	(0.180)
1	0.323	0.325	0.293	0.273	0.273	0.269	0.272	0.270	0.254
		(0.022)	(0.097)	(0.146)	(0.147)	(0.149)	(0.155)	(0.149)	(0.183)
4	0.304	0.305	0.274	0.252	0.252	0.248	0.252	0.249	0.234
		(0.020)	(0.098)	(0.149)	(0.150)	(0.152)	(0.158)	(0.152)	(0.185)
8	0.281	0.281	0.248	0.223	0.223	0.219	0.223	0.221	0.205
		(0.017)	(0.098)	(0.151)	(0.153)	(0.154)	(0.160)	(0.155)	(0.190)
12	0.259	0.259	0.225	0.198	0.198	0.194	0.197	0.195	0.180
		(0.014)	(0.101)	(0.151)	(0.152)	(0.153)	(0.160)	(0.155)	(0.190)
16	0.239	0.239	0.206	0.177	0.177	0.173	0.176	0.174	0.159
		(0.012)	(0.099)	(0.151)	(0.153)	(0.155)	(0.161)	(0.155)	(0.189)
20	0.220	0.220	0.189	0.161	0.161	0.157	0.160	0.158	0.144
		(0.011)	(0.102)	(0.150)	(0.152)	(0.153)	(0.159)	(0.154)	(0.188)
24	0.203	0.202	0.174	0.143	0.143	0.139	0.143	0.141	0.127
		(0.010)	(0.100)	(0.151)	(0.153)	(0.153)	(0.159)	(0.154)	(0.188)
28	0.187	0.186	0.156	0.127	0.127	0.123	0.127	0.125	0.112
		(0.010)	(0.101)	(0.152)	(0.153)	(0.152)	(0.158)	(0.153)	(0.187)
32	0.173	0.172	0.141	0.112	0.112	0.108	0.112	0.110	0.097
		(0.010)	(0.101)	(0.148)	(0.150)	(0.148)	(0.154)	(0.150)	(0.185)
36	0.159	0.158	0.127	0.095	0.095	0.091	0.094	0.092	0.079
		(0.010)	(0.100)	(0.150)	(0.152)	(0.149)	(0.158)	(0.150)	(0.186)

Figure C.1: Estimated policy score functions with different first-stage model specifications. Propensity score function is plotted at different quatiles of covariate X.

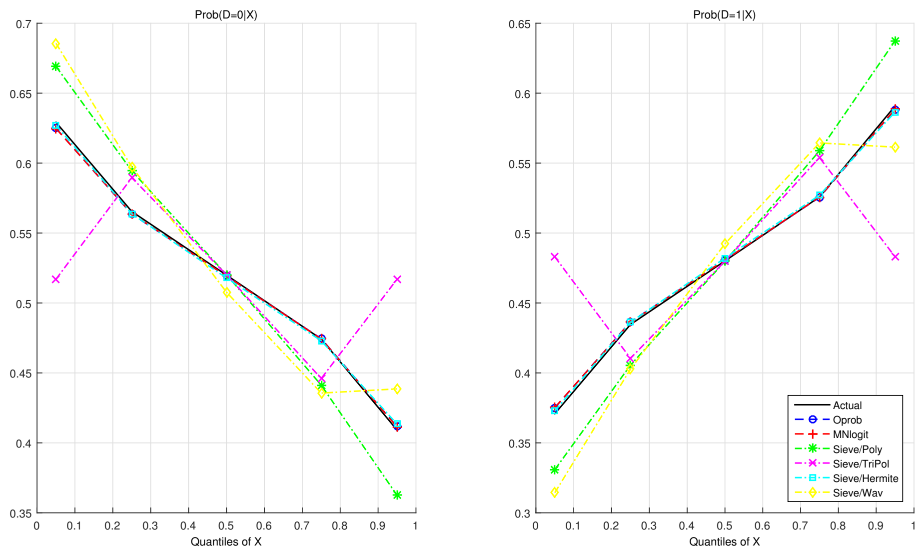


Table C.4: Mean squared error of point estimates of the impulse-response functions at persistence level $\rho = 0.70$. Each column indicate the following estimation methods: (P1) parametric estimates, (P2) local projection, (OP) ordered probit, (MN) multinomial logit, (Pl) sieve approximation with polynomial basis, (TP) sieve approximation with trigonometric basis, (HM) sieve approximation with Hermite polynomial basis, and (Wav) sieve approximation with wavelet basis.

	Parametric		Semi-parametric					
Periods	(P1)	(P2)	(OP)	(MN)	(Pl)	(TP)	(HM)	(Wav)
0	0.0004	0.0010	0.0017	0.0017	0.0017	0.0018	0.0017	0.0019
1	0.0001	0.0008	0.0025	0.0025	0.0025	0.0026	0.0025	0.0027
4	0.0001	0.0008	0.0024	0.0024	0.0024	0.0025	0.0024	0.0025
8	0.0000	0.0007	0.0020	0.0020	0.0020	0.0021	0.0020	0.0022
12	0.0000	0.0007	0.0022	0.0022	0.0022	0.0023	0.0022	0.0023
16	0.0000	0.0007	0.0022	0.0022	0.0023	0.0023	0.0023	0.0024
20	0.0000	0.0007	0.0021	0.0021	0.0021	0.0021	0.0021	0.0022
24	0.0000	0.0007	0.0021	0.0021	0.0021	0.0022	0.0022	0.0023
28	0.0000	0.0007	0.0024	0.0024	0.0024	0.0025	0.0025	0.0025
32	0.0000	0.0007	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
36	0.0000	0.0007	0.0020	0.0020	0.0020	0.0021	0.0020	0.0021

Table C.5: Mean squared error of point estimates of the impulse-response functions at persistence level $\rho = 0.85$. Each column indicate the following estimation methods: (P1) parametric estimates, (P2) local projection, (OP) ordered probit, (MN) multinomial logit, (Pl) sieve approximation with polynomial basis, (TP) sieve approximation with trigonometric basis, (HM) sieve approximation with Hermite polynomial basis, and (Wav) sieve approximation with wavelet basis.

	Parametric		Semi-parametric					
Periods	(P1)	(P2)	(OP)	(MN)	(Pl)	(TP)	(HM)	(Wav)
0	0.0005	0.0015	0.0037	0.0037	0.0036	0.0038	0.0036	0.0041
1	0.0003	0.0014	0.0044	0.0045	0.0044	0.0045	0.0044	0.0048
4	0.0001	0.0014	0.0046	0.0046	0.0046	0.0047	0.0046	0.0049
8	0.0001	0.0015	0.0045	0.0045	0.0045	0.0046	0.0045	0.0048
12	0.0001	0.0015	0.0040	0.0040	0.0041	0.0042	0.0040	0.0044
16	0.0000	0.0014	0.0040	0.0040	0.0040	0.0042	0.0040	0.0044
20	0.0000	0.0014	0.0040	0.0040	0.0040	0.0040	0.0040	0.0042
24	0.0000	0.0014	0.0037	0.0037	0.0037	0.0038	0.0037	0.0042
28	0.0000	0.0014	0.0038	0.0038	0.0038	0.0039	0.0038	0.0043
32	0.0000	0.0014	0.0038	0.0038	0.0038	0.0039	0.0038	0.0043
36	0.0000	0.0013	0.0040	0.0040	0.0040	0.0041	0.0040	0.0043

Table C.6: Mean squared error of point estimates of the impulse-response functions at persistence level $\rho = 0.98$. Each column indicate the following estimation methods: (P1) parametric estimates, (P2) local projection, (OP) ordered probit, (MN) multinomial logit, (Pl) sieve approximation with polynomial basis, (TP) sieve approximation with trigonometric basis, (HM) sieve approximation with Hermite polynomial basis, and (Wav) sieve approximation with wavelet basis.

	Parametric		Semi-parametric					
Periods	(P1)	(P2)	(OP)	(MN)	(Pl)	(TP)	(HM)	(Wav)
0	0.0005	0.0105	0.0224	0.0229	0.0238	0.0254	0.0237	0.0369
1	0.0005	0.0104	0.0238	0.0243	0.0252	0.0267	0.0251	0.0383
4	0.0004	0.0105	0.0249	0.0253	0.0262	0.0277	0.0262	0.0393
8	0.0003	0.0107	0.0262	0.0267	0.0276	0.0289	0.0276	0.0417
12	0.0002	0.0113	0.0265	0.0270	0.0278	0.0294	0.0280	0.0423
16	0.0002	0.0108	0.0268	0.0272	0.0283	0.0297	0.0283	0.0419
20	0.0001	0.0113	0.0260	0.0265	0.0275	0.0290	0.0275	0.0414
24	0.0001	0.0110	0.0265	0.0270	0.0274	0.0290	0.0276	0.0412
28	0.0001	0.0112	0.0266	0.0271	0.0271	0.0285	0.0274	0.0407
32	0.0001	0.0112	0.0257	0.0263	0.0261	0.0275	0.0264	0.0399
36	0.0001	0.0110	0.0268	0.0273	0.0269	0.0291	0.0271	0.0409

Figure C.2: Evolution of mean squared errors and absolute biases of point estimates with different estimation methods.

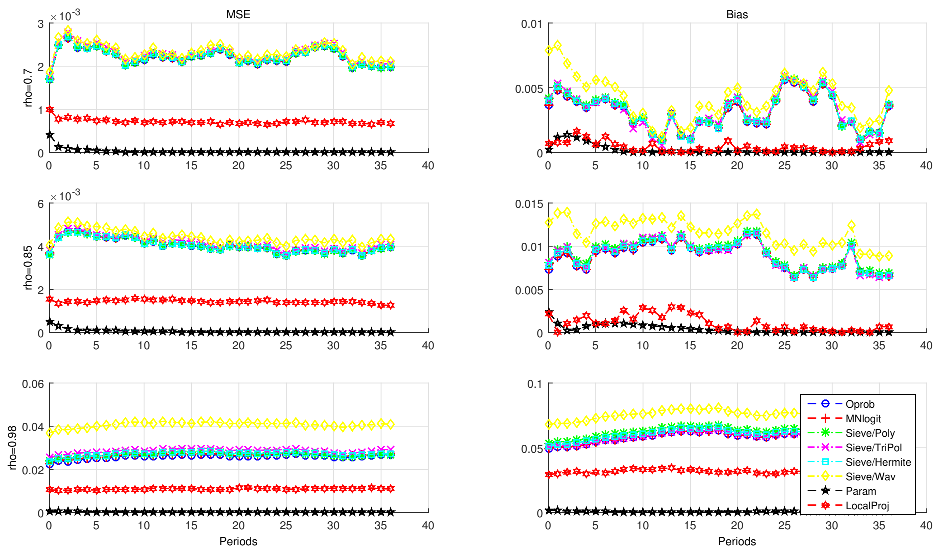


Table C.7: Empirical size of t-test the point estimates of impulse-response functions with underlying process simulated at persistence level of $\rho = 0.70$.

	Semi-parametric					
Periods	(OP)	(MN)	(Pl)	(TP)	(HM)	(Wav)
0	0.066	0.067	0.068	0.061	0.060	0.076
1	0.076	0.075	0.075	0.079	0.065	0.076
4	0.076	0.075	0.074	0.082	0.067	0.081
8	0.050	0.050	0.053	0.059	0.050	0.056
12	0.061	0.061	0.064	0.066	0.054	0.053
16	0.075	0.075	0.080	0.072	0.059	0.062
20	0.050	0.050	0.053	0.058	0.043	0.047
24	0.050	0.050	0.056	0.061	0.046	0.053
28	0.068	0.068	0.073	0.077	0.056	0.069
32	0.045	0.046	0.050	0.046	0.041	0.040
36	0.055	0.058	0.059	0.057	0.047	0.051

Table C.8: Empirical size of t-test the point estimates of impulse-response functions with underlying process simulated at persistence level of $\rho = 0.85$.

	Semi-parametric					
Periods	(OP)	(MN)	(Pl)	(TP)	(HM)	(Wav)
0	0.080	0.079	0.079	0.084	0.057	0.094
1	0.083	0.082	0.086	0.090	0.068	0.101
4	0.077	0.079	0.077	0.077	0.062	0.091
8	0.093	0.092	0.092	0.096	0.060	0.092
12	0.075	0.075	0.076	0.081	0.053	0.077
16	0.069	0.070	0.074	0.073	0.047	0.073
20	0.063	0.064	0.067	0.065	0.043	0.054
24	0.061	0.060	0.068	0.062	0.043	0.071
28	0.067	0.068	0.063	0.055	0.046	0.066
32	0.064	0.064	0.068	0.066	0.041	0.070
36	0.069	0.069	0.069	0.073	0.051	0.052

Table C.9: Empirical size of t-test the point estimates of impulse-response functions with underlying process simulated at persistence level of $\rho = 0.98$.

	Semi-parametric					
Periods	(OP)	(MN)	(Pl)	(TP)	(HM)	(Wav)
0	0.080	0.079	0.079	0.075	0.029	0.104
1	0.085	0.083	0.087	0.074	0.037	0.114
4	0.090	0.086	0.090	0.091	0.039	0.124
8	0.112	0.109	0.108	0.098	0.040	0.134
12	0.131	0.129	0.130	0.120	0.048	0.142
16	0.139	0.137	0.133	0.125	0.047	0.144
20	0.134	0.132	0.134	0.119	0.046	0.147
24	0.144	0.141	0.143	0.128	0.046	0.154
28	0.136	0.134	0.134	0.122	0.053	0.138
32	0.138	0.133	0.133	0.115	0.046	0.151
36	0.140	0.138	0.130	0.117	0.044	0.155

Bibliography

Abadie, A. and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235–267.

Abbring, J. H. and J. J. Heckman (2007): "Chapter 72 Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in *Handbook of Econometrics*, ed. by James J. Heckman and Edward E. Leamer, Elsevier, vol. 6, Part B, 5145–5303.

Abraham, K. G. and H. S. Farber (1987): "Job Duration, Seniority, and Earnings," *The American Economic Review*, 77, 278–297.

Abraham, K. G., J. C. Haltiwanger, K. Sandusky, and J. Spletzer (2016): "The Consequences of Long Term Unemployment: Evidence from Matched Employer-Employee Data," Working Paper 22665, National Bureau of Economic Research.

Acemoglu, D. and D. Autor (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 4, Part B, 1043–1171.

Ackerberg, D., X. Chen, and J. Hahn (2012): "A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators," *Review of Economics and Statistics*, 94, 481–498.

Ackerberg, D., X. Chen, J. Hahn, and Z. Liao (2014): "Asymptotic Efficiency of Semiparametric Two-step GMM," *The Review of Economic Studies*, 81, 919–943.

Altonji, J. G. and R. A. Shakotko (1987): "Do Wages Rise with Job Seniority?" *The Review of Economic Studies*, 54, 437–459.

Altonji, J. G. and N. Williams (2005): "Do Wages Rise with Job Seniority? A Reassessment," *ILR Review*, 58, 370–397.

Andrews, D. W. (1994): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica: Journal of the Econometric Society*, 43–72.

Andrews, D. W. and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica: Journal of the Econometric Society*, 953–966.

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Angrist, J. D., O. Jordà, and G. M. Kuersteiner (2016): "Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited," *Journal of Business & Economic Statistics*, 0–0.

Angrist, J. D., . Jord, and G. M. Kuersteiner (2013): "Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited," Working Paper 19355, National Bureau of Economic Research.

Angrist, J. D. and G. M. Kuersteiner (2004): "Semiparametric Causality Tests Using the Policy Propensity Score," Working Paper 10975, National Bureau of Economic Research.

——— (2011): "Causal Effects of Monetary Shocks: Semiparametric Conditional Independence Tests with a Multinomial Propensity Score," *Review of Economics and Statistics*, 93, 725–747.

Arellano, M. and S. Bonhomme (2012): "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," *The Review of Economic Studies*, 79, 987–1020.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008): "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics*, 90, 300–323.

Bassett, G. and R. Koenker (1982): "An Empirical Quantile Function for Linear Models with iid Errors," *Journal of the American Statistical Association*, 77, 407–415.

Becker, G. S. (1964): *Human capital*, New York: Columbia University Press.

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015):

"Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *Journal of Econometrics*, 186, 345–366.

Billingsley, P. (2008): *Probability and measure*, John Wiley & Sons.

Bonhomme, S. and J.-M. Robin (2010): "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics," *The Review of Economic Studies*, 77, 491–533.

Böttcher, A. (1996): "Infinite Matrices and Projection Methods," *Lectures on operator theory and its applications*, 3, 1–72.

Canay, I. A. (2011): "A simple approach to quantile regression for panel data," *The Econometrics Journal*, 14, 368–386.

Carrington, W. J. (1993): "Wage Losses for Displaced Workers: Is It Really the Firm That Matters?" *The Journal of Human Resources*, 28, 435–462.

Carroll, R. J., A. Delaigle, and P. Hall (2009): "Nonparametric Prediction in Measurement Error Models," *Journal of the American Statistical Association*, 104, 993–1003.

Carroll, R. J. and P. Hall (1988): "Optimal Rates of Convergence for Deconvolving a Density," *Journal of the American Statistical Association*, 83, 1184–1186.

Cattaneo, M. D. (2010): "Efficient Semiparametric Estimation of Multi-valued Treatment Effects under Ignorability," *Journal of Econometrics*, 155, 138–154.

Chen, X. and Z. Liao (2015): "Sieve Semiparametric Two-step GMM under Weak Dependence," *Journal of Econometrics*, 189, 163–186.

Chen, X. and X. Shen (1998): "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica*, 66, 289–314.

Chernozhukov, V. (2005): "Extremal Quantile Regression," *The Annals of Statistics*, 33, 806–839.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010): "Quantile and Probability Curves without Crossing," *Econometrica*, 78, 1093–1125.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205–2268.

Chernozhukov, V., I. Fernández-Val, B. Melly, and K. Wüthrich (2016): "Generic Inference on Quantile and Quantile Effect Functions for Discrete Outcomes," *arXiv:1608.05142 [stat]*, arXiv: 1608.05142.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Chiburis, R. C. (2010): "Semiparametric bounds on treatment effects," *Journal of Econometrics*, 159, 267–275.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary Policy Shocks: What Have We Learned and to What End?" *Handbook of macroeconomics*, 1, 65–148.

Christiano, L. J., M. Trabandt, and K. Walentin (2010): "DSGE Models for Monetary Policy Analysis," in *Handbook of monetary economics*, Elsevier, vol. 3, 285–367.

Couch, K. A. and D. W. Placzek (2010): "Earnings Losses of Displaced Workers Revisited," *The American Economic Review*, 100, 572–589.

Dattner, I., A. Goldenshluger, and A. Juditsky (2011): "On Deconvolution of Distribution Functions," *The Annals of Statistics*, 39, 2477–2501.

Davis, S. J. and T. M. von Wachter (2011): "Recessions and the Cost of Job Loss," Working Paper 17638, National Bureau of Economic Research.

Doksum, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *The Annals of Statistics*, 2, 267–277.

Doukhan, P., P. Massart, and E. Rio (1994): "The Functional Central Limit Theorem for Strongly Mixing Processes," in *Annales de l'IHP Probabilités et statistiques*, Gauthier-Villars, vol. 30, 63–82.

——— (1995): "Invariance Principles for Absolutely Regular Empirical Processes," in *Annales de l'IHP Probabilits et statistiques*, Gauthier-Villars, vol. 31, 393–427.

Dustmann, C. and C. Meghir (2005): "Wages, Experience and Seniority," *The Review of Economic Studies*, 72, 77–108.

Evdokimov, K. (2010): "Identification and estimation of a nonparametric panel data model with unobserved heterogeneity," *Department of Economics, Princeton University*.

FAN, J. (1991a): "Asymptotic Normality for Deconvolution Kernel Density Estimators," *Sankhy: The Indian Journal of Statistics, Series A (1961-2002)*, 53, 97–110.

——— (1991b): "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *The Annals of Statistics*, 19, 1257–1272.

FAN, J. AND Y. K. TRUONG (1993): "Nonparametric Regression with Errors in Variables," *The Annals of Statistics*, 21, 1900–1925.

FAN, Y. AND S. S. PARK (2010): "Sharp Bounds on the Distribution of Treatment Effects and their Statistical Inference," *Econometric Theory*, 26, 931–951.

——— (2014): "Nonparametric inference for counterfactual means: Bias-correction, confidence sets, and weak IV," *Journal of Econometrics*, 178, Part 1, 45–56.

FAN, Y., R. SHERMAN, AND M. SHUM (2014): "Identifying Treatment Effects Under Data Combination," *Econometrica*, 82, 811–822.

FARBER, H. S. (2011): "Job Loss in the Great Recession: Historical Perspective from the Displaced Workers Survey, 1984-2010," Working Paper 17040, National Bureau of Economic Research.

——— (2015): "Job Loss in the Great Recession and its Aftermath: U.S. Evidence from the Displaced Workers Survey," Working Paper 21216, National Bureau of Economic Research.

Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259–276.

Gallant, A. R., P. E. Rossi, and G. Tauchen (1993): "Nonlinear Dynamic Structures," *Econometrica*, 61, 871–907.

Gautier, E. and S. Hoderlein (2011): "A triangular treatment effect model with random coefficients in the selection equation," *arXiv:1109.0362 [math, stat]*, arXiv: 1109.0362.

Geman, S. and C.-R. Hwang (1982): "Nonparametric Maximum Likelihood Estimation by the Method of Sieves," *The Annals of Statistics*, 10, 401–414.

Gil-Pelaez, J. (1951): "Note on the inversion theorem," *Biometrika*, 38, 481–482.

Grafakos, L. (2008): *Classical Fourier Analysis*, vol. 2, New York, NY: Springer.

Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315–331.

Hall, P. and S. N. Lahiri (2008): "Estimation of distributions, moments and quantiles in deconvolution problems," *The Annals of Statistics*, 36, 2110–2134.

Hansen, B. E. (2015): "A Unified Asymptotic Distribution Theory for Parametric and Non-Parametric Least Squares," Tech. rep., Working paper.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Heckman, J. J. and R. Robb (1985): "Alternative methods for evaluating the impact of interventions," *Journal of Econometrics*, 30, 239–267.

Heckman, J. J., J. A. Smith, and N. Clements (1997): "Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *The Review of Economic Studies*, 64, 487–535.

Heckman, J. J., S. Urzúa, and E. Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 88, 389–432.

Heredia, D. M., G. Rucci, F. Saltiel, and S. Urzúa (2017): "Should I Stay or Should I Go? Career Choices for Young Workers in Latin America," Working Paper.

Hirano, K., G. W. Imbens, and G. Ridder (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189.

Horn, R. A. and C. R. Johnson (1985): *Matrix Analysis*, Cambridge [Cambridgeshire] ;: Cambridge University Press.

Horowitz, J. L. and M. Markatou (1996): "Semiparametric Estimation of Regression Models for Panel Data," *The Review of Economic Studies*, 63, 145–168.

Horvitz, D. G. and D. J. Thompson (1952): "A Generalization of Sampling Without Replacement from a Finite Universe," *Journal of the American Statistical Association*, 47, 663–685.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993): "Earnings Losses of Displaced Workers," *The American Economic Review*, 83, 685–709.

Johannes, J. (2009): "Deconvolution with Unknown Error Distribution," *The Annals of Statistics*, 37, 2301–2323.

Jordà, O. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *The American Economic Review*, 95, 161–182.

Kambourov, G. and I. Manovskii (2009): "Occupational Specificity of Human Capital," *International Economic Review*, 50, 63–115.

Kim, J. H. (2014): "Identifying the Distribution of Treatment Effects under Support Restrictions," *arXiv:1410.5885 [stat]*, arXiv: 1410.5885.

Kletzer, L. G. (1989): "Returns to Seniority After Permanent Job Loss," *The American Economic Review*, 79, 536–543.

Kuersteiner, G. M. (2016): "Invariance Principles for Dependent Processes Indexed by Besov Classes with an Application to a Hausman Test for Linearity," *arXiv:1603.07978 [stat]*, arXiv: 1603.07978.

Lewis, R. and G. C. Reinsel (1985): "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of multivariate analysis*, 16, 393–411.

Magnus, J. R. and H. Neudecker (1980): "The Elimination Matrix: Some Lemmas and Applications," *SIAM Journal on Algebraic Discrete Methods*, 1, 422–449.

Maiorov, V. and R. S. Meir (1998): "Approximation bounds for smooth functions in C (R/sup d/) by neural and mixture networks," *IEEE Transactions on Neural Networks*, 9, 969–978.

Makarov, G. D. (1982): "Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed," *Theory of Probability & its Applications*, 26, 803–806.

Mhaskar, H. N. (1996): "Neural networks for optimal approximation of smooth and analytic functions," *Neural computation*, 8, 164–177.

Mincer, J. and B. Jovanovic (1979): "Labor Mobility and Wages," Working Paper 357, National Bureau of Economic Research.

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

Mortensen, D. T. (1978): "Specific Capital and Labor Turnover," *The Bell Journal of Economics*, 9, 572–586.

Neal, D. (1995): "Industry-Specific Human Capital: Evidence from Displaced Workers," *Journal of Labor Economics*, 13, 653–677.

Nelsen, R. B. (2007): *An Introduction to Copulas*, Springer Science & Business Media.

Neumann, M. H. (2007): "Deconvolution from panel data with unknown error distribution," *Journal of Multivariate Analysis*, 98, 1955–1968.

Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica: Journal of the Econometric Society*, 1349–1382.

——— (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79, 147–168.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

——— (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *The Review of Economic Studies*, 61, 631–653.

Pavan, R. (2011): "Career Choice and Wage Growth," *Journal of Labor Economics*, 29, 549–587.

Rio, E. (1993): "Covariance Inequalities for Strongly Mixing Processes," *Ann. Inst. H. Poincar Probab. Statist*, 29, 587–597.

Rosenbaum, P. R. and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

Rothe, C. (2010): "Nonparametric estimation of distributional policy effects," *Journal of Econometrics*, 155, 56–70.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

Rudelson, M. (1999): "Random Vectors in the Isotropic Position," *Journal of Functional Analysis*, 164, 60–72.

Ruhm, C. J. (1991): "Are Workers Permanently Scarred by Job Displacements?" *The American Economic Review*, 81, 319–324.

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

Sklar, M. (1959): *Fonctions de Répartition à n Dimensions et Leurs Marges*, Universit Paris 8.

Stevens, A. H. (1997): "Persistent Effects of Job Displacement: The Importance of Multiple Job Losses," *Journal of Labor Economics*, 15, 165–188.

Taupin, M.-L. (2001): "Semi-Parametric Estimation in the Nonlinear Structural Errors-in-Variables Model," *The Annals of Statistics*, 29, 66–93.

Topel, R. (1991): "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," *Journal of Political Economy*, 99, 145–176.

Zimmer, D. M. and P. K. Trivedi (2006): "Using Trivariate Copulas to Model

Sample Selection and Treatment Effects," *Journal of Business & Economic Statistics*, 24, 63–76.