Accepted manuscript. Published in: Journal of Econometrics 256 (2026) 106272.
Final published version: https://doi.org/10.1016/j.jeconom.2026.106272


# Identification and Estimation in a Time-Varying Endogenous Random Coefficient Panel Data Model*

Ming Li†

June 12, 2026

**Abstract**

This paper proposes a correlated random coefficient linear panel data model, where regressors can be correlated with time-varying and individual-specific random coefficients through both a fixed effect and a time-varying random shock. I develop a new panel data-based method to identify the average partial effect and the local average response function. The identification strategy employs a sufficient statistic to control for the fixed effect and a control variable for the random shock. Conditional on these two controls, the residual variation in the regressors is driven solely by the exogenous instrumental variables, and thus can be exploited to identify the parameters of interest. The constructive identification analysis leads to three-step series estimators, for which I establish rates of convergence and asymptotic normality. To illustrate the method, I estimate a heterogeneous Cobb-Douglas production function for manufacturing firms in China, finding substantial variations in output elasticities across firms that can be related to various firm characteristics.

**Keywords:** Correlated random coefficients, panel data, time-varying endogeneity, semiparametric estimation.

*I am deeply grateful to Donald W. K. Andrews, Yuichi Kitamura, and Steven T. Berry for their invaluable guidance and support. I also thank Stéphane Bonhomme, Xiaohong Chen, Áureo de Paula, Wayne Gao, Guido Imbens, Whitney Newey, Alexandre Poirier, Liangjun Su, Frank Vella, and participants of conferences and seminars at the Duke Class of 2020 and 2021 Microeconometrics Conference, Georgetown, Harvard and MIT joint seminar, ITAM, NTU, NUS, PKU, PSU, THU, UCL, UGA, UQ, Yale, and various Econometric Society Meetings for their helpful comments. I am also grateful to the co-editor, the associate editor, and two anonymous referees for suggestions that greatly improved the paper. All remaining errors are my own.

†Department of Economics and Risk Management Institute, National University of Singapore, 1 Arts Link, Singapore 117570, mli@nus.edu.sg.

# 1 Introduction

Correlated random coefficient (CRC) linear panel models have proven useful due to their ability to accommodate complex forms of unobserved heterogeneity that are empirically relevant (Wooldridge (2010), Hsiao (2022)). A crucial consideration in these models is the correlation between random coefficients and regressors. Classical methods typically address this issue by allowing a time-invariant, individual-specific fixed effect—either in the form of an additive intercept or an individual-specific coefficient—to be correlated with the regressors (e.g., Hausman and Taylor (1981), Hsiao and Pesaran (2008)). While convenient, this approach may not fully capture agents' optimization behavior. For example, even high-capability firms may cut inputs in some periods when input prices are flat because input choices also respond to transitory, time-varying shocks to output efficiency or productivity. Throughout the sample, I assume capability is time-invariant. Thus, both persistent capability and transitory shocks can affect realized output efficiency and may be reflected in input choices, thereby generating endogeneity that is not fully absorbed by a fixed effect.

In this paper, I aim to address this gap by proposing a new time-varying endogenous random coefficient (TERC) linear panel model, where regressors can be correlated with time-varying and individual-specific random coefficients not only through a fixed effect but, more importantly, through a time-varying random shock. Specifically, the baseline TERC model consists of two equations:

$$Y_{it} = X_{it}'\beta\left(A_i, \varepsilon_{it}\right), \tag{1.1}$$

$$X_{it} = g\left(Z_{it}, A_i, \eta_{it}\right). \tag{1.2}$$

The vector of random coefficients $\beta_{it} = \beta\left(A_i, \varepsilon_{it}\right)$ is modeled as a vector of unknown functions $\beta$ of a fixed effect $A_i$ and a time-varying random shock $\varepsilon_{it}$. $Y_{it}$ is then determined by the inner product between $X_{it}$ and $\beta_{it}$. In (1.2), I model the vector of regressors $X_{it}$ as a vector-valued, time-varying function $g$ of the vector of exogenous instrumental variables (IV) $Z_{it}$, fixed effect $A_i$, and per-period information $\eta_{it}$ about $\beta_{it}$.

The motivation for (1.2) is that the agent $i$ in period $t$ optimally chooses values of her regressors $X_{it}$ by solving an optimization problem (e.g., firm profit maximization), with $Z_{it}$, $A_i$, and $\eta_{it}$ in her information set, leading to (1.2). All of $A_i$, $\varepsilon_{it}$, and $\eta_{it}$ can be scalar- or vector-valued, and any pair of the three variables can be correlated. As an analyst, I observe $\{X_{it}, Y_{it}, Z_{it}\}$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$ and aim to identify the pooled average partial effect $T^{-1}\sum_{t=1}^{T}\mathbb{E}\beta_{it}$ (APE, see Graham and Powell (2012)) and the local average response function $\mathbb{E}[\beta_{it} | X_{it}]$ (LAR, see Altonji and Matzkin (2005)).[1]

A key feature of the TERC model is that $X_{it}$ can be correlated with $\beta_{it}$ through both $A_i$ and $\eta_{it}$, potentially in a complicated manner. I refer to this feature as "time-varying endogeneity through the random coefficients." Allowing such correlation is important to applications when agent $i$ in period $t$ has information about $\beta_{it}$ through both $A_i$ and $\eta_{it}$ when optimally deciding $X_{it}$. In Section 2, I present three empirical examples—heterogeneous production function estimation, a labor supply model, and demand estimation—that demonstrate the validity of this correlation (Deaton (1986), Blundell, MaCurdy, and Meghir (2007), Li and Sasaki (2024), Keiller, de Paula, and Van Reenen (2024)). The associated technical challenge is to control for the time-varying endogeneity through the random coefficients when both $A_i$ and $\eta_{it}$ enter $g$ in a nonlinear and nonseparable manner in (1.2). Moreover, I do not impose parametric assumptions on the distributions of $A_i$, $\varepsilon_{it}$, or $\eta_{it}$, nor do I specify the functional form of $\beta$ or $g$. In this sense, the TERC model is a fixed-effects panel data model, and addressing these challenges requires a new method.

I propose a new panel data-based method for identifying the APE and LAR in the TERC model. The idea is to control for $A_i$ and $\eta_{it}$ via a *sufficient statistic* and *a generalized residual*, respectively, such that, conditional on these two controls, the residual variation in $X_{it}$ is driven solely by the exogenous IV $Z_{it}$. In other words, the residual variation in $X_{it}$ is causal and thus can be exploited to identify the parameters of interest. Specifically, I first impose an index exclusion assumption that supplies a sufficient statistic $W_i$ for $A_i$. I present justifications for this assumption from the literature. Next, I construct a feasible conditional cumulative distribution function (CDF) $V_{it} = F_{X_{it}|Z_{it},W_i}(X_{it}| Z_{it}, W_i)$ to control for $\eta_{it}$. Given $(A_i, W_i)$, I establish a one-to-one mapping between $V_{it}$ and $\eta_{it}$. Finally, conditioning on these two controls—which effectively fixes both $A_i$ and $\eta_{it}$—(1.2) implies that the residual variation in $X_{it}$ is driven solely by $Z_{it}$. This residual, instrument-induced variation can therefore be exploited to identify the APE and LAR via perturbation arguments and iterated expectations. At the end of Section 3, I present three extensions: (i) allowing $g$ to depend on multiple components of $\eta_{it}$, (ii) identifying higher-order moments of $\beta_{it}$, and (iii) adding ex-post shocks to $\beta_{it}$ and $Y_{it}$, and including exogenous covariates into $X_{it}$.

I estimate the parameters of interest using a three-step series procedure. In the first step, I estimate the conditional CDF $V_{it}$, which serves as a control for $\eta_{it}$. In the second step, I exploit the linear structure of (1.1) to estimate $\mathbb{E}[Y_{it}| X_{it}, V_{it}, W_i]$. In the third step, I use these estimates to recover the APE and LAR. When working with large datasets, the

[1] Note that the joint distribution of $(\varepsilon_{it}, \eta_{it})$ may vary over $t$. Consequently, any functionals of their distribution (e.g., $\mathbb{E}\beta_{it}$ and $\mathbb{E}[\beta_{it}| X_{it}]$) are in principle time-varying. I suppress the subscript $t$ whenever it is clear from context.

main computational challenge comes from the first-step estimation of $V_{it}$ and the second-step estimation of $\mathbb{E}[Y_{it}|\, X_{it}, V_{it}, W_i]$. I address these issues in Remark 3.

Inference is less standard because these estimators are multi-step: they rely on intermediate nonparametric estimates, include derivatives of an earlier-step estimate, and then apply an additional unknown but estimable mapping to a conditional expectation. As a result, sampling error from the first and second steps can affect the limiting distribution of the final estimators, so valid standard errors must account for how uncertainty propagates across steps. I therefore derive convergence rates and asymptotic normality by explicitly tracking the contribution of each step, building on existing results for multi-step series estimators (Andrews (1991), Newey (1997), Imbens and Newey (2009)). A discussion of how the errors from each step aggregate into the final standard error is provided at the end of Subsection 4.2.

To illustrate the method, I estimate a Cobb–Douglas production function with heterogeneous output elasticities using data on Chinese manufacturing firms. The instruments are inspired by Berry, Levinsohn, and Pakes (1995) and constructed as competitors' leave-one-out weighted-average input prices within the same city and industry. The resulting APE estimates are broadly consistent with those obtained from classical constant-coefficient approaches applied to the same data (Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, Caves, and Frazer (2015)). I also estimate the LAR functions evaluated at each firm's realized input choices, and I find substantial cross-firm variation. I then relate heterogeneity in output elasticities to firm characteristics and obtain economically intuitive patterns. Finally, I conduct extensive robustness checks in Appendix B and find that the results are stable.

To complement the empirical analysis, Appendix C presents Monte Carlo simulations designed to mirror the empirical environment; the results show that the proposed estimator performs well and reinforce the empirical findings. In particular, in the baseline design ($n = 1,000$, $T = 2$, and 1,000 replications, using second-degree spline bases), the APE estimator exhibits small normalized bias (about 2.8–3.2% relative to the length of the true parameter) and low normalized rMSE (about 2.9–3.7%) across coefficients. Estimating the function-valued LAR is naturally less precise, but still yields normalized rMSEs on the order of 3.4–5.6% for all the coefficients. Performance improves with larger $n$ and additional time periods $T$. Robustness checks also indicate that second-degree splines outperform first- or third-degree bases, and that adding ex-post shocks increases rMSEs only slightly.

Before turning to the related literature, I note two limitations of the paper. First, the baseline model assumes each coordinate of $g$ is monotone in a single coordinate of $\eta_{it}$, which (as in Imbens and Newey (2009)) rules out certain nonseparable supply-and-demand systems.

Subsection 3.3 discusses remedies under additional, empirically motivated assumptions. Second, the approach requires IVs and a tractable way to summarize the fixed effect $A_i$ via a sufficient statistic $W_i$; while I provide examples for the applications considered, finding credible instruments (and justifying the required structure) may be challenging in some settings.

## 1.1 Related Literature

This paper contributes to the literature on CRC linear panel models. Related work includes Chamberlain (1992) on regular identification of first moments when $T > d_X$; Graham and Powell (2012) on irregular identification when $T = d_X$ using movers/stayers; and Arellano and Bonhomme (2012) on identifying first (and, under additional assumptions, higher) moments and distributions via within–between variation and restrictions on residual dependence. Wooldridge (2005, 2009) develops estimation strategies for APEs—respectively under mean-independence conditions and with sample selection in unbalanced panels—while Murtazashvili and Wooldridge (2008) study consistency of fixed-effects IV estimators in time-invariant CRC models. Additional contributions include identification of quantiles (Graham, Hahn, Poirier, and Powell (2018)), random-coefficient simultaneous equations (Masten (2018)), and a two-step identification approach with additive fixed effects and time-invariant random coefficients (Laage (2024)).

My contribution to the CRC literature is to study a model, (1.1) and (1.2), with time-varying random coefficients that can be correlated with regressors through both fixed effects and time-varying shocks. The closest paper to my work is Graham and Powell (2012), who identify moments by exploiting movers and stayers under a time-stationarity restriction. This time-stationarity restriction conditions on the regressors for all periods and thus effectively rules out the time-varying endogeneity central here. I instead introduce controls for $A_i$ and importantly $\eta_{it}$ to identify the APE and LAR. The two approaches rely on non-nested assumptions; I view this paper as a step toward addressing time-varying endogeneity through random coefficients.

In addition to the aforementioned papers, my work is also related to the literature on: (i) time-varying parameter panel data models (e.g., Li, Chen, and Gao (2011), Li, Qian, and Su (2016)), (ii) latent structure panel data models (e.g., Su, Shi, and Phillips (2016), Su, Wang, and Jin (2019), Wang, Phillips, and Su (2024)), and (iii) linear panel data model with interactive fixed effects, which can be treated as imposing a specific structure on the random coefficient associated with the constant term (e.g., Pesaran (2006), Bai (2009), Moon and Weidner (2015)). The modeling technique and identification method of my paper are different from these seminal papers.

The second line of research concerns the techniques used in this paper. For nonseparable panel models with generalized fixed effects, restrictions on the fixed effect are generally needed to identify global objects such as the APE and LAR (Hoderlein and White (2012, p.2); see also Gao, Li, and Xu (2023) and Gao and Li (2026) on dealing with the issue of fixed effects in network formation and panel multinomial choice models, respectively). My method requires a sufficient statistic $W_i$ for $A_i$ and a control variable $V_{it}$ for $\eta_{it}$. To justify the sufficiency of $W_i$ for $A_i$, I adapt several techniques from the literature (e.g., Mundlak (1978), Altonji and Matzkin (2005), Bester and Hansen (2009), Wooldridge (2019), Arkhangelsky and Imbens (2022)). See Liu, Poirier, and Shiu (2025) for a discussion on the index sufficiency condition in nonlinear semiparametric panel data models. To justify $V_{it}$ as a valid control for $\eta_{it}$, I generalize the technique of Imbens and Newey (2009) in two nontrivial ways and discuss them in relation to (3.15). More recently, Nagasawa (2024) studies treatment effect estimation with noisy conditioning variables using control variables from the joint distribution of noisy proxy measures.

The asymptotic analysis builds on foundational results for series and control-function estimators. Andrews (1991) and Newey (1997) provide conditions for convergence rates and asymptotic normality of series estimators, which I use to handle vector-valued functionals of regression functions. Newey, Powell, and Vella (1999) establish asymptotic normality for two-step nonparametric estimators in triangular models with a separable first stage, while Imbens and Newey (2009) extend the triangular framework to nonseparable first stages and derive rates and asymptotic normality for known functionals of conditional expectations. I build on Imbens and Newey (2009) to obtain asymptotic normality for unknown but estimable functionals of conditional expectation functions.

**Organization**. The rest of the paper is organized as follows. Section 2 formally introduces the TERC model and provides several empirical examples that fit within its framework. Section 3 outlines the identification idea and presents the assumptions, key identification theorem, and three extensions. Section 4 presents the series estimators for the APE and LAR and establishes their asymptotic properties. Section 5 provides an empirical illustration. Finally, Section 6 concludes. Proofs, additional empirical results, and a simulation study are included in Appendix A, B, and C, respectively.

**Notation**. Let $i \in \{1, \ldots, n\}$ index agents and $t \in \{1, \ldots, T\}$ index periods, with finite $T \geq 2$. I use boldface upper case letters for random matrices, regular upper case letters for random variables and random vectors, and lower case letters for their values. Let $\mathbf{X}_i = (X_{i1}, ..., X_{iT})'$, $Y_i = (Y_{i1}, ..., Y_{iT})'$, $\mathbf{Z}_i = (Z_{i1}, ..., Z_{iT})'$, $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$, and $\eta_i = (\eta_{i1}, ..., \eta_{iT})'$. I assume $(\mathbf{X}_i, \mathbf{Z}_i, A_i, \varepsilon_i, \eta_i)$ are i.i.d. across $i$. Let $d_X$ be the dimension of $X_{it}$ and $X_{it,l}$ denote the $l^{\text{th}}$ coordinate of vector $X_{it}$. Similar notation is used for all other variables. I use $\coloneqq$ to define a new random variable, $\sim_d$ to indicate that two random variables are identically distributed, $F$ for CDF, $f$ for probability density function (PDF), $\mathbb{E}$ for expectation, $\mathbb{V}$ for variance, $I_d$ for $d \times d$ identity matrix, $\xrightarrow{p}$ for convergence in probability, and $\xrightarrow{d}$ for convergence in distribution. All convergence results are stated under $n \to \infty$.

# 2 Model

Consider the following baseline TERC model

$$Y_{it} = X_{it}' \beta \left( A_i, \varepsilon_{it} \right), \tag{2.1}$$

$$X_{it} = g \left( Z_{it}, A_i, \eta_{it} \right), \tag{2.2}$$

where $\beta_{it} = \beta \left( A_i, \varepsilon_{it} \right) \in \mathbb{R}^{d_X}$, the central object of interest, is a vector of random coefficients modeled as an unknown vector-valued function $\beta$ of $\left( A_i, \varepsilon_{it} \right)$. Here, $A_i$ is a fixed effect and $\varepsilon_{it}$ governs the time-varying behavior of $\beta_{it}$; both may have arbitrary dimension. The mapping $\beta$ itself can be time-varying, because $\varepsilon_{it}$ captures time variation in the functional form. $\eta_{it} \in \mathbb{R}^{d_\eta}$ is a continuously distributed, time-varying random vector that may be correlated with $A_i$ and $\varepsilon_{it}$; its coordinates may also be mutually correlated. This correlation can be important in applications; see the examples below. Let $X_{it} \in \mathbb{R}^{d_X}$ denote choice variables (e.g., capital and labor), $Y_{it} \in \mathbb{R}$ the outcome, and $Z_{it} \in \mathbb{R}^{d_Z}$ instruments. Finally, $g$ is an unknown vector-valued function of $\left( Z_{it}, A_i, \eta_{it} \right)$ that determines each coordinate of $X_{it}$.

The TERC model is widely used in empirical studies. It extends the classic constant-coefficient linear panel model while preserving directly interpretable targets—such as the APE and LAR—that carry clear policy relevance. It also accommodates natural extensions, including higher-order moments of the APE and LAR. In what follows, I describe three applications of the TERC model.

**Example 1** (Production Function Estimation)**.** Suppose firm $i$ in year $t$ has a heterogeneous Cobb-Douglas production function in the form of (2.1). The capital and labor elasticities $\beta_{it} = \beta \left( A_i, \varepsilon_{it} \right) \in \mathbb{R}^2$ are modeled as a two-dimensional function of the firm fixed effect $A_i$ (e.g., manager ability that is assumed to be stable over time) and a time-varying random shock $\varepsilon_{it}$ (e.g., realized per-period technology shock). When deciding its $X_{it} \in \mathbb{R}^2$ of capital and labor, assume firm $i$ knows $A_i$ and $\eta_{it}$ (e.g., expected per-period technology shock). Clearly, $\eta_{it}$ is correlated with $A_i$ and $\varepsilon_{it}$. Assume firm $i$ also observes $Z_{it}$, which captures local input market dynamics that affect $X_{it}$—for example, competitors' debt-weighted interest rates and employment-weighted wages. Let the cost function be given by $c \left( x, z \right)$. Then, firm

$i$ selects $X_{it}$ by maximizing its expected profit with the knowledge of $(Z_{it}, A_i, \eta_{it})$, i.e.,

$$X_{it} = \arg\max_{x \in \mathbb{R}^2} \left[ \mathbb{E} \left[ x' \beta \left( A_i, \varepsilon_{it} \right) \middle| Z_{it}, A_i, \eta_{it} \right] - c \left( x, Z_{it} \right) \right],$$

leading to (2.2). Subsequently, firm $i$ obtains its output $Y_{it}$ via (2.1).

**Example 2** (Labor Supply)**.** Suppose worker $i$ has a linear labor supply function of the form (2.1), where $Y_{it}$ is the number of hours worked and $X_{it}$ includes the endogenous hourly wage ($X_{it,1}$) along with other exogenous demographic variables. The coordinate of $\beta_{it}$ corresponding to the wage is the key object of interest because it quantifies the elasticity of labor supply with respect to the wage rate. Given exogenous factors $Z_{it}$ (e.g., county minimum-wage changes or earned income tax credit expansions), individual ability $A_i$, and shock $\eta_{it}$ (e.g., her expectation of a near-term surge in labor demand in her sector, which increases the expected marginal product of labor and wage offers), she chooses a job with the wage that maximizes her expected utility, resulting in (2.2). Finally, worker $i$ commits $Y_{it}$ hours of working via (2.1).

**Example 3** (Almost Ideal Demand System (AIDS))**.** Suppose household $i$'s gasoline budget share $Y_{it}$ at time $t$ depends on the gasoline price $X_{it,1}$ and total expenditure $X_{it,2}$ as in (2.1). Let $\beta_{it} \in \mathbb{R}^2$ denote the possibly heterogeneous elasticities with respect to $X_{it}$, modeled as unknown functions of a household fixed effect $A_i$ (e.g., a persistent propensity to drive versus use transit) and a time-varying wealth shock $\varepsilon_{it}$. Given instruments $Z_{it}$ (e.g., head of household's earned income and gas tax change), $A_i$, and a signal $\eta_{it}$ about $\beta_{it}$ (e.g., expected near-term income change such as upcoming bonus), the household searches for gasoline prices and sets its total expenditure budget by maximizing expected utility, which yields (2.2).

The time-varying correlation between $X_{it}$ and $\beta_{it}$ illustrates a distinct source of endogeneity: optimization-induced selection into regressors through random coefficients. Classic work (Manski (1987), Altonji and Matzkin (2005), Graham and Powell (2012), Chernozhukov, Fernández-Val, Hahn, and Newey (2013)) allows $X_{it}$ to correlate with $A_i$ (e.g., better-managed firms choose higher inputs). But empirically, $X_{it}$ can also co-move with $\beta_{it}$ beyond $A_i$, because time variation in choices is rarely fully spanned by instrument variation $Z_{it}$. For example, even with little movement in input prices, a well-managed firm may temporarily scale inputs up or down in response to private information about returns. This motivates a time-varying shock $\eta_{it}$ that affects $X_{it}$ and is informative about $\beta_{it}$.

A second departure from standard triangular models (Newey, Powell, and Vella (1999), Imbens and Newey (2009)) is that $A_i$ enters both the outcome equation and the first-step choice equation: $X_{it} = g\left(Z_{it}, A_i, \eta_{it}\right)$, not $X_{it} = g\left(Z_{it}, \eta_{it}\right)$. Consequently, residual- or CDF-based control-function approaches that rely on a one-to-one mapping between the control and $\eta_{it}$ no longer apply. Moreover, $g$ is nonseparable and can depend on $A_i$ nonlinearly, as implied by optimization, so demeaning or first-differencing cannot eliminate $A_i$. I instead use an index exclusion condition to handle the fixed effects.

# 3 Identification

## 3.1 Identification Idea

### Why Standard Linear Identification Fails

The main obstacle to identifying the APE and the LAR function in the TERC model using standard linear identification is that the regressor is chosen by the agent using information that is also informative about the random coefficients. To see where the standard argument breaks down, take conditional expectations in (2.1) given $X_{it}$:

$$\mathbb{E}[Y_{it}|\, X_{it}] = X_{it}'\mathbb{E}[\beta_{it}|\, X_{it}]. \tag{3.1}$$

In a constant-coefficient model, the term $\mathbb{E}[\beta_{it}|\, X_{it}]$ would be a constant vector and could be recovered from moment conditions. Here, however, $\mathbb{E}[\beta_{it}|\, X_{it}]$ is generally a function of $X_{it}$. As a result, the familiar argument based on multiplying both sides of (3.1) by $X_{it}$, taking expectations, and inverting $\mathbb{E}[X_{it}X_{it}']$ does not identify $\mathbb{E}[\beta_{it}|\, X_{it}]$:

$$\mathbb{E}[X_{it}Y_{it}] = \mathbb{E}\left[X_{it}X_{it}'\mathbb{E}[\beta_{it}|\, X_{it}]\right], \tag{3.2}$$

because $X_{it}X_{it}'$ and $\mathbb{E}[\beta_{it} \mid X_{it}]$ move together inside the expectation. Likewise, the "derivative/perturbation" argument for linear models does not isolate $\mathbb{E}[\beta_{it}|\, X_{it}]$, since differentiating (3.1) yields

$$\frac{\partial \mathbb{E}[Y_{it}|\, X_{it}]}{\partial X_{it}} = \mathbb{E}[\beta_{it}|\, X_{it}] + \left(\frac{\partial \mathbb{E}[\beta_{it}|\, X_{it}]}{\partial X_{it}}\right)' X_{it}, \tag{3.3}$$

and the second term captures how selection into different $X_{it}$ changes the conditional distribution of $\beta_{it}$.

### Motivation and Interpretation of the Two Control Variables

The identification strategy in this paper is to recreate, using observables, the thought experiment "vary $X_{it}$ while holding fixed everything the agent knows that is related to $\beta_{it}$." Doing so requires two controls—one for the time-invariant heterogeneity $A_i$ and one for the time-varying information $\eta_{it}$.

The first control is a statistic $W_i = W(\mathbf{X}_i, \mathbf{Z}_i)$ constructed from the individual's panel history. The maintained index exclusion condition implies that once I condition on $W_i$, the remaining variation in $(X_{it}, Z_{it})$ for a fixed $t$ contains no additional information about $A_i$:

$$A_i | \left(X_{it}, Z_{it}, W_i\right) \sim_d A_i | W_i. \tag{3.4}$$

Intuitively, $W_i$ is chosen to summarize the persistent, individual-specific component of behavior that is linked to $A_i$. For example, in many panel applications a time average of $X_{it}$ (possibly together with time averages of instruments) is a natural candidate: firms with higher managerial ability or individuals with higher baseline productivity tend to exhibit persistently different choices, and $W_i$ is designed to capture that persistent component.

Even after accounting for $A_i$, endogeneity remains because agents may respond to time-varying information $\eta_{it}$ when choosing $X_{it}$. The second control is constructed as the conditional CDF (a "rank" or "generalized residual"):

$$V_{it} := F_{X_{it}|Z_{it}, W_i}\left(X_{it} | Z_{it}, W_i\right). \tag{3.5}$$

The interpretation is that $V_{it}$ records where the realized $X_{it}$ sits in the distribution of $X_{it}$ among agents with the same $(Z_{it}, W_i)$. Since $A_i$ and $Z_{it}$ appear in the first-step equation (1.2) and $W_i$ controls for $A_i$ by (3.4), the residual variation of $X_{it}$ given $(Z_{it}, W_i)$ is only driven by $\eta_{it}$ which is assumed to enter (1.2) strictly monotonically. For example, consider two firms with the same observed average inputs over all periods ($W_i$) and input prices in period $t$ ($Z_{it}$). If firm 1's choice of capital ($X_{it}$) in this period is higher than that of firm 2, it can be inferred that firm 1 receives a larger productivity shock ($\eta_{it}$) to its capital choice function in period $t$ than firm 2 since (3.4) ensures that the residual variation in $A_i$ given $W_i$ is not informative about the choice of $X_{it}$.

As a result, $V_{it}$ records where the realized $\eta_{it}$ sits in the distribution of $\eta_{it}$ among agents with the same $(Z_{it}, A_i, W_i)$. By an exogeneity assumption between $Z_{it}$ and $\eta_{it}$, it drops out from the conditioning set and hence $V_{it}$ controls for $\eta_{it}$ given $(A_i, W_i)$. Thus, conditioning on $(V_{it}, W_i)$ is intended to hold fixed both the individual's persistent type (through $W_i$) and the individual's period-$t$ information that drives endogenous adjustment (through $V_{it}$).

This two-control structure matches the economics of the motivating examples in Section 2. In the production function estimation example, $W_i$ captures persistent firm heterogeneity (e.g., management skill) that affects both average input choices and average elasticities, while $V_{it}$ captures time-$t$ private information or expectations about technology that shift input choices and are correlated with realized productivity shocks. In the labor supply application, $W_i$ summarizes persistent ability/tastes and $V_{it}$ captures time-varying job-match information that affects wage choice. In the AIDS application, $W_i$ captures persistent household heterogeneity and $V_{it}$ captures time-varying wealth or liquidity information that affects expenditure plans.

### Steps to Identify APE and LAR

Given the two controls $V_{it}$ and $W_i$, the central step to identification of the APE and LAR is to show that once I condition on the two controls, $X_{it}$ no longer carries additional information about $\beta_{it}$. The argument proceeds in two layers.

The first step tackles the *time-varying* part of the endogeneity by assuming $A_i$ is known and showing

$$\mathbb{E}[\beta_{it}|\, X_{it}, A_i, V_{it}, W_i] = \mathbb{E}[\beta_{it}|\, A_i, V_{it}, W_i]. \tag{3.6}$$

The key fact behind (3.6), established in the proof of Theorem 1, is that $(A_i, V_{it}, W_i)$ uniquely pins down $(A_i, \eta_{it}, W_i)$. Hence, equation (1.2) implies that the remaining variation in $X_{it}$ is generated only by $Z_{it}$. Under the maintained IV exogeneity condition, this residual variation is independent of $\beta_{it}$. This is the point at which the time-varying part of the endogeneity problem is resolved: within cells defined by $(A_i, V_{it}, W_i)$, movements in $X_{it}$ are "as good as randomized" by $Z_{it}$. In the production function estimation example, if two firms have the same management ability ($A_i$) and expected per-period technology shock ($\eta_{it}$) but selects different capital or labor, such difference can only be explained by the variation in exogenous $Z_{it}$, which is causal and thus can be exploited to identify the APE and LAR.

The second step tackles the *time-invariant* part of the endogeneity. Since $A_i$ is not observed, it is not feasible to condition on it in (3.6). I deal with unknown $A_i$ by integrating it out. This involves iterating expectations and the index exclusion restriction of $W_i$. Specifically, I show

$$\mathbb{E}[\mathbb{E}[\beta_{it}|\, A_i, V_{it}, W_i]|\, X_{it}, V_{it}, W_i] = \mathbb{E}[\beta_{it}|\, V_{it}, W_i]. \tag{3.7}$$

Notice that on the left-hand side of (3.7), $X_{it}$ affects the conditional expectation only through the conditional density $f_{A_i|X_{it},V_{it},W_i}$. The index exclusion restriction (Assumption 2) implies that, conditional on $W_i$, $A_i$ is independent of $(X_{it}, Z_{it})$ and thus of $V_{it}$. For instance, in the production function estimation example, $W_i$ captures the firm's time-invariant management ability $A_i$, so once $W_i$ is held fixed, period-$t$ inputs $X_{it}$ and control for productivity shock $V_{it}$ only reflect transitory variation and add no extra information about $A_i$. Therefore, I have

$$f_{A_i|X_{it},V_{it},W_i}(A_i|\, X_{it}, V_{it}, W_i) = f_{A_i|W_i}(A_i|\, W_i), \tag{3.8}$$

which ensures that (3.7) holds.

Combining (3.6) and (3.7) gives

$$\mathbb{E}[\beta_{it}|\, X_{it}, V_{it}, W_i] = \mathbb{E}\left[\mathbb{E}[\beta_{it}|\, X_{it}, A_i, V_{it}, W_i]|\, X_{it}, V_{it}, W_i\right] = \mathbb{E}[\beta_{it}|\, V_{it}, W_i]. \tag{3.9}$$

Equation (3.9) is the key identification result: after conditioning on $(V_{it}, W_i)$, the dependence of $\beta_{it}$ on $X_{it}$ is removed.

Given (3.9), equation (2.1) implies a conditional linear representation

$$\mathbb{E}[Y_{it}|\, X_{it}, V_{it}, W_i] = X_{it}'\mathbb{E}[\beta_{it}|\, V_{it}, W_i]. \tag{3.10}$$

With sufficient residual variation in $X_{it}$ conditional on $(V_{it}, W_i)$, $\mathbb{E}[\beta_{it}|\, V_{it}, W_i]$ can be identified either from conditional second moments

$$\mathbb{E}[\beta_{it}|\, V_{it}, W_i] = \mathbb{E}[X_{it}X_{it}'|\, V_{it}, W_i]^{-1}\mathbb{E}[X_{it}Y_{it}|\, V_{it}, W_i], \tag{3.11}$$

or from the derivative of the conditional mean

$$\mathbb{E}[\beta_{it}|\, V_{it}, W_i] = \frac{\partial \mathbb{E}[Y_{it}|\, X_{it}, V_{it}, W_i]}{\partial X_{it}}. \tag{3.12}$$

Finally, identification of the APE and LAR follows by the law of iterated expectations

$$\bar{b} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\mathbb{E}[\beta_{it}|\, V_{it}, W_i]\right], \quad \text{and} \quad b_t(X_{it}) = \mathbb{E}\left[\mathbb{E}[\beta_{it}|\, V_{it}, W_i]|\, X_{it}\right]. \tag{3.13}$$

In the production function estimation example, $\bar{b}$ summarizes output elasticities by averaging across firms and time, so it plays the same role as the constant coefficient in standard linear panel regressions. $b_t(x)$ denotes the period-$t$ average elasticity conditional on operating at input level $x$, a policy-relevant local parameter when elasticities are systematically related to firms' input choices. Analogous interpretations apply to labor supply (average and local wage elasticities) and demand estimation (average and local expenditure/price elasticities). The purpose of $(W_i, V_{it})$ is precisely to make these interpretations credible by ensuring that the remaining variation used to identify (3.13) is driven by the exogenous instruments rather than by unobserved, coefficient-relevant information.

## 3.2 Assumptions

Next, I provide a list of assumptions on the model primitives required for the identification analysis and discuss them in relation to model (2.1)–(2.2).

**Assumption 1 (Componentwise Monotonicity).** *For every $(Z_{it}, A_i)$ and each $j \in \{1, \ldots, d_X\}$, $g_j(Z_{it}, A_i, \eta_{it})$ depends on $\eta_{it}$ only through $\eta_{it,j}$ and is strictly monotone in $\eta_{it,j}$.*

The role of Assumption 1 is to deliver the invertibility/control-function step used in the identification argument. This assumption is mild when $\eta_{it}$ is scalar: it is automatically satisfied in additive triangular specifications (e.g., Newey, Powell, and Vella (1999)), and in the nonseparable scalar case it is essentially the same as the monotonicity condition in Imbens and Newey (2009, eq. (2)), except that I also allow for time-invariant heterogeneity $A_i$. When $d_X > 1$, as in production-function estimation (Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, Caves, and Frazer (2015)), one can further weaken the requirement to monotonicity in only one endogenous input, mirroring the identifying restriction in those classic approaches.

Assumption 1 requires justification when $\eta_{it}$ is vector-valued and each coordinate of $X_{it}$ admits one distinct coordinate of $\eta_{it}$. In the production function estimation example, this is plausible if firms make hiring and investment decisions in different units (e.g., HR and finance), each responding primarily to its own market-specific signal—labor-market conditions for hiring and financial conditions for investment—satisfying the "single-coordinate" dependence. Moreover, positive shocks in each market naturally raise the corresponding choice (more hiring when labor conditions improve, more investment when financing conditions improve), implying that $g_j$ is strictly monotone in $\eta_{it,j}$ for each coordinate $j$ and thus satisfying componentwise monotonicity.

Importantly, Assumption 1 places no restrictions on the dependence structure across coordinates of $\eta_{it}$, nor on the dependence between $A_i$ and $\eta_{it}$. Accordingly, without loss of generality, one may normalize $\eta_{it}$ so that $d_\eta = d_X$ and each coordinate of $g$ depends on the corresponding coordinate of $\eta_{it}$.

*Remark* 1. Assumption 1 excludes settings where multiple components of $\eta_{it}$ enter a single coordinate of $X_{it}$, as in simultaneous supply–demand systems. Subsection 3.3 relaxes this restriction via timing assumptions, global univalence conditions, or single-index structure.

**Assumption 2 (Index Exclusion).** *Let $W_i := W(\mathbf{X}_i, \mathbf{Z}_i)$, where $W : \mathbb{R}^{T \times (d_X + d_Z)} \mapsto \mathbb{R}^{d_W}$ is known. Suppose $A_i | (X_{it}, Z_{it}, W_i) \sim_d A_i | W_i$ for each $t$.*

Assumption 2 serves two roles. First, it makes $W_i$ a sufficient statistic for $A_i$, which is used to establish (3.9). Second, it is used to construct $V_{it}$ and establish its connection with

$\eta_{it}$ by holding $A_i$ fixed. Under this assumption, the channel of endogeneity operating through $A_i$ is handled entirely via $W_i$, while the remaining "time-varying endogeneity through the random coefficients" is driven by the relationship between $\eta_{it}$ and $\varepsilon_{it}$ and is addressed via $V_{it}$. Assumption 2 is similar to Assumption 2.1 of Altonji and Matzkin (2005) and Assumption A2 of Liu, Poirier, and Shiu (2025). Considering the heterogeneity and endogeneity afforded in the TERC model, some restriction that isolates $A_i$ is typically needed.

To motivate $W_i$ in empirical settings, consider steady-state input policies with convex adjustment costs in the production application. In standard investment and hiring models with convex (e.g., quadratic) adjustment costs (Cooper and Haltiwanger (2006)), optimality implies convergence to a firm-specific long-run target $X_i^\star(A_i)$ that is increasing in ability $A_i$. Observed inputs fluctuate around this target due to transitory shocks and gradual adjustment. Averaging $X_{it}$ over time filters out these high-frequency deviations and provides a consistent estimator of $X_i^\star(A_i)$. Because $X_i^\star(\cdot)$ is monotone, conditioning on $\overline{X}_i = T^{-1}\sum_{t=1}^{T} X_{it}$ induces a control for $A_i$, satisfying Assumption 2.

Next, I provide three sufficient conditions for Assumption 2.

First, Assumption 2 can be justified by nonparametrically generalizing equation (2.4) of Mundlak (1978) to be $A_i = h\left(\overline{X}_i, \nu_i\right)$, where $\nu_i \perp (X_{it}, Z_{it})$ for all $t$. Here, the functional form of $h$ and distribution of $\nu_i$ can be unknown. Then, Assumption 2 is satisfied by taking $W_i$ to be $\overline{X}_i$. This is true because conditioning on $W_i$, any $t$-specific $X_{it}$ and $Z_{it}$ does not affect the distribution of $A_i$ as the residual randomness in $A_i$ is only driven by $\nu_i$.

Second, insights from treatment assignment models in panel data (e.g., Arkhangelsky and Imbens (2022)) can be exploited to find candidate $W_i$. For example, if $(X_{it}, Z_{it})$ conditional on $A_i$ is a two-dimensional normal random vector i.i.d. through time with known covariance and $\mathbf{Z}_i \perp A_i$, then the density of $(X_{it}, Z_{it})$ given $A_i$ depends on $A_i$ only through $\overline{X}_i$. Hence, by the sufficient statistic for bi-variate normal random vectors with known covariance matrix (Hogg, McKean, and Craig (2019, ch. 7)), $W_i = (\overline{X}_i, \overline{Z}_i)$ suffices for Assumption 2.

Third, the nonparametric exchangeability condition (e.g., Altonji and Matzkin (2005)) can be adapted to justify Assumption 2. I present its details in Proposition 1 of Appendix A.2. The key idea is time exchangeability: conditional on $A_i$, the joint behavior of $(X_{it}, Z_{it})$ is unchanged if I relabel periods, so the time order contains no information about $A_i$. Hence only order-invariant summaries of the panel matter for learning about $A_i$, and symmetric functions of the observed pairs $(X_{it}, Z_{it})$ can be used as $W_i$ to satisfy Assumption 2.

Next, I control for the time-varying $\eta_{it}$. When $A_i$ is not present in (2.2) and under

Assumption 1, Imbens and Newey (2009) propose

$$V_{it}^{\mathrm{IN09}} \coloneqq F_{X_{it}|Z_{it}}\left(X_{it}|\, Z_{it}\right) \overset{\text{by IN09}}{=} F_{\eta_{it}}\left(\eta_{it}\right) \tag{3.14}$$

as a control for $\eta_{it}$. The intuition is that, holding input prices $Z_{it}$ fixed, a firm choosing a higher capital $X_{it}$ must be experiencing a higher productivity shock $\eta_{it}$; otherwise the higher input choice would not be optimal. However, this argument breaks down in the TERC model because the unobserved, time-invariant component $A_i$ also shifts input choices: firm 1's higher $X_{it}$ could reflect greater managerial ability rather than a larger transitory shock $\eta_{it}$. Therefore $A_i$ must also be conditioned on, but this makes $F_{X_{it}|A_i,Z_{it}}\left(X_{it}|\, A_i, Z_{it}\right)$ infeasible since $A_i$ is unobserved.

I proceed in two steps to generalize the argument of Imbens and Newey (2009), adapting it to the TERC model. First, to make the control for $\eta_{it}$ feasible, I replace $A_i$ with the observable proxy $W_i$ and define

$$V_{it} = F_{X_{it}|Z_{it},W_i}\left(X_{it}|\, Z_{it}, W_i\right). \tag{3.15}$$

Then, I show in (A.1) that, under the next Assumption, $V_{it}$ controls for $\eta_{it}$ conditional on $A_i$ and $W_i$. Second, I further extend their analysis to allow coordinates of $g$ to depend on multiple elements of $\eta_{it}$ in Subsection 3.3.

**Assumption 3** (**Control for $\boldsymbol{\eta_{it}}$**)**.** *Suppose the following conditions hold:*

(a) $Z_{it} \perp (\eta_{it}, \varepsilon_{it})|\, (A_i, W_i)$.

(b) *For every $(A_i, W_i)$ and each $l = 1, ..., d_X$, $F_{\eta_{it,l}\left|A_i,W_i\right.}(\eta_{it,l}|\, A_i, W_i)$ is strictly increasing in $\eta_{it,l}$.*

Assumption 3 enables construction of a feasible control variable $V_{it}$ for $\eta_{it}$ given $A_i$ and $W_i$. To present it clearly, suppose $d_X = 1$ and recall that $V_{it} = F_{X_{it}|Z_{it},W_i}\left(X_{it}|\, Z_{it}, W_i\right)$. Then,

$$\begin{aligned} V_{it} &= F_{X_{it}|Z_{it},A_i,W_i}\left(X_{it}|\, Z_{it}, A_i, W_i\right) = F_{\eta_{it}|Z_{it},A_i,W_i}\left(\eta_{it}|\, Z_{it}, A_i, W_i\right) \\ &= F_{\eta_{it}|A_i,W_i}\left(\eta_{it}|\, A_i, W_i\right), \end{aligned} \tag{3.16}$$

where the first equality holds by Assumption 2, the second equality holds by Assumption 1 and a change of variable argument, and the last equality holds by Assumption 3(a). Thus, $V_{it}$ uniquely determines $\eta_{it}$ given $A_i$ and $W_i$ by Assumption 3(b). Note that (3.16) relies on Assumption 1; when multiple elements of $\eta_{it}$ appear in one coordinate of $X_{it}$, the one-to-one relationship does not hold in general, and additional structure or assumption is needed to recover (3.16). I discuss them in Subsection 3.3.

Assumption 3(a) is similar to condition (i) of Theorem 1 in Imbens and Newey (2009), except that I further condition on $(A_i, W_i)$. Since $W_i$ can be viewed as summarizing all the time-invariant information about $A_i$ in the data, Assumption 3(a), loosely speaking, requires $Z_{it} \perp (\varepsilon_{it}, \eta_{it}) |\, A_i$, which is already implied by the standard exogeneity condition $Z_{it} \perp (A_i, \varepsilon_{it}, \eta_{it})$. Note that $Z_{it} \perp (A_i, \varepsilon_{it}, \eta_{it})$ is stronger than Assumption 3(a) as the former rules out the possibility that $Z_{it}$ and $A_i$ are correlated.[2] Nonetheless, the unconditional exogeneity condition remains a useful benchmark since it is widely used in applied work.

Assumption 3(a) can be justified in many applications with the usual choice of IVs from the literature. For example, in labor-supply applications, $Z_{it}$ can be county minimum-wage changes or EITC expansions—policy-driven shocks that, conditional on time-invariant ability $A_i$ and its sufficient statistic $W_i$, vary independently of expected sectoral labor-demand shocks $\eta_{it}$ and labor-supply-elasticity shocks $\varepsilon_{it}$. In gasoline-demand estimation, $Z_{it}$ can be the head of household's earned income or gas-tax changes, assumed (conditional on a household fixed effect $A_i$ and $W_i$) to be independent of anticipated near-term income changes $\eta_{it}$ (e.g., an upcoming bonus) and transitory wealth shocks $\varepsilon_{it}$.

For the empirical application of Section 5, I construct the instruments toward BLP-style (Berry, Levinsohn, and Pakes (1995)) cost shifters: leave-one-out, competitor-weighted input-price (wages and interest rates) averages at the industry–city–year level. These instruments are relevant because firms compete locally for labor and capital, so competitors' input prices shift firm $i$'s cost environment and input choices. They are plausibly exogenous to firm $i$'s idiosyncratic productivity shocks conditional on the fixed effect and its control variable because firm $i$'s own input prices are excluded from the construction, so identification comes from competitors' input-price variation rather than firm $i$'s unobserved shocks. Robustness checks in Appendix B provide evidence that these IVs are plausibly valid and satisfy Assumption 3(a).

Assumption 3(b) is analogous to condition (ii) of Theorem 1 in Imbens and Newey (2009). It is mild because it concerns a conditional CDF, which is typically strictly increasing for continuous random variables.

The last assumption concerns the residual variation in $X_{it}$ given $V_{it}$ and $W_i$, which is used to identify the APE and LAR by (3.11).

[2] For instance, Bartik instruments use national industry shocks, weighted by pre-existing local industry shares. The intuition is the national shocks are plausibly "as-good-as-random" conditional on the fixed effects and controls, so the conditional exogeneity is plausible. But the baseline shares can reflect deep local traits—so unconditional exogeneity is harder to defend.

**Assumption 4 (Residual Variation in $\boldsymbol{X_{it}}$).** *There are at least $d_X$ linearly independent points in the support of $X_{it}$ conditional on $(V_{it}, W_i)$ almost surely.*

Assumption 4 ensures that $\mathbb{E}\left[X_{it}X_{it}'\middle| V_{it}, W_i\right]$ in (3.11) is invertible. Thus, the APE and LAR can be identified by (3.11) and the law of iterated expectations.

Assumption 4 is best interpreted as a conditional support (non-degeneracy) requirement: for each $t$, $X_{it}$ must retain sufficient variation conditional on $(V_{it}, W_i)$ to identify the target objects. Requiring residual variation in $X_{it}$ conditional on $V_{it}$ is generally not restrictive (Imbens and Newey (2009)). Essentially, it requires $Z_{it}$ to induce enough variation in $X_{it}$ so that, for each $v \in \mathrm{Supp}(V_{it})$, the set $\{X_{it} : V_{it} = v\}$ is not a singleton. Therefore, it corresponds to the classical IV relevance condition of $\partial g(z, a, \eta)/\partial z \neq 0$.

The restriction on the support of $X_{it}$ in Assumption 4 mainly comes from conditioning on $W_i$, which also concerns Assumption 2. There is an inherent trade-off between Assumptions 2 and 4. To see this, suppose $W_i = (\mathbf{X}_i, \mathbf{Z}_i)$ which satisfies Assumption 2 trivially. However, because $X_{it}$ is included in $W_i$, it makes the support of $X_{it}$ given $W_i$ a singleton. Hence, Assumption 4 does not hold unless $X_{it}$ is a scalar and the point in the support of $X_{it}$ given $W_i$ is nonzero. More generally, including more elements in $W_i$ tends to make Assumption 2 easier to satisfy while making Assumption 4 harder to maintain. Because of this trade-off, I describe next how the conditions used to justify Assumption 2 affect the plausibility of Assumption 4, and give corresponding sufficient conditions.

When justifications for Assumption 2 are used so that $W_i$ only includes averages of $X_{it}$ and/or $Z_{it}$ through time (e.g., Mundlak (1978), Arkhangelsky and Imbens (2022), Liu, Poirier, and Shiu (2025)), Assumption 4 is not restrictive. For instance, suppose $d_X = 2$, $T = 2$, and $W_i = \overline{X}_i$. Conditional on $(V_{it}, W_i)$, the remaining variation in $X_{it}$ is induced by $Z_{it}$ through (1.2) and the relevance condition above. Assumption 4 therefore holds provided the serial dependence in $Z_{it}$ is not so high that the value of $Z_{it}$ (hence of $X_{it}$) for any $t$ is uniquely determined once $(V_{it}, W_i)$ is fixed. In particular, $X_{i1}$ and $X_{i2}$ cannot be perfectly linearly dependent, since otherwise $\mathrm{Supp}(X_{it} \mid \overline{X}_i)$—and *a fortiori* $\mathrm{Supp}(X_{it} \mid V_{it}, \overline{X}_i)$—would be degenerate.

By contrast, when a high-dimensional $W_i$ is used to justify Assumption 2 (e.g., via the exchangeability condition of Altonji and Matzkin (2005)), Assumption 4 becomes more restrictive. Intuitively, a high-dimensional $W_i$ imposes many constraints on the path $(X_{i1}, \ldots, X_{iT})$, so one typically needs longer panel to leave enough feasible configurations and satisfy Assumption 4.

In this case, one possibility is to explore the symmetry through time in the solution to $\{\mathbf{X}_i : V_{it} = v, W_i = w\}$. In particular, when $W_i$ includes averages through time of the polynomials of $(X_{it}, Z_{it})$ up to the $T^{\text{th}}$ order, Assumption 2 is satisfied under a nonparametric exchangeability condition on $f_{\eta_i|A_i}$ by adapting the proof of Altonji and Matzkin (2005) in Proposition 1. Then, since $W_i$ is symmetric in $(X_{it}, Z_{it})$ through time, one may permute the order of $(X_{it}, Z_{it})$ in time without changing the value of $W_i$. This creates enough linearly independent points in the support of $X_{it}$ given $W_i$ when $T \geq d_X$. When such permutation also does not change the value of $V_{it}$ which is usually satisfied under the relevance condition, Assumption 4 is satisfied. I provide an example in Appendix A.3.

Assumption 4 pertains to the inversion-based argument in (3.11). To facilitate the derivative-based argument in (3.12), I introduce the next assumption, which requires more residual variation of $X_{it}$ conditional on $V_{it}$ and $W_i$ than Assumption 4.

**Assumption 4′ (More Variation in $\boldsymbol{X_{it}}$).** *The support of $X_{it}$ conditional on $(V_{it}, W_i)$ contains an open ball of positive radius almost surely.*

Assumption 4′ allows one to perturb $X_{it}$ in $\mathbb{E}[Y_{it}| X_{it}, V_{it}, W_i]$ when $(V_{it}, W_i)$ is fixed so that conditional moments of $\beta_{it}$ can be recovered by (3.12). It is similar to Assumption 2.2 of Altonji and Matzkin (2005).

The same trade-off between Assumptions 2 and 4 also applies to Assumption 4′. When Assumption 2 is justified such that $W_i$ is low-dimensional (e.g., $\overline{X}_i$ as in Mundlak (1978)), Assumption 4′ is generally not restrictive. In essence, $Z_{it}$ must vary enough to generate sufficient residual variation in $X_{it}$ conditional on $(V_{it}, W_i)$.

However, when Assumption 2 is justified using nonparametric arguments (e.g., the exchangeability condition of Altonji and Matzkin (2005)), $d_W$ is large and Assumption 4′ can be restrictive. To address this issue, I present two sets of solutions in Appendix A.3. First, the *unconditional* variation of exogenous regressors outside of $W_i$ can be leveraged. Second, following the suggestions of Altonji and Matzkin (2005), one may restrict how $W_i$ enters $f_{A_i|W_i}$ so that the relevant conditional expectations only concern a subset of $W_i$ or a linear index of its components.

Assumption 4′ allows the straightforward perturbation-based method (3.12) to identify the APE and LAR without requiring the computation of the inverse of $\mathbb{E}\left[X_{it}X_{it}'\middle| V_{it}, W_i\right]$. When residual variation is not a concern—such as when $W_i$ contains only the mean of $X_{it}$ through time or when exogenous regressors are included in (2.1)—Assumption 4′ is preferred for simpler analysis. I compare these two approaches in Remark 2 below.

The next theorem summarizes the main identification result of this paper.

**Theorem 1 (Identification).** *If Assumptions 1–3 and either Assumption 4 or 4′ are satisfied, then the APE $\bar{b} = T^{-1}\sum_{t=1}^{T}\mathbb{E}\beta_{it}$ and the LAR function $b_t(x) = \mathbb{E}\left[\beta_{it}\middle| X_{it} = x\right]$ are both identified.*

### 3.3 Extensions

**Vector-Valued $\boldsymbol{\eta_{it}}$**

I present four methods for incorporating vector-valued $\eta_{it}$ into $X_{it}$ under additional assumptions. To illustrate the idea clearly, I assume $d_X = d_\eta = 2$.

First, timing assumptions on the choice of regressors can be exploited. For instance, in production function estimation, choices of later inputs such as capital typically do not depend on random shocks to earlier chosen inputs such as labor after conditioning on the level of the earlier chosen inputs; see Ackerberg, Caves, and Frazer (2015, (Appendix 1 of 2006 Working Paper)) for a similar idea. Under this timing assumption, I modify (2.1)–(2.2) to be:

$$\begin{aligned} Y_{it} &= X_{it}'\beta\left(A_i, \varepsilon_{it}\right), \\ X_{it,1} &= g_1\left(Z_{it}, A_i, \eta_{it,1}\right), \text{ and } X_{it,2} = g_2\left(X_{it,1}, Z_{it}, A_i, \eta_{it,2}\right). \end{aligned} \tag{3.17}$$

Note that now $X_{it,2}$ depends on the whole vector of $\eta_{it}$ if $X_{it,1} = g_1\left(Z_{it}, A_i, \eta_{it,1}\right)$ is substituted into $g_2$. I first use $V_{it,2} \coloneqq F_{X_{it,2}|X_{it,1},Z_{it},W_i}\left(X_{it,2}|\, X_{it,1}, Z_{it}, W_i\right)$ to control for $\eta_{it,2}$ first. Then, I use $F_{X_{it,1}|Z_{it},W_i}\left(X_{it,1}|\, Z_{it}, W_i\right)$ to control for $\eta_{it,1}$.

Second, interaction across different coordinates of $X_{it}$ can provide useful information about $\eta_{it}$. One possibility is to assume that, although the full vector $\eta_{it}$ enters each coordinate of $X_{it}$, the ratio $X_{it,1}/X_{it,2}$ depends only on a single component of $\eta_{it}$. For example, in production function estimation, $\eta_{it,1}$ represents the Hicks neutral productivity and $\eta_{it,2}$ denotes the labor-augmenting technology. Although both capital and labor choices depend on vector $\eta_{it}$, it is plausible that *capital per worker* depends only on the labor-augmenting technology $\eta_{it,2}$. See Proposition 2.1 of Demirer (2025) for more discussions about this assumption. Given this assumption, I can proceed to control for $\eta_{it,2}$ first by

$$V_{it,2} \coloneqq F_{X_{it,1}/X_{it,2}|Z_{it},W_i}\left(X_{it,1}/X_{it,2}|\, Z_{it}, W_i\right), \tag{3.18}$$

and then for $\eta_{it,1}$ by

$$V_{it,1} \coloneqq F_{X_{it,1}|Z_{it},V_{it,2},W_i}\left(X_{it,1}|\, Z_{it}, V_{it,2}, W_i\right). \tag{3.19}$$

Third, general "global univalence" results—such as Gale and Nikaido (1965) or Palais (1959), Hadamard (1906a,b)—suffice to establish the invertibility required here. In production function estimation, the inverse isotonicity approach in Berry, Gandhi, and Haile (2013) appears promising to obtain global injectivity of $g$ in $\eta_{it}$. In particular, suppose each input choice increases in its own component of $\eta_{it}$, while cross-partials are weakly negative due to substitutability. This happens when, for example, inputs are substitutes in the firm's optimal input demand; a favorable shock to one input in the form of higher efficiency or lower effective cost leads the firm to substitute toward that input and away from others. Then, the Jacobian matrix of $g$ with respect to $\eta_{it}$, denoted by $J_g(\eta_{it})$, has positive diagonal entries and nonpositive off-diagonal entries, i.e., the $Z$-matrix sign pattern required for an $M$-matrix. If, in addition, the own effects are strong enough that the matrix is strictly diagonally dominant, then $J_g(\eta_{it})$ is an $M$-matrix. This, in turn, guarantees a unique monotone inverse mapping from observables back to the vector $\eta_{it}$.

Finally, if $\eta_{it}$ enters the structural function $g$ through a scalar index $\eta_{it}'\tau$ and $g$ is monotone in that index (Ichimura and Lee (1991)), then the effective unobservable is scalar, satisfying Assumption 1. This aligns with common practice in asset pricing, where a scalar market factor emerges as an index of aggregate shocks and affects stock returns monotonically. A similar structure arises in production function estimation, where scalar productivity summarizes richer underlying states and affects input choices strictly increasingly.

### Higher-Order Moments of $\beta_{it}$

I identify the $P^{\text{th}}$-order moments of $\beta_{it}$. For clarity of illustration, suppose the regressor is $X_{it} = (K_{it}, L_{it})'$ and $\beta_{it} = (\beta_{it,K}, \beta_{it,L})'$. The extension to $d_X > 2$ is straightforward. Since the residual variation in $X_{it}$ given $(V_{it}, W_i)$ is driven only by exogenous $Z_{it}$, (3.9) holds for any measurable function $h$ of $\beta_{it}$:

$$\mathbb{E}\left[h(\beta_{it}) \mid X_{it}, V_{it}, W_i\right] = \mathbb{E}\left[h(\beta_{it}) \mid V_{it}, W_i\right]. \tag{3.20}$$

In particular, let $h = (h_1, \ldots, h_P)'$ and $h_q(\beta_{it}) = \beta_{it,K}^q \beta_{it,L}^{P-q}$ for $q = 0, \ldots, P$. By (1.1), the conditional $P^{\text{th}}$-order moment of $Y_{it}$ given $(X_{it}, V_{it}, W_i)$ is

$$\mathbb{E}\left[Y_{it}^P \mid X_{it}, V_{it}, W_i\right] = \sum_{q=0}^{P} \binom{P}{q} K_{it}^q L_{it}^{P-q} \mathbb{E}\left[\beta_{it,K}^q \beta_{it,L}^{P-q} \mid V_{it}, W_i\right]. \tag{3.21}$$

Under $P$-times differentiability of $\mathbb{E}\left[Y_{it}^P \mid X_{it}, V_{it}, W_i\right]$ in $X_{it}$, the coefficients on $K_{it}^P$ and $L_{it}^P$ are identified via differentiation, yielding

$$\mathbb{E}\left[\beta_{it,K}^q \beta_{it,L}^{P-q} \mid V_{it}, W_i\right] = \frac{1}{P!} \frac{\partial^P}{\partial k^q \partial l^{P-q}} \mathbb{E}\left[Y_{it}^P \mid K_{it} = k, L_{it} = l, V_{it}, W_i\right], \text{ for } q = 0, \ldots, P, \tag{3.22}$$

and hence the $P^{\text{th}}$-order moments of $\beta_{it}$ follow by the law of iterated expectations

$$\mathbb{E}[\beta_{it,K}^{q}\beta_{it,L}^{P-q}] = \mathbb{E}\left[\frac{1}{P!}\frac{\partial^P}{\partial k^q \partial l^{P-q}}\mathbb{E}\left[Y_{it}^P \mid K_{it}=k, L_{it}=l, V_{it}, W_i\right]\right], \text{ for } q=0,\ldots,P. \quad (3.23)$$

Repeating this argument for all $P \in \mathbb{N}$ identifies the entire moment sequence of $\beta_{it}$. Under standard moment-determinacy conditions (e.g., Stoyanov (2000)), this sequence uniquely determines the distribution of $\beta_{it}$.

### Exogenous Shocks and Covariates

The identification argument goes through when I include ex-post shocks $\upsilon_{it}$ to $\beta_{it}$ (i.e., $\beta_{it} = \beta(A_i, \varepsilon_{it}, \upsilon_{it})$) and $\widetilde{\epsilon}_{it}$ to $Y_{it}$ (i.e., $Y_{it} = X_{it}'\beta_{it} + \widetilde{\epsilon}_{it}$ with $\mathbb{E}\widetilde{\epsilon}_{it} = 0$), where $\upsilon_{it}$ and $\widetilde{\epsilon}_{it}$ are independent of all other variables. Note that $\beta(\cdot)$ may a priori be time-varying and the joint distribution of $(\varepsilon_{it}, \eta_{it})$ may also vary with $t$. Introducing ex-post shocks $\upsilon_{it}$ provides an additional and economically distinct source of time variation (e.g., unexpected technology shock). While the identification result remains valid with the inclusion of $\upsilon_{it}$, I present the required changes to the proof at the end of the proof of Theorem 1. I also include $\upsilon_{it}$ into $\beta_{it}$ when estimating the APE and LAR and examine its impact via simulations. As for $\widetilde{\epsilon}_{it}$, since it is independent of all other variables and enters (2.1) in an additive way, equations (3.9)–(3.13) hold as before. Thus, the identification analysis is unaffected.

The presence of exogenous covariates in $X_{it}$ can strengthen identification. The key point is that Assumptions 1–3 pertain only to the endogenous covariates of $X_{it}$. In particular, the control variables $V_{it}$ and $W_i$ depend only on the endogenous covariates of $X_{it}$, which alleviates concerns that the controls may be high-dimensional and makes the required residual variation in $X_{it}$ easier to satisfy. Moreover, as discussed in Appendix A.3, unconditional variation in exogenous covariates of $X_{it}$ can be directly exploited to make Assumptions 4 and 4′ easier to satisfy. Details on how to adapt the analysis to incorporate exogenous covariates are provided at the end of the proof of Theorem 1.

# 4 Estimation and Large Sample Theory

In this section, I first describe how to estimate the parameters using three-step series estimators. I then establish the convergence rates of the proposed estimators, and finally show that they are asymptotically normal.

## 4.1 Estimation

The parameters of interest are

$$\bar{b} = T^{-1} \sum_{t=1}^{T} \mathbb{E}\beta_{it} \quad \text{and} \quad b_t(x) \coloneqq \mathbb{E}\left[\beta_{it} \middle| X_{it} = x\right]. \tag{4.1}$$

$\bar{b}$ is the pooled APE across all firms and periods. $b_t(x)$ is the LAR function for a subpopulation characterized by $X_{it} = x$ in period $t$ and is also useful for answering policy-related questions. For example, plugging realizations $x_{it}$ of $X_{it}$ into $b_t(x)$ provides a fine approximation to $\beta_{it}$.

I propose to estimate the parameters in (4.1) with three-step series estimators. For clarity of exposition, I set $d_X = 1$ and highlight the modifications required for $d_X > 1$ when needed. To fix ideas, I follow Liu, Poirier, and Shiu (2025) to let $W_i = T^{-1} \sum_{t=1}^{T} X_{it}$, which can be motivated by generalizing the method of Mundlak (1978).

First, for each $t$, I estimate $V_t(x, z, w) \coloneqq F_{X_{it}|Z_{it},W_i}(x| z, w)$ by regressing $\mathbb{1}\{X_{it} \leq x\}$ on the basis functions $q^{M_1}$ of $(Z_{it}, W_i)$ with trimming function $\tau$:

$$\begin{aligned} \widehat{V}_t(x, z, w) &= \tau\left(\widehat{F}_{X_{it}|Z_{it},W_i}(x| z, w)\right) \\ &= \tau\left(q^{M_1}(z, w)' n^{-1} \widehat{Q}_t^{-1} \sum_{j=1}^{n} q_{jt} \mathbb{1}\{x_{jt} \leq x\}\right) \\ &\eqqcolon \tau\left(q^{M_1}(z, w)' \widehat{\gamma}_t^{M_1}(x)\right), \end{aligned} \tag{4.2}$$

where $\widehat{Q}_t \coloneqq n^{-1} \sum_{i=1}^{n} q_{it} q_{it}'$ and $q_{it} \coloneqq q^{M_1}(z_{it}, w_i)$. Examples of $q^{M_1}$ include power series and spline functions. When $d_X > 1$, I regress $\mathbb{1}\{X_{it,l} \leq x_l\}$ on the basis functions $q^{M_1}$ of $(Z_{it}, W_i)$ with trimming function $\tau$ for each $l$ and obtain $\widehat{V}_t(x, z, w) = \left(\widehat{V}_{t,1}(x, z, w), \ldots, \widehat{V}_{t,d_X}(x, z, w)\right)'$. I highlight two properties of $\widehat{V}_t(x, z, w)$. First, the regression coefficient $\widehat{\gamma}_t^{M_1}(x)$ in (4.2) depends on $x$ because the dependent variable $\mathbb{1}\{X_{it} \leq x\}$ is a function of $x$. This fact causes the convergence rate of $\widehat{V}_t(x, z, w)$ to be slower than the standard rates for series estimators (Imbens and Newey (2009)). Second, a trimming function $\tau$ is needed since I am estimating a conditional CDF. An example of $\tau$ is $\tau(x) = \mathbb{1}\{x > 0\} \cdot \min(x, 1)$.

Next, define $S \coloneqq (X, V, W)$ and let $\mathcal{X}$, $\mathcal{Z}$, $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{S}$ denote the supports of $X$, $Z$, $V$, $W$, and $S$, respectively. Let $V_{it} \coloneqq V_t(X_{it}, Z_{it}, W_i)$ ,$\widehat{V}_{it} \coloneqq \widehat{V}_t(X_{it}, Z_{it}, W_i)$, and $\widehat{v}_{it} \coloneqq \widehat{V}_t(x_{it}, z_{it}, w_i)$. For any $s = (x, v, w) \in \mathcal{S}$, I estimate $G_t(s) \coloneqq \mathbb{E}\left[Y_{it} \middle| S_{it} = s\right]$ by regressing $Y_{it}$ on the basis functions $p^{M_2}$ of $\left(X_{it}, \widehat{V}_{it}, W_i\right)$:

$$\widehat{G}_t\left(s\right) = p^{M_2}\left(s\right)' n^{-1} \widehat{P}_t^{-1} \widehat{p}_t' y_t =: p^{M_2}\left(s\right)' \widehat{\alpha}_t^{M_2}, \tag{4.3}$$

where $\widehat{P}_t := n^{-1} \sum_{i=1}^{n} \widehat{p}_{it} \widehat{p}_{it}'$, $\widehat{p}_{it} := p^{M_2}\left(x_{it}, \widehat{v}_{it}, w_i\right)$, $\widehat{p}_t$ is an $n \times M_2$ matrix with $\widehat{p}_{it}'$ being its $i^{\text{th}}$ row, and $y_t$ is the vector of $y_{it}$s. Following Newey, Powell, and Vella (1999), I let

$$p^{M_2}\left(s\right) = x \otimes p^{m_2}\left(v, w\right) \tag{4.4}$$

by exploiting the index structure of (2.1). Hence, the effective degree that matters for the convergence rate is $m_2$ since $M_2 = d_X \times m_2$ and $d_X$ is finite.

Finally, I exploit the index structure of (2.1) again to estimate $b_{1t}\left(v, w\right) := \mathbb{E}\left[\beta_{it} \middle| V_{it} = v, W_i = w\right]$. When (3.11) is used, I estimate $b_{1t}\left(v, w\right)$ by

$$\widetilde{b}_{1t}\left(v, w\right) = \left(\widehat{\mathbb{E}}\left[X_{it} X_{it}' \middle| \widehat{V}_{it} = v, W_i = w\right]\right)^{-1} \widehat{\mathbb{E}}\left[X_{it} Y_{it} \middle| \widehat{V}_{it} = v, W_i = w\right], \tag{4.5}$$

where $\widehat{\mathbb{E}}\left[X_{it} Y_{it} \middle| \widehat{V}_{it}, W_i\right]$ is obtained by regressing each coordinate of $X_{it} Y_{it}$ on the basis functions of $\left(\widehat{V}_{it}, W_i\right)$ and similarly for $\widehat{\mathbb{E}}\left[X_{it} X_{it}' \middle| \widehat{V}_{it}, W_i\right]$. When (3.12) is used, I estimate $b_{1t}\left(v, w\right)$ by

$$\widehat{b}_{1t}\left(v, w\right) = \partial \widehat{G}_t\left(s\right) / \partial x = \left(I_{d_X} \otimes p^{m_2}\left(v, w\right)\right)' \widehat{\alpha}_t^{M_2} =: \overline{p}^{M_2}\left(s\right)' \widehat{\alpha}_t^{M_2}, \tag{4.6}$$

where the second equality holds by the chain rule. I follow (4.6) in what follows, as it simplifies both implementation and the derivation of asymptotic properties.

*Remark* 2 (**Compare $\widetilde{b}_{1t}\left(v, w\right)$ with $\widehat{b}_{1t}\left(v, w\right)$**). While both approaches are valid, this paper recommends using the derivative-based estimator $\widehat{b}_{1t}(v, w)$ rather than the inversion-based estimator $\widetilde{b}_{1t}(v, w)$ in practice when Assumption $4'$ holds. The derivative-based route operates through the estimated one-dimensional regression function $\widehat{G}_t(x, v, w)$ and obtains $\widehat{b}_{1t}(v, w)$ via the analytic partial derivative with respect to $x$. This approach exploits the linear random-coefficient structure directly, avoids an additional "matrix estimation + inversion" step, and leads to a cleaner and more stable asymptotic argument in the paper's framework. By contrast, $\widetilde{b}_{1t}(v, w)$ requires estimating and inverting the conditional second-moment matrix $\mathbb{E}[X_{it} X_{it}' \mid V_{it} = v, W_i = w]$, whose sample analogue can be ill-conditioned in finite samples and hence numerically unstable (Carrasco, Florens, and Renault (2007)). This inversion also complicates the asymptotic analysis, because one must control how estimation error in the conditional second moment propagates through a matrix inverse.

To estimate $\overline{b}$ and $b_t\left(x\right)$, by the law of iterated expectations I regress $\widehat{b}_{1t}\left(\widehat{V}_{it}, W_i\right)$ on the basis function $r^{M_3}$ of constant one and $X_{it}$, respectively:

$$\widehat{\bar{b}} = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\widehat{b}_{1t}\left(\widehat{v}_{it}, w_i\right) \quad \text{and} \quad \widehat{b}_t\left(x\right) = r^{M_3}\left(x\right)' n^{-1}\widehat{R}_t^{-1} r_t' \widehat{B}_t =: r^{M_3}\left(x\right)' \widehat{\rho}_t^{M_3}, \tag{4.7}$$

where $\widehat{R}_t := n^{-1}\sum_{i=1}^{n} r_{it} r_{it}'$, $r_{it} := r^{M_3}(x_{it})$, $r_t$ is an $n \times M_3$ matrix with $r_{it}'$ being its $i^{\text{th}}$ row, and $\widehat{B}_t$ is an $n \times d_X$ matrix with $\widehat{b}_{1t}(\widehat{v}_{it}, w_i)'$ being its $i^{\text{th}}$ row.

*Remark* 3 (**Computational Considerations**). While the three-step estimators are based on standard series regressions and are typically computationally manageable, two implementation aspects are worth noting for large datasets. First, when estimating $V_t\left(x, z, w\right)$, a key computational simplification is that the regressors—the basis functions $q^{M_1}(Z_{it}, W_i)$—are identical across all thresholds, with only the dependent variable $\mathbb{1}\{X_{it} \leq x\}$ changing in $x$. Hence, I can compute the inverse Gram matrix $\widehat{Q}_t^{-1}$ once and reuse it to obtain the coefficient vector $\widehat{\gamma}_t^{M_1}\left(x\right)$ for every threshold $x_{it}$. This reuse is substantially cheaper than running $n$ fully separate regressions. Second, when estimating $G_t\left(s\right)$, I exploit the linear structure in $X_{it}$ using the separable specification $p^{M_2}\left(s\right) = x \otimes p^{m_2}\left(v, w\right)$, which reduces the effective regressor dimensionality and alleviates the associated computational burden.

## 4.2 Convergence Rates and Asymptotic Normality

Since I let $n \to \infty$ for each fixed $t$ in the asymptotic analysis, the $t$-subscript is suppressed when it is clear. Let $p_i := p^{M_2}\left(X_i, V_i, W_i\right)$ and $P := \mathbb{E} p_i p_i'$. In the analysis, I use results from Imbens and Newey (2009) to establish convergence rates for $\widehat{V}\left(x, z, w\right)$ and $\widehat{G}\left(s\right)$. The following assumption is needed.

**Assumption 5.** *Suppose the following conditions hold:*

(a) $V\left(x, z, w\right)$ *is continuously differentiable of order* $d_1$ *on the support with derivatives uniformly bounded in* $(x, z, w)$ *and* $\mathcal{Z} \times \mathcal{W} \subset \mathbb{R}^{j_1}$.

(b) $p^{m_2}\left(v, w\right) = p^{m_{2v}}\left(v\right) \otimes p^{m_{2w}}\left(w\right)$ *and there exist constants* $C, \theta > 0$ *such that* $\lambda_{\min}\left(P\right) \geq C$ *and*
$$\inf_{w \in \mathcal{W}} f_{V,W}\left(v, w\right) \geq C\left[v\left(1 - v\right)\right]^{\theta}.$$

(c) $G\left(s\right)$ *is continuously differentiable of order* $d_2 > 1$ *on* $\mathcal{S} \subset \mathbb{R}^{j_2}$.

(d) $\mathbb{V}\left(Y | X, Z, W\right)$ *is bounded.*

Let $\zeta\left(m_2\right) := m_{2v}^{\theta} m_2$ and $\zeta_1\left(m_2\right) := m_{2v}^{\theta+2} m_2$. With Assumption 5 in place, the next lemma follows directly from Theorem 12 of Imbens and Newey (2009). Let $\Delta_{1n}^2 := n^{-1}M_1 + M_1^{1-2d_1/j_1}$ and $\Delta_{2n}^2 := \Delta_{1n}^2 + n^{-1}m_2 + m_2^{-2d_2/j_2}$ in the next lemma.

**Lemma 1** (**Convergence Rates of $\widehat{V}$ and $\widehat{G}$**). *If the conditions of Theorem 1 and Assumption 5 are satisfied, and $m_2^2 m_{2v}^{\theta+2}\left(n^{-1}M_1 + M_1^{1-2d_1/j_1}\right) \to 0$, then*

$$\mathbb{E}\left[n^{-1}\sum_i \left(\widehat{V}_i - V_i\right)^2\right] = O\left(\Delta_{1n}^2\right),$$
$$\int \left[\widehat{G}(s) - G(s)\right]^2 dF(s) = O_p\left(\Delta_{2n}^2\right), \quad \textit{and} \quad \sup_{s\in\mathcal{S}}\left|\widehat{G}(s) - G(s)\right| = O_p\left(\zeta(m_2)\Delta_{2n}\right).$$

Lemma 1 states that the mean-square convergence rate for $\widehat{G}(s)$ is the sum of the first-step rate $\Delta_{1n}^2$, the variance term $n^{-1}m_2$, and the squared bias term $m_2^{-2d_2/j_2}$. $d_1/j_1$ and $d_2/j_2$ are the uniform approximation rates that govern how well the unknown functions $V(x,z,w)$ and $G(s)$ can be approximated with $\widehat{V}(x,z,w)$ and $\widehat{G}(s)$, respectively; see Assumptions 3 and 5 of Imbens and Newey (2009) for the details.

Let $\xi_i := b_1(V_i, W_i) - b(X_i)$ and $\xi := (\xi_1, \ldots, \xi_n)'$. To derive the convergence rates for the APE and LAR estimators, I impose the following assumption.

**Assumption 6.** *Suppose the following conditions hold:*

(a) *$b(x)$ is continuously differentiable of order $d_3$ on $\mathcal{X} \subset \mathbb{R}^{j_3}$.*

(b) *There is a constant $C > 0$ and $\zeta(M_3)$, such that for each $M_3$ there exists a non-singular constant matrix $N_r$ such that $\tilde{r}^{M_3}(x) := N_r r^{M_3}(x)$ satisfies $\lambda_{\min}\left(\mathbb{E}\tilde{r}^{M_3}(X_i)\tilde{r}^{M_3}(X_i)'\right) \geq C$ and $\sup_{x\in\mathcal{X}}\left\|\tilde{r}^{M_3}(x)\right\| \leq C\zeta(M_3)$.*

(c) *$\mathbb{E}[\xi\xi'|\mathbf{X}]$ is bounded.*

Assumption 6 imposes conditions on the degree of smoothness of $b(x)$, the normalization of basis functions $r^{M_3}(x)$, and the boundedness of the second moment of $\xi_i$, similar to those in Assumption 5. Since $\widehat{b}_1(v,w)$ and $\widehat{G}(s)$ share the same series regression coefficient $\widehat{\alpha}^{M_2}$, the convergence rate of $\widehat{b}_1(v,w)$ to $b_1(v,w)$ is the same as that of $\widehat{G}(s)$ to $G(s)$. I use this result to prove the convergence rates of $\widehat{\bar{b}}$ and $\widehat{b}(x)$, both of which are unknown but estimable functionals of $\widehat{G}(s)$. Let $\Delta_{3n}^2 := n^{-1}M_3 + M_3^{-2d_3/j_3} + \Delta_{2n}^2$ in the next theorem.

**Theorem 2** (**Convergence Rates of $\widehat{\bar{b}}$ and $\widehat{b}(x)$**). *If the conditions of Lemma 1 and Assumption 6 are satisfied, and $n^{-1}M_3\zeta(M_3)^2 \to 0$, then*

$$\left\|\widehat{\bar{b}} - \bar{b}\right\|^2 = O_p\left(\Delta_{2n}^2\right),$$
$$\int \left\|\widehat{b}(x) - b(x)\right\|^2 dF(x) = O_p\left(\Delta_{3n}^2\right), \quad \textit{and} \quad \sup_{x\in\mathcal{X}}\left\|\widehat{b}(x) - b(x)\right\| = O_p\left(\zeta(M_3)\Delta_{3n}\right).$$

The 2002 working paper version of Imbens and Newey (2009) (henceforth IN02) has obtained asymptotic normality for estimators of known and scalar-valued linear functionals of $G(s)$. I build on their results to analyze vector-valued functionals of $G(s)$ via a Cramér–Wold device and prove asymptotic normality for $\widehat{b}_1(v,w)$. To simplify the notation, I take $\widetilde{\overline{p}}^{M_2}(s)$ and $\widetilde{r}^{M_3}(x)$ that satisfy Assumptions 5 and 6 as $\overline{p}^{M_2}(s)$ and $r^{M_3}(x)$ in what follows.

**Assumption 7.** *Suppose the following conditions hold:*

(a) *There is a constant $C > 0$ and $\zeta(M_1)$, such that for each $M_1$ there exists a non-singular constant matrix $N_q$ such that $\widetilde{q}^{M_1}(z,w) := N_q q^{M_1}(z,w)$ satisfies $\lambda_{\min}\left(\mathbb{E}\widetilde{q}^{M_1}(Z_i,W_i)\,\widetilde{q}^{M_1}(Z_i,W_i)'\right) \geq C$ and $\sup_{(z,w)\in\mathcal{Z}\times\mathcal{W}}\left\|\widetilde{q}^{M_1}(z,w)\right\| \leq C\zeta(M_1)$.*

(b) *$G(s)$ is twice continuously differentiable with bounded first and second derivatives. For functional $a$ of $G$ and some constant $C > 0$, it is true that $|a(G)| \leq C\sup_s|G(s)|$ and either (i) there is $\delta(s)$ and $\widetilde{\alpha}^{M_2}$ such that $\mathbb{E}\delta(S_i)^2 < \infty$, $a\left(p_m^{M_2}\right) = \mathbb{E}\delta(S_i)\,p_m^{M_2}(S_i)$ for all $m = 1,...,M_2$, $a(G) = \mathbb{E}\delta(S_i)\,G(S_i)$, and $\mathbb{E}\left(\delta(S_i) - p^{M_2}(S_i)'\,\widetilde{\alpha}^{M_2}\right)^2 \to 0$; or (ii) for some $\widetilde{\alpha}^{M_2}$, $\mathbb{E}\left[p^{M_2}(S_i)'\,\widetilde{\alpha}^{M_2}\right]^2 \to 0$ and $a\left(p^{M_2\prime}\widetilde{\alpha}^{M_2}\right)$ is bounded away from zero as $M_2 \to \infty$.*

(c) *$\mathbb{E}\left[(Y - G(S))^4\middle| X,Z,W\right]$ is bounded and $\mathbb{V}(Y|\,X,Z,W)$ is bounded away from zero.*

(d) *$nM_1^{1-2d_1/j_1}$, $nM_2^{-2d_2/j_2}$, $nM_3^{-2d_3/j_3}$, $n^{-1}M_1^2M_2\zeta_1(m_2)^2$, $n^{-1}M_1M_3\zeta_1(m_2)^2$, $n^{-1}M_1M_3\zeta(M_3)^2$, $n^{-1}M_1^2\zeta(M_3)^2$, $n^{-1}M_2\zeta(M_3)^2$, $n^{-1}M_1\zeta(M_1)^4\zeta(m_2)^4$, $n^{-1}M_1\zeta(M_1)^4\zeta(M_3)^4$, and $n^{-1}M_1^4\zeta(m_2)^6$ are all $o(1)$.*

(e) *There exist $d_4$ and $\overline{\alpha}^{M_2}$ such that for each element $s_j$ of $s = (x,v,w) \in \mathcal{S} \subset \mathbb{R}^{j_4}$:*

$$\sup\left\{\sup_{s\in\mathcal{S}}\left|G(s) - p^{M_2}(s)'\,\overline{\alpha}^{M_2}\right|, \sup_{s\in\mathcal{S}}\left|\partial\left(G(s) - p^{M_2}(s)'\,\overline{\alpha}^{M_2}\right)/\partial s_j\right|\right\} = O\left(M_2^{-d_4/j_4}\right).$$

*Also, $nM_2^{-2d_4/j_4}$ and $M_1M_2^{-2d_4/j_4}\zeta_1(m_2)^2$ are $o(1)$.*

(f) *(Assumption J(iii) of Andrews (1991)) For a bounded sequence of constants $\{c_{1n} : n \geq 1\}$ and constant pd matrix $\overline{\Omega}_1$, it is true that $c_{1n}\Omega_1 \xrightarrow{p} \overline{\Omega}_1$, where $\Omega_1$ is defined in (A.12).*

Assumption 7(a)–(e) is also imposed by IN02 and is a regularity condition required for the asymptotic normality of $\widehat{b}_1(v,w)$. Assumption 7(f) concerns the asymptotic covariance matrix of $\widehat{b}_1(v,w)$ and is used by Andrews (1991). It guarantees that the normality result of IN02 applies to vector-valued functionals of $G(s)$. Essentially, Assumption 7(f) requires that all the coordinates of $\widehat{b}_1(v,w)$ converge at the same speed, which is mild because ex-ante I do not distinguish any coordinate of $\beta_{it}$ from the others.

**Lemma 2** (**Asymptotic Normality of $\widehat{\boldsymbol{b}}_1(\boldsymbol{v},\boldsymbol{w})$**). *If the conditions of Theorem 2 and Assumption 7 are satisfied, then*

$$\sqrt{n}\,\Omega_1^{-1/2}\left(\widehat{b}_1(v,w)-b_1(v,w)\right)\xrightarrow{d}N(0,I),$$

*where $\Omega_1$ is defined in (A.12).*

*Furthermore,*

$$\widehat{\Omega}_1\Omega_1^{-1}\xrightarrow{p}I,$$

*where $\widehat{\Omega}_1$ is defined in (A.13).*

Lemma 2 concerns $b_1(v,w)$, a *known* functional of $G(s)$. Therefore, the results of IN02 directly apply and I omit its proof in this paper. However, the results of IN02 do not directly apply to $\widehat{\overline{b}}$ and $\widehat{b}(x)$ because $\overline{b}$ and $b(x)$ are *unknown* functionals of $G(s)$. To explain, notice that by the law of iterated expectations

$$\overline{b}=T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[\partial G(S)/\partial X\right],\quad\text{and}\quad b(x)=\mathbb{E}\left[\partial G(S)/\partial X\right|X=x],\tag{4.8}$$

both of which involve integrating $b_1(V,W)=\partial G(S)/\partial X$ with respect to the unknown but estimable distribution of $(V,W)$. Therefore, I need to estimate the unknown functionals in (4.8) and correctly account for the additional estimation bias in the asymptotic analysis.

**Assumption 8.** *Suppose the following conditions hold:*

(a) $\mathbb{E}\overline{p}_i r_i'$ *has full column rank.*

(b) $\mathbb{E}\left[\|\xi\|^4\middle|\mathbf{X}\right]$ *is bounded and* $\mathbb{E}\left[\xi\xi'\middle|\mathbf{X}\right]$ *is bounded away from zero.*

(c) *For a sequence of bounded constants $\{c_{2n},c_{3n}:n\geq 1\}$ and some constant pd matrix $\overline{\Omega}_2$ and $\overline{\Omega}_3$, $c_{2n}\Omega_2\xrightarrow{p}\overline{\Omega}_2$ and $c_{3n}\Omega_3\xrightarrow{p}\overline{\Omega}_3$, where $\Omega_2$ and $\Omega_3$ are defined in (A.18) and (A.64), respectively.*

(d) $\mathbb{E}\left\|b_1(v,w)-\overline{b}\right\|^4<\infty$.

Assumption 8(a) is needed to show that the asymptotic covariance matrix $\Omega_2$ of $\sqrt{n}(\widehat{b}(x)-b(x))$ is positive definite and similarly for $\Omega_3$. Assumption 8(b) is a regularity condition imposed for the Lindeberg–Feller central limit theorem (CLT). Assumption 8(c) is similar to Assumption 7(f) and is needed to prove that the asymptotic normality result holds for vector-valued functionals of $G(s)$ by the Cramér–Wold device.

**Theorem 3** (**Asymptotic Normality of $\widehat{\bar{b}}$ and $\widehat{b}\,(x)$**)**.** *If the conditions of Lemma 2 and Assumption 8 are satisfied, then*

$$\sqrt{n}\,\Omega_2^{-1/2}(\widehat{b}\,(x) - b\,(x)) \xrightarrow{d} N\,(0, I) \quad \text{and} \quad \sqrt{n}\,\Omega_3^{-1/2}(\widehat{\bar{b}} - \bar{b}) \xrightarrow{d} N(0, I),$$

*where $\Omega_2$ and $\Omega_3$ are defined in (A.18) and (A.64), respectively.*

*Furthermore,*

$$\widehat{\Omega}_2\Omega_2^{-1} \xrightarrow{p} I \quad \text{and} \quad \widehat{\Omega}_3\Omega_3^{-1} \xrightarrow{p} I,$$

*where $\widehat{\Omega}_2$ and $\widehat{\Omega}_3$ are defined in (A.48) and (A.67), respectively.*

I present the main idea to the proof of Theorem 3, focusing on how the error propagates through the three-step analysis and the roles of $\widehat{\Omega}_2$ and $\widehat{\Omega}_3$. Define the functionals

$$\begin{aligned}
a\,(b_1, V) &:= \mathbb{E}\,[b_1\,(V, W)|\, X = x] = b\,(x)\,, \\
\widehat{a}\,(b_1, V) &:= \widehat{\mathbb{E}}\,[b_1\,(V, W)|\, X = x]\,, \\
\widehat{a}\left(b_1, \widehat{V}\right) &:= \widehat{\mathbb{E}}\left[b_1\left(\widehat{V}, W\right)\Big|\, X = x\right], \text{ and} \\
\widehat{a}\left(\widehat{b}_1, \widehat{V}\right) &:= \widehat{\mathbb{E}}\left[\widehat{b}_1\left(\widehat{V}, W\right)\Big|\, X = x\right] = \widehat{b}\,(x)\,.
\end{aligned}$$

Then

$$\widehat{b}(x) - b(x) = \underbrace{\widehat{a}(b_1, \widehat{V}) - \widehat{a}(b_1, V)}_{\text{(i) first-step error in } V} + \underbrace{\widehat{a}(\widehat{b}_1, \widehat{V}) - \widehat{a}(b_1, \widehat{V})}_{\text{(ii) second-step error in } b_1} + \underbrace{\widehat{a}(b_1, V) - a(b_1, V)}_{\text{(iii) third-step sampling error}}\,. \tag{4.9}$$

This decomposition makes clear that the asymptotic covariance of $\sqrt{n}\,(\widehat{b}(x) - b(x))$, denoted by $\Omega_2$ as in (A.18), must account for errors of all three sources. In the proof of Theorem 3, I show that

$$\sqrt{n}\,\Omega_2^{-1/2}\Big(\widehat{b}(x) - b(x)\Big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1}^{3}\psi_{ki} + o_p(1), \tag{4.10}$$

where the influence functions $\psi_{1i}$, $\psi_{2i}$, and $\psi_{3i}$ correspond exactly to the three error terms on the right-hand side of (4.9) up to proper normalization. Once (4.10) is established, I verify the Lindeberg-Feller condition on $\Psi_{in} := n^{-1/2}\sum_{k=1}^{3}\psi_{ki}$, which proves its convergence to the standard normal distribution and that $\Omega_2$ is indeed the appropriate asymptotic covariance associated with $\sqrt{n}\,(\widehat{b}(x) - b(x))$.

For feasible inference, a consistent estimator for $\Omega_2$ is needed, and I show in the proof of Theorem 3 that $\widehat{\Omega}_2$ is consistent. Note that $\Omega_2$ may diverge because $\widehat{b}(x)$ converges more slowly than $n^{-1/2}$ as shown in Theorem 2. Hence, all convergence results are expressed in self-normalized way, for example,

$$\sqrt{n}\,\Omega_2^{-1/2}\Big(\widehat{b}(x) - b(x)\Big) \xrightarrow{d} N(0, I) \quad \text{and} \quad \widehat{\Omega}_2\Omega_2^{-1} \xrightarrow{p} I. \tag{4.11}$$

The same argument applies to $\widehat{\Omega}_3$ with the basis function $r^{M_3}(x) = 1$ and the influence function capturing serial correlation of estimating the APE across periods.

# 5 Empirical Application

In this section, I apply my procedure to estimate a heterogeneous Cobb-Douglas production function for each of the five largest manufacturing sectors in China. I obtain unconditional means of the output elasticities (APE) and compare them with those derived using classic methods on the same data set. Furthermore, I estimate conditional means of the elasticities given the regressors (LAR). Results show that there are significant across-firm variations in the output elasticities. I also examine how heterogeneity in output elasticities relates to observed characteristics and obtain intuitive results.

In Appendix B, I conduct several robustness checks—including alternative price deflators, sample trimming, alternative IV and $W_i$ constructions, city–year fixed effects, and comparison with OLS and naive IV estimates—and the results remain stable. In Appendix C, I conduct a production-function motivated simulation study to support the findings in this application.

## 5.1 Data and Methodology

I use the China Annual Survey of Industrial Firms (CASIF), a longitudinal micro-level dataset collected by the National Bureau of Statistics of China that includes information on all state-owned industrial firms and non-state-owned firms with annual sales above 5 million RMB (~US$770K). According to Brandt, Van Biesebroeck, Wang, and Zhang (2017), they account for 91% of the gross output, 71% of employment, 97% of exports, and 91% of total fixed assets in 2004, and thus are representative of industrial activities in China. Many papers on topics such as firm behavior, international trade, and growth theory have used the CASIF data (e.g., Hsieh and Klenow (2009)).

I focus on the five largest 2-digit sectors in terms of the number of firms between 2004 and 2007. Note that the CASIF dataset spans between 1998 and 2007. I choose year 2004 to 2007 to (i) ensure data consistency due to the change in the Chinese Industry Classification codes in 2003, (ii) avoid major structural breaks in the early 2000s (e.g., China joined the WTO in 2001), and (iii) use the most recent data.

Table 1: Summary Statistics

| Variables | N | mean | sd | min | max |
|---|---|---|---|---|---|
| $y_{it} = \ln(\text{value-added output})$ | 45,268 | 9.55 | 1.34 | 2.24 | 16.96 |
| $k_{it} = \ln(\text{capital})$ | 45,268 | 9.17 | 1.56 | 0.98 | 16.84 |
| $l_{it} = \ln(\text{labor})$ | 45,268 | 5.08 | 1.08 | 2.08 | 11.97 |
| $Z_{it,1} = \ln \overline{r}^{\text{real},(-i)}_{cs(i),t}$ | 45,268 | 0.73 | 0.45 | -6.02 | 6.94 |
| $Z_{it,2} = \ln \overline{w}^{\text{real},(-i)}_{cs(i),t}$ | 45,268 | 2.68 | 0.37 | 0.33 | 5.28 |
| Year | 4 | - | - | 2004 | 2007 |
| Firm ID | 11,317 | - | - | - | - |
| Industry Code | 5 | - | - | - | - |

*Notes:*
(i) Output is measured as firm-level real value added in constant 1998 RMB (10,000 yuan). Capital is measured as firm-level real fixed capital stock (book value net of depreciation), deflated to constant 1998 RMB (10,000 yuan). Labor is measured as the number of employees (persons). The real interest rate is defined as real interest expenditures divided by total debt (percent). The real wage is average real salary per worker per year in constant 1998 RMB (10,000 yuan).
(ii) Sample sizes by industry are: textile (3,072), chemicals (2,131), nonmetallic mineral products (1,758), general equipment (2,828), and transportation equipment (1,528).
(iii) $Z_{it,1}$ denotes the natural logarithm of debt-weighted average real interest rate of competitors in the same city–industry–year. $Z_{it,2}$ denotes the natural logarithm of employment-weighted average real wage of competitors in the same city–industry–year.

Following Brandt, Van Biesebroeck, and Zhang (2014), appropriate price deflators for inputs and outputs are applied separately. I preprocess the data so that firms with strictly positive amounts of capital, employment, value-added output, real wage expense, and real interest rate are used for estimation. The final dataset consists of a balanced panel of 11,317 firms over four years across five sectors. Summary statistics are presented in Table 1.

The value-added heterogeneous production function is

$$\begin{aligned} y_{it} &= k_{it}\beta_{it,K} + l_{it}\beta_{it,L} + \omega_{it} + \widetilde{\epsilon}_{it}, \\ \beta_{it,K} &= \beta_K\left(A_i, \varepsilon_{it}, \upsilon_{it}\right), \ \beta_{it,L} = \beta_L\left(A_i, \varepsilon_{it}, \upsilon_{it}\right), \ \omega_{it} = \omega\left(A_i, \varepsilon_{it}, \upsilon_{it}\right), \\ k_{it} &= g_K\left(Z_{it}, A_i, \eta_{it,K}\right), \ l_{it} = g_L\left(Z_{it}, A_i, \eta_{it,L}\right). \end{aligned} \tag{5.1}$$

I highlight two features of model (5.1). First, output elasticities $\beta_{it} := (\beta_{it,K}, \beta_{it,L})'$ are allowed to differ across firms and through time. Second, input choices $X_{it} := (k_{it}, l_{it})'$ can be correlated with $\beta_{it}$ through their dependence on the pairwise correlated random variables $A_i$, $\eta_{it}$, and $\varepsilon_{it}$.

I use leave-one-out industry–city–year cell averages of real wages and real interest rates faced by other firms in the same industry and city, weighted by competitors' employment (for wages) and total debt (for interest rates) as IVs. I write them as $Z_{it} :=$ $(\ln \overline{r}^{\text{real},(-i)}_{cs(i),t}, \ln \overline{w}^{\text{real},(-i)}_{cs(i),t})'$, where $cs(i)$ represents city and sector of firm $i$. These instruments are relevant because firms compete locally for labor and capital, so variation in competitors' cost conditions shifts firm $i$'s effective input prices and hence its optimal input choices. At the same time, they are plausibly exogenous to firm $i$'s idiosyncratic productivity shocks conditional on the fixed effect and its control variable because firm $i$'s own input prices are excluded from the construction; identification therefore relies on competitors' cost variation rather than firm-level wage or interest-rate realizations. In Appendix B, I examine the exogeneity issue by using a more exogenous IV constructed at the provincial level, and the results remain stable.

For estimation, I take $W_i$ to be the mean through time of each coordinate of $X_{it}$ by adapting the argument of Mundlak (1978) nonparametrically. I also examine the effect of further including the mean through time of each coordinate of $Z_{it}$ and the results reported in Table A.3 are similar. For all series estimation steps, I use second-degree polynomial spline basis functions with knots at the median, and a robustness check to this choice is included in Appendix B. First, I estimate each coordinate of $V_{it} := (F_{k_{it}|Z_{it},W_i}(k_{it}|\, Z_{it}, W_i), F_{l_{it}|Z_{it},W_i}(l_{it}|\, Z_{it}, W_i))'$ by regressing $\mathbb{1}(k_{it} \leq k)$ and $\mathbb{1}(l_{it} \leq l)$ on the basis functions of $(Z_{it}, W_i)$, respectively. Next, I estimate $G_{it} := \mathbb{E}[y_{it}|\, X_{it}, V_{it}, W_i]$ by regressing $y_{it}$ on the basis functions of $(X_{it}, \widehat{V}_{it}, W_i)$. Estimation of $b_{1t}(V_{it}, W_i) := \mathbb{E}[\beta_{it}|\, V_{it}, W_i]$ is then obtained by taking the partial derivative of $\widehat{G}_{it}(X_{it}, \widehat{V}_{it}, W_i)$ with respect to $X_{it}$. Finally, I estimate the pooled APE $\bar{b}$ by simply averaging $\widehat{b}_{1t}(\widehat{V}_{it}, W_i)$ over $i$ and $t$. For the LAR, as in (4.7), I regress $\widehat{b}_{1t}(\widehat{V}_{it}, W_i)$ on the basis functions of $X_{it}$ for each $t$, then average over $t$ to obtain $\widehat{\beta}_{i,K}$ and $\widehat{\beta}_{i,L}$ as a measure of heterogeneity in output elasticities across firms.

## 5.2 Results

### Average Output Elasticities

I compare my TERC estimates of average output elasticities with those obtained from OP (Olley and Pakes (1996)), LP (Levinsohn and Petrin (2003)), and ACF (Ackerberg, Caves, and Frazer (2015)) applied to the same dataset. Table 2 contains the estimation results.

The $\widehat{\bar{b}}_K$ estimates from TERC fall within [0.371, 0.463] across the five sectors, broadly consistent with the results obtained from applying OP, LP, and ACF to the same dataset. In contrast, the $\widehat{\bar{b}}_L$ estimates exhibit greater discrepancies across methods. The TERC estimates of $\widehat{\bar{b}}_L$ lie in $[0.311, 0.588]$ across the five sectors. OP's labor elasticity estimates are close to TERC's for the first three sectors but are smaller for the last two. LP seems to produce smaller $\widehat{\bar{b}}_L$ estimates across all sectors, whereas ACF generates larger $\widehat{\bar{b}}_L$ estimates for the general equipment and transportation equipment sectors. It is worth noting that the ACF approach tends to produce wider confidence intervals in certain sectors, which may partly

Table 2: Estimates of Average Output Elasticities

| Textile | OP | LP | ACF | TERC |
|---|---|---|---|---|
| Capital Elasticity | 0.359 | 0.252 | 0.300 | 0.415 |
| 95% CI | [0.300, 0.417] | [0.206, 0.297] | [0.227, 0.373] | [0.393, 0.441] |
| Labor Elasticity | 0.470 | 0.175 | 0.567 | 0.452 |
| 95% CI | [0.442, 0.498] | [0.154, 0.196] | [0.440, 0.695] | [0.415, 0.485] |
| Chemical | OP | LP | ACF | TERC |
| Capital Elasticity | 0.294 | 0.288 | 0.344 | 0.463 |
| 95% CI | [0.223, 0.365] | [0.228, 0.348] | [0.155, 0.533] | [0.430, 0.497] |
| Labor Elasticity | 0.296 | 0.113 | 0.378 | 0.311 |
| 95% CI | [0.253, 0.339] | [0.084, 0.143] | [-0.039, 0.795] | [0.263, 0.361] |
| Nonmetallic Mineral | OP | LP | ACF | TERC |
| Capital Elasticity | 0.697 | 0.311 | 0.236 | 0.371 |
| 95% CI | [0.499, 0.895] | [0.263, 0.358] | [0.038, 0.434] | [0.332, 0.410] |
| Labor Elasticity | 0.353 | 0.071 | 0.601 | 0.311 |
| 95% CI | [0.311, 0.394] | [0.051, 0.091] | [-0.151, 1.354] | [0.255, 0.365] |
| General Equipment | OP | LP | ACF | TERC |
| Capital Elasticity | 0.416 | 0.246 | 0.176 | 0.433 |
| 95% CI | [0.267, 0.565] | [0.202, 0.291] | [0.016, 0.337] | [0.402, 0.469] |
| Labor Elasticity | 0.444 | 0.071 | 0.927 | 0.521 |
| 95% CI | [0.406, 0.482] | [0.053, 0.089] | [0.635, 1.218] | [0.479, 0.564] |
| Transportation Equipment | OP | LP | ACF | TERC |
| Capital Elasticity | 0.523 | 0.281 | 0.217 | 0.425 |
| 95% CI | [0.397, 0.649] | [0.216, 0.346] | [0.052, 0.381] | [0.393, 0.459] |
| Labor Elasticity | 0.523 | 0.137 | 1.042 | 0.588 |
| 95% CI | [0.473, 0.573] | [0.100, 0.173] | [0.771, 1.313] | [0.536, 0.644] |

*Notes:* For TERC, I implement the three-step series estimator as described above. To compute the CI, I use a subsampling method to draw $b = \lfloor 4n_s^{3/4} \rfloor$ firms from each sector without replacement and repeat this 1,000 times. For the other methods, I use the Stata commands (`prodest`, Rovigatti and Mollisi (2018)) for OP and LP and (`acfest`, Manjón and Manez (2016)) for ACF.

reflect numerical instability in the Stata implementation, an issue also discussed by Keiller, de Paula, and Van Reenen (2024). The 95% CIs for TERC estimates are reasonably tight across all sectors.

I consider the estimation results from TERC reasonable based on two pieces of empirical evidence. First, it is well documented in the literature that output elasticities in Cobb-Douglas production function estimation usually lie within $[0, 1]$. Second, using Chinese manufacturing data, Hsieh and Klenow (2009) show that roughly half of output accrues to capital. Under profit maximization with a Cobb-Douglas production function, output elasticities correspond to factor revenue shares. The magnitudes of the estimated capital and labor elasticities can therefore be interpreted through this lens, and my estimates are consistent with this result.

The broad similarity of average elasticities across methods is not unexpected, as OP, LP, ACF, and the TERC approach are all designed to address simultaneity between input choices and unobserved productivity. While OP, LP, and ACF provide consistent estimates under a constant-coefficient production function, TERC adds value by relaxing this restriction and allowing output elasticities to vary across firms while still addressing simultaneity via a control-function approach.

*Remark* 4 (**Addressing Simultaneity Issue**). OP, LP, and ACF address simultaneity via a proxy-variable strategy: they invert a strictly monotone proxy input demand—investment in OP and intermediate inputs in LP/ACF—to recover a scalar productivity term $\omega_{it}$, then impose a Markov-style law of motion for $\omega_{it}$ to form identifying moments. This requires (i) strict monotonicity in $\omega_{it}$ and (ii) that $\omega_{it}$ is the only econometric unobservable entering the proxy demand, which rules out random coefficients that directly affect firms' input choices. In contrast, my baseline TERC model allows input choices to correlate with heterogeneous output elasticities through an arbitrary-dimensional fixed effect $A_i$ and a time-varying shock $\eta_{it}$, with productivity modeled flexibly as $\omega_{it} = \omega(A_i, \varepsilon_{it})$. Identification proceeds via a control-function strategy using a sufficient statistic $W_i$ for $A_i$ and a period-specific control $V_{it}$ for $\eta_{it}$, so that conditional on $(V_{it}, W_i)$ the remaining input variation is driven by exogenous instruments without taking a Markov law for $\omega_{it}$ as a primitive assumption.

### Cross-Firm Heterogeneity in Output Elasticities

Next, I examine the distribution of conditional mean output elasticities given the regressors. These conditional means summarize average elasticities for firm subgroups defined by their capital and labor levels, and thus provide a useful basis for designing more targeted, sector-specific policies. They are also informative about the underlying distribution of true elasticities. For instance, by the law of total variance, the variance of these conditional means

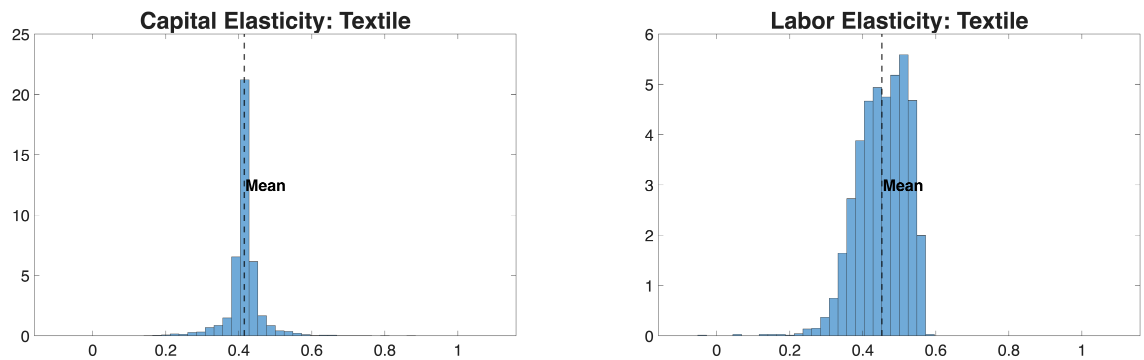


Figure 1: Distribution of Conditional Means of Output Elasticities: Textile

is a lower bound on the variance of the true output elasticities. Demirer (2025) argues that the heterogeneity in output elasticities is largely explained by across-firm variation. To study this issue, I average $\widehat{b}_{t,K}(X_{it})$ and $\widehat{b}_{t,L}(X_{it})$ for each firm through time and denote them by $\widehat{\beta}_{i,K}$ and $\widehat{\beta}_{i,L}$, respectively. These $\widehat{\beta}_i$ estimates can be considered as a proxy for their output elasticities. Then, I plot the histograms of $\widehat{\beta}_{i,K}$ and $\widehat{\beta}_{i,L}$ across firms for each sector.

In Figure 1, I present the histograms of $\widehat{\beta}_{i,K}$ and $\widehat{\beta}_{i,L}$ for the textile sector and indicate the corresponding pooled APE labeled by "Mean" in the figure. The left subplot of Figure 1 corresponds to the histogram of $\widehat{\beta}_{i,K}$. All of its probability mass lies between zero and one, with its mode at 0.4. Over 95% of its probability mass lies between 0.3 and 0.5. The distribution of $\widehat{\beta}_{i,K}$ is concentrated around the mean and symmetric, suggesting that these firms tend to be homogeneous in capital efficiency. The right subplot of Figure 1 is the histogram of $\widehat{\beta}_{i,L}$. Again, the majority of its probability mass lies between zero and one. The distribution of $\widehat{\beta}_{i,L}$ is more dispersed than that of $\widehat{\beta}_{i,K}$ and is slightly left-skewed, with a small number of textile firms exhibiting low labor efficiency.

Figure 2 presents the histograms of $\widehat{\beta}_{i,K}$ and $\widehat{\beta}_{i,L}$ for the other sectors. I draw the following two conclusions. First, for all sectors, the majority of $\widehat{\beta}_{i,K}$ and $\widehat{\beta}_{i,L}$ lie between zero and one, consistent with the empirical evidence in Hsieh and Klenow (2009). Second, there is substantial across-firm variations in both capital and labor elasticities within each sector. The extent of heterogeneity, however, differs across elasticities and sectors. For example, in the left panel of Figure 2, capital elasticity exhibits greater dispersion among chemical firms than among firms in the nonmetallic minerals sector.

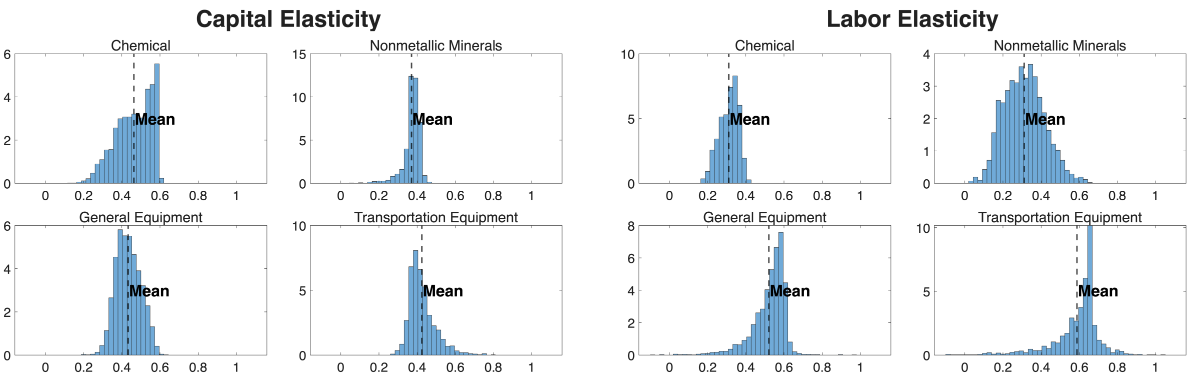


Figure 2: Distribution of Conditional Means of Output Elasticities: Other Sectors

### Explaining the Heterogeneity

I examine cross-firm heterogeneity in output elasticities by regressing the estimated $\widehat{\beta}_{i,K}$ and $\widehat{\beta}_{i,L}$ on firm characteristics. Specifically, using the CASIF dataset, I construct five covariates: firm size, leverage, export status, industrial-park location, and state ownership. All specifications include industry and city fixed effects, and I report heteroskedasticity-robust standard errors.

The results summarized in Table 3 show that larger firms have significantly higher capital elasticities, while leverage is positively associated with labor elasticities. Exporters exhibit modestly higher capital elasticities, and state-owned enterprises have substantially lower labor elasticities. Conditional on these controls, the industrial-park indicator is small. Industry and location fixed effects are jointly significant, indicating that sectoral and regional factors account for an important share of the heterogeneity. The adjusted-$R^2$s are 0.58 and 0.57 for $\widehat{\beta}^K$ and $\widehat{\beta}^L$, respectively, suggesting that over half of the variation in estimated elasticities can be explained by observable firm characteristics and fixed effects.

The results are generally intuitive and broadly consistent with existing empirical evidence. Larger firms, for instance, often operate in more capital-intensive segments or adopt technologies with stronger scale economies, automation, and standardized production processes, so output tends to be more responsive to capital at the margin (Syverson (2011)). Leverage may also be associated with more labor-intensive operating choices—for example, delaying or outsourcing capital investment, relying on older or rented equipment, or placing greater emphasis on variable inputs (Kalemli-Özcan, Laeven, and Moreno (2022)). To the extent that highly leveraged firms substitute away from capital toward labor, the estimated production relationship will place greater weight on labor. Finally, exporters tend to use

Table 3: Explaining the Heterogeneity

| | $\widehat{\beta}^K$ | $\widehat{\beta}^L$ |
|---|---|---|
| Firm Size | 0.0328 | 0.0002 |
| | (0.0006) | (0.0012) |
| Leverage | 0.0004 | 0.0261 |
| | (0.0024) | (0.0044) |
| Export Status | 0.0098 | 0.0023 |
| | (0.0010) | (0.0018) |
| Industrial Park | -0.0014 | -0.0043 |
| | (0.0013) | (0.0026) |
| State Ownership | 0.0050 | -0.0534 |
| | (0.0040) | (0.0076) |

*Notes:*

(i) Size is defined as the natural logarithm of total average assets. Leverage is measured as the debt-to-asset ratio. Exporting firms are defined as those with positive export delivery value. Industrial park firms are identified based on firm address information containing keywords such as “development zone” or “industrial park.” State-owned enterprises (SOEs) are defined as firms whose state capital accounts for more than 50 percent of total paid-in capital.

(ii) The variables are measured as follows: levels for the elasticities; the natural logarithm of 10,000 yuan for firm size; a ratio for leverage; and binary indicators (0/1) for the last three variables in the table.

(iii) The numbers in parentheses in the second row of each cell represent the standard errors.

capital more efficiently (Bernard and Jensen (1999)), whereas state-owned enterprises may employ labor in excess of the efficient level due to political considerations (Wen (2025)), which is consistent with their substantially lower labor elasticities relative to private firms.

# 6 Conclusion

This paper proposes a new TERC model in which regressors are correlated with random coefficients through not only a fixed effect but also a time-varying shock—an empirically relevant feature consistent with optimizing behavior in many applications. I construct feasible control variables for both the fixed effect and the time-varying shock, and use the resulting residual variation in regressors to identify the APE and LAR. I then develop three-step series estimators and establish their convergence rates and asymptotic normality. In an application to Chinese manufacturing, the estimates reveal substantial cross-firm dispersion in output elasticities, part of which is explained by observable firm characteristics.

I propose two directions for future research. First, beyond the low-order moments studied here, policymakers may be interested in the full distribution of random coefficients. One—admittedly demanding—route is to identify moments of all orders by induction and recover the distribution via moment determinacy (Stoyanov (2000)). Second, it remains open whether elements of the approach can be extended to dynamic linear or nonlinear panel models, such as those in Marx, Tamer, and Tang (2024) and Liu, Poirier, and Shiu (2025). Doing so would likely require additional structure governing how lagged outcomes (or state variables) co-move with the random coefficients.

# References

Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification properties of recent production function estimators," *Econometrica*, 83, 2411–2451.

Altonji, J. G. and R. L. Matzkin (2005): "Cross section and panel data estimators for nonseparable models with endogenous regressors," *Econometrica*, 73, 1053–1102.

Andrews, D. W. K. (1991): "Asymptotic normality of series estimators for nonparametric and semiparametric regression models," *Econometrica*, 59, 307–345.

Arellano, M. and S. Bonhomme (2012): "Identifying distributional characteristics in random coefficients panel data models," *Review of Economic Studies*, 79, 987–1020.

Arkhangelsky, D. and G. Imbens (2022): "The role of the propensity score in fixed effect models," Working Paper.

Bai, J. (2009): "Panel data models with interactive fixed effects," *Econometrica*, 77, 1229–1279.

Bernard, A. B. and J. B. Jensen (1999): "Exceptional exporter performance: cause, effect, or both?" *Journal of International Economics*, 47, 1–25.

Berry, S., A. Gandhi, and P. Haile (2013): "Connected substitutes and invertibility of demand," *Econometrica*, 81, 2087–2111.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

Bester, C. A. and C. Hansen (2009): "Identification of marginal effects in a nonparametric correlated random effects model," *Journal of Business & Economic Statistics*, 27, 235–250.

Blundell, R., T. MaCurdy, and C. Meghir (2007): "Labor supply models: unobserved heterogeneity, nonparticipation and dynamics," *Handbook of Econometrics*, 6, 4667–4775.

Brandt, L., J. Van Biesebroeck, L. Wang, and Y. Zhang (2017): "WTO accession and performance of Chinese manufacturing firms," *American Economic Review*, 107, 2784–2820.

Brandt, L., J. Van Biesebroeck, and Y. Zhang (2014): "Challenges of working with the Chinese NBS firm-level data," *China Economic Review*, 30, 339–352.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization," *Econometrica*, 75, 491–533.

Chamberlain, G. (1992): "Efficiency bounds for semiparametric regression," *Econometrica*, 60, 567–596.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81, 535–580.

Cooper, R. W. and J. C. Haltiwanger (2006): "On the nature of capital adjustment costs," *The Review of Economic Studies*, 73, 611–633.

Deaton, A. (1986): "Demand analysis," *Handbook of Econometrics*, 3, 1767–1839.

Demirer, M. (2025): "Production function estimation with factor-augmenting technology: an application to markups," Working Paper.

Gale, D. and H. Nikaido (1965): "The Jacobian matrix and global univalence of mappings," *Mathematische Annalen*, 159, 81–93.

Gao, W. Y. and M. Li (2026): "Identification of semiparametric panel multinomial choice models with infinite-dimensional fixed effects," *The Review of Economics and Statistics*, 1–45.

Gao, W. Y., M. Li, and S. Xu (2023): "Logical differencing in dyadic network formation models with nontransferable utilities," *Journal of Econometrics*, 235, 302–324.

Graham, B. S., J. Hahn, A. Poirier, and J. L. Powell (2018): "A quantile correlated random coefficients panel data model," *Journal of Econometrics*, 206, 305–335.

Graham, B. S. and J. L. Powell (2012): "Identification and estimation of average partial effects in irregular correlated random coefficient panel data models," *Econometrica*, 80, 2105–2152.

Hadamard, J. (1906a): "Sur les transformations planes." *C. R. Acad. Sci., Paris*, 142, 74–77.

——— (1906b): "Sur les transformations ponctuelles." *Bull. Soc. Math. Fr.*, 34, 71–84.

Hausman, J. A. and W. E. Taylor (1981): "Panel data and unobservable individual effects," *Econometrica*, 49, 1377–1398.

Hoderlein, S. and H. White (2012): "Nonparametric identification in nonseparable panel data models with generalized fixed effects," *Journal of Econometrics*, 168, 300–314.

Hogg, R. V., J. W. McKean, and A. T. Craig (2019): *Introduction to mathematical statistics*, Pearson.

Hsiao, C. (2022): *Analysis of panel data*, Cambridge University Press.

Hsiao, C. and M. H. Pesaran (2008): "Random coefficient models," in *The econometrics of panel data*, Springer, 185–213.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, 1403–1448.

Ichimura, H. and L.-F. Lee (1991): "Semiparametric least squares estimation of multiple index models: single equation estimation," in *Nonparametric and semiparametric methods in econometrics and statistics: proceedings of the fifth international symposium in economic theory and econometrics*, Cambridge University Press, 3–49.

Imbens, G. W. and W. K. Newey (2009): "Identification and estimation of triangular simultaneous equations models without additivity," *Econometrica*, 77, 1481–1512.

Kalemli-Özcan, Ş., L. Laeven, and D. Moreno (2022): "Debt overhang, rollover risk, and corporate investment: evidence from the European crisis," *Journal of the European Economic Association*, 20, 2353–2395.

Keiller, A. N., Á. de Paula, and J. Van Reenen (2024): "Production function estimation using subjective expectations data," Working Paper.

Laage, L. (2024): "A correlated random coefficient panel model with time-varying endogeneity," *Journal of Econometrics*, 242, 105804.

Levinsohn, J. and A. Petrin (2003): "Estimating production functions using inputs to control for unobservables," *Review of Economic Studies*, 70, 317–341.

Li, D., J. Chen, and J. Gao (2011): "Non-parametric time-varying coefficient panel data models with fixed effects," *The Econometrics Journal*, 14, 387–408.

Li, D., J. Qian, and L. Su (2016): "Panel data models with interactive fixed effects and multiple structural breaks," *Journal of the American Statistical Association*, 111, 1804–1819.

Li, T. and Y. Sasaki (2024): "Identification of heterogeneous elasticities in gross-output production functions," *Journal of Econometrics*, 238, 105637.

Liu, L., A. Poirier, and J.-L. Shiu (2025): "Identification and estimation of partial effects in nonlinear semiparametric panel models," *Journal of Econometrics*, 252, 105860.

Manjón, M. and J. Manez (2016): "Production function estimation in Stata using the Ackerberg–Caves–Frazer method," *Stata Journal*, 16, 900–916.

Manski, C. F. (1987): "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica*, 55, 357–362.

Marx, P., E. Tamer, and X. Tang (2024): "Heterogeneous intertemporal treatment effects via dynamic panel data models," Working Paper.

Masten, M. A. (2018): "Random coefficients on endogenous variables in simultaneous equations models," *Review of Economic Studies*, 85, 1193–1250.

Moon, H. R. and M. Weidner (2015): "Linear regression for panel with unknown number of factors as interactive fixed effects," *Econometrica*, 83, 1543–1579.

Mundlak, Y. (1978): "On the pooling of time series and cross section data," *Econometrica*, 46, 69–85.

Murtazashvili, I. and J. M. Wooldridge (2008): "Fixed effects instrumental variables estimation in correlated random coefficient panel data models," *Journal of Econometrics*, 142, 539–552.

Nagasawa, K. (2024): "Treatment effect estimation with noisy conditioning variables," Working Paper, Department of Economics, University of Warwick.

Newey, W. K. (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79, 147–168.

Newey, W. K., J. L. Powell, and F. Vella (1999): "Nonparametric estimation of triangular simultaneous equations models," *Econometrica*, 67, 565–603.

Olley, G. S. and A. Pakes (1996): "The dynamics of productivity in the telecommunications equipment industry," *Econometrica*, 64, 1263–1297.

Palais, R. S. (1959): "Natural operations on differential forms," *Transactions of the American Mathematical Society*, 92, 125–141.

Pesaran, M. H. (2006): "Estimation and inference in large heterogeneous panels with a multifactor error structure," *Econometrica*, 74, 967–1012.

Rovigatti, G. and V. Mollisi (2018): "Theory and practice of total-factor productivity estimation: the control function approach using Stata," *Stata Journal*, 18, 618–662.

Stoyanov, J. (2000): "Krein condition in probabilistic moment problems," *Bernoulli*, 6, 939–949.

Su, L., Z. Shi, and P. C. Phillips (2016): "Identifying latent structures in panel data," *Econometrica*, 84, 2215–2264.

Su, L., X. Wang, and S. Jin (2019): "Sieve estimation of time-varying panel data models with latent structures," *Journal of Business & Economic Statistics*, 37, 334–349.

Syverson, C. (2011): "What determines productivity?" *Journal of Economic Literature*, 49, 326–365.

Wang, Y., P. C. Phillips, and L. Su (2024): "Panel data models with time-varying latent group structures," *Journal of Econometrics*, 240, 105685.

Wen, J. Y. (2025): "State employment as a strategy of autocratic control in China," *Review of Economics and Statistics*, 1–45.

Weyl, H. (1939): *The classical groups: their invariants and representations*, Princeton University Press.

Wooldridge, J. M. (2005): "Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models," *Review of Economics and Statistics*, 87, 385–390.

——— (2009): "On estimating firm-level production functions using proxy variables to control for unobservables," *Economics Letters*, 104, 112–114.

——— (2010): *Econometric analysis of cross section and panel data*, MIT press.

——— (2019): "Correlated random effects models with unbalanced panels," *Journal of Econometrics*, 211, 137–150.

# Appendix

Appendix A provides proofs for all theorems and lemmas in the main text, along with additional theoretical results. Appendix B reports additional empirical results. Appendix C presents the simulation results.

## A Proofs and Supplementary Theoretical Results

Appendix A.1 contains proofs of all the lemmas and theorems in the main text. Appendix A.2 introduces Proposition 1, which provides sufficient conditions for Assumption 2 by adapting the nonparametric exchangeability condition of Altonji and Matzkin (2005). Appendix A.3 further discusses Assumptions 2, 4, and 4′.

### A.1 Proofs

***Proof of Theorem 1.*** I first prove that $V_{it}$ is a control for $\eta_{it}$ given $A_i$ and $W_i$. Then, I show $\mathbb{E}\left[\beta_{it}\middle| X_{it}, V_{it}, W_i\right]$ does not depend on $X_{it}$ via the law of iterated expectations. Finally, I identify $\mathbb{E}\beta_{it}$ and $\mathbb{E}\left[\beta_{it}\middle| X_{it}\right]$ by leveraging the residual variation in $X_{it}$ given $V_{it}$ and $W_i$. I present how the inclusion of exogenous shocks $\upsilon_{it}$ and $\widetilde{\epsilon}_{it}$ as well as exogenous coordinates in the regressors affects the analysis at the end of this proof.

Without loss of generality, I assume that $d_\eta = d_X$ and that each coordinate of $\eta_{it}$ enters the corresponding coordinate of $X_{it}$.[3] By Assumption 2, I have $A_i \perp (X_{it}, Z_{it})\middle| W_i$, which implies $X_{it} \perp A_i\middle| (Z_{it}, W_i)$. Thus, for each $l \in \{1, \ldots, d_X\}$ and any on-support $(x_l, z, a, w)$, I have

$$\begin{aligned}
& F_{X_{it,l}\middle|Z_{it},W_i}\left(x_l\middle| z, w\right) \\
&= F_{X_{it,l}\middle|Z_{it},A_i,W_i}\left(x_l\middle| z, a, w\right) \\
&= \mathbb{P}\left(g_l\left(z, a, \eta_{it,l}\right) \le x_l\middle| Z_{it} = z, A_i = a, W_i = w\right) \\
&= \mathbb{P}\left(\eta_{it,l} \le g_l^{-1}\left(x_l, z, a\right)\middle| A_i = a, W_i = w\right) \\
&= F_{\eta_{it,l}\middle|A_i,W_i}\left(g_l^{-1}\left(x_l, z, a\right)\middle| a, w\right),
\end{aligned} \tag{A.1}$$

where the first equality holds by $X_{it} \perp A_i\middle| (Z_{it}, W_i)$, the second uses (2.2), the third holds by Assumptions 1 and 3(a), and the last holds by definition. By (2.2), the random variable

[3] This is without loss of generality because under Assumption 1, I can always redefine $\widetilde{\eta}$ to be the vector that collects the coordinates of $\eta$ that enters each coordinate of $g$ function.

$\eta_{it,l} = g_l^{-1}(X_{it,l}, Z_{it}, A_i)$ for each $l$, so that plugging in gives

$$V_{it,l} \coloneqq F_{X_{it,l}|Z_{it},W_i}\left(X_{it,l}|\, Z_{it}, W_i\right) = F_{\eta_{it,l}|A_i,W_i}\left(\eta_{it,l}|\, A_i, W_i\right), \tag{A.2}$$

which establishes a one-to-one mapping to $\eta_{it,l}$ given $A_i$ and $W_i$ under Assumption 3(b). Let $V_{it} \coloneqq (V_{it,1}, \ldots, V_{it,d_X})'$. By (A.2), $V_{it}$ uniquely determines the vector of $\eta_{it}$ given $A_i$ and $W_i$.

Next, I have

$$\begin{aligned}
&\mathbb{E}\left[\beta_{it}|\, X_{it}, A_i, V_{it}, W_i\right] \\
&= \mathbb{E}\left[\beta(A_i, \varepsilon_{it})|\, g(Z_{it}, A_i, \eta_{it}), A_i, V_{it}, W_i\right] \\
&= \mathbb{E}\left[\beta(A_i, \varepsilon_{it})|\, A_i, V_{it}, W_i\right] \\
&\eqqcolon b_2(A_i, V_{it}, W_i),
\end{aligned} \tag{A.3}$$

where the second equality holds because conditioning on $(A_i, V_{it}, W_i)$ is equivalent to fixing $(A_i, \eta_{it}, W_i)$ by (A.2), and thus the residual variation in $X_{it}$ is driven solely by $Z_{it}$ which is independent of $\varepsilon_{it}$ given $(A_i, \eta_{it}, W_i)$ by Assumption 3(a). I exclude $X_{it}$ from the conditioning set of $\mathbb{E}[\beta_{it}|\, X_{it}, A_i, V_{it}, W_i]$ and write it out as $b_2(A_i, V_{it}, W_i)$.

Then, by (A.3), I have

$$\begin{aligned}
&\mathbb{E}\left[\beta_{it}|\, X_{it}, V_{it}, W_i\right] \\
&= \mathbb{E}\left[b_2(A_i, V_{it}, W_i)|\, X_{it}, V_{it}, W_i\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[b_2(A_i, V_{it}, W_i)|\, X_{it}, Z_{it}, V_{it}, W_i\right]|\, X_{it}, V_{it}, W_i\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[b_2(A_i, V_{it}, W_i)|\, X_{it}, Z_{it}, W_i\right]|\, X_{it}, V_{it}, W_i\right] \\
&= \mathbb{E}\left[\int b_2(a, V_{it}, W_i)\, f_{A_i|X_{it},Z_{it},W_i}(a|\, X_{it}, Z_{it}, W_i)\, \mu(da)\right| X_{it}, V_{it}, W_i\Big] \\
&= \mathbb{E}\left[\int b_2(a, V_{it}, W_i)\, f_{A_i|W_i}(a|\, W_i)\, \mu(da)\right| X_{it}, V_{it}, W_i\Big] \\
&\eqqcolon b_1(V_{it}, W_i),
\end{aligned} \tag{A.4}$$

where the first and second equalities hold by the law of iterated expectations, the third holds because $V_{it}$ is a measurable function of $X_{it}$, $Z_{it}$, and $W_i$, all of which are also conditioned on, and the fifth equality holds by $(X_{it}, Z_{it}) \perp A_i|\, W_i$ by Assumption 2.

Finally, given (A.4), I have

$$\mathbb{E}\left[Y_{it}|\, X_{it}, V_{it}, W_i\right] = X_{it}' b_1(V_{it}, W_i). \tag{A.5}$$

When Assumption 4 is satisfied, I pre-multiply both sides of (A.5) by $X_{it}$ and take conditional expectation of both sides conditioning on $V_{it}$ and $W_i$:

$$\mathbb{E}\left[\mathbb{E}\left[X_{it}Y_{it}\right|X_{it},V_{it},W_i\right]\left|V_{it},W_i\right]=\mathbb{E}\left[X_{it}X_{it}'\right|V_{it},W_i\right]b_1\left(V_{it},W_i\right),$$

which identifies $b_1\left(V_{it},W_i\right)$

$$b_1\left(V_{it},W_i\right)=\left(\mathbb{E}\left[X_{it}X_{it}'\right|V_{it},W_i\right])^{-1}\mathbb{E}\left[X_{it}Y_{it}\right|V_{it},W_i\right].$$

When Assumption 4′ is used, I take partial derivative of both sides of (A.5) with respect to $X_{it}$ and obtain

$$b_1\left(V_{it},W_i\right)=\partial\mathbb{E}\left[Y_{it}\right|X_{it},V_{it},W_i\right]/\partial X_{it}.$$

Given $b_1\left(V_{it},W_i\right)$, I use the law of iterated expectations to identify $\bar{b}$ and $b(X_{it})$ by

$$\bar{b}=T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[b_1\left(V_{it},W_i\right)\right]\text{ and }b\left(X_{it}\right)=\mathbb{E}\left[b_1\left(V_{it},W_i\right)\right|X_{it}\right],\tag{A.6}$$

respectively.

When ex-post shock $\widetilde{\epsilon}_{it}$ is included additively in (2.1) and $\mathbb{E}\widetilde{\epsilon}_{it}=0$, equations (A.4) and (A.5) still hold, so the identification result (A.6) holds without any changes. When ex-post shock $\upsilon_{it}$ is also included in $\beta_{it}\coloneqq\beta\left(A_i,\varepsilon_{it},\upsilon_{it}\right)$, since $b(X_{it})$ is already a time-varying function of $X_{it}$ due to the distribution of $\left(\varepsilon_{it},\eta_{it}\right)$ being possibly time-varying, adding $\upsilon_{it}$ which is independent of everything else does not affect the identification proof.

Finally, when exogenous regressors are included in (2.1), I let $X_{it}=\left(U_{it}',Z_{it,1}'\right)'$ and $Z_{it}=\left(Z_{it,1}',Z_{it,2}'\right)'$, and rewrite model (2.1)–(2.2) to be

$$\begin{aligned}Y_{it}&=X_{it}'\beta\left(A_i,\varepsilon_{it}\right),\\U_{it}&=g\left(Z_{it},A_i,\eta_{it}\right).\end{aligned}$$

Then, I replace $X_{it}$ by $U_{it}$ in Assumptions 1–3 as well as in the definition of $V_{it}$, and the analysis goes through as before. Note that I keep $X_{it}$ in Assumption 4 or 4′ because I need to perturb the whole $X_{it}$ vector instead of just endogenous coordinates $U_{it}$ in (A.5) to identify $b_1\left(V_{it},W_i\right)$. As discussed in Appendix A.3, the residual variation requirement of Assumption 4 or 4′ is easier to be satisfied when exogenous $Z_{it,1}$ is included in $X_{it}$ as the support of $Z_{it,2}$ is unaffected by $V_{it}$ and $W_i$. □

***Proof of Theorem 2.*** I denote $\sum_{i=1}^{n}$ by $\sum_i$ and omit all $t$-subscripts. As the result for a finite-dimensional vector-valued $\beta$ can be established by proving for each of its coordinates and combining the results using the triangle inequality, I assume $\beta$ is a scalar in this proof.

I focus on $\widehat{b}(x)$, since the result for $\widehat{\widehat{b}}$ follows immediately by setting $r^{M_3} \equiv 1$. Let $p_i^{m_2} := p^{m_2}(V_i, W_i)$, $r_i := r^{M_3}(X_i)$, and $\overline{p}_i = \overline{p}^{M_2}(S_i)$. Following Imbens and Newey (2009), I can normalize $\mathbb{E} p_i^{m_2} p_i^{m_2\prime} = I_{m_2}$ by Assumption 5 and $R := \mathbb{E} r_i r_i' = I_{M_3}$ by Assumption 6, which imply $\overline{P} := \mathbb{E}\overline{p}_i\overline{p}_i' = \mathbb{E}\left[I_{d_X} \otimes (p_i^{m_2} p_i^{m_2\prime})\right] = I_{M_2}$. Furthermore, I have $\lambda_{\min}\left(\widehat{R}\right) \geq C > 0$ with probability approaching one by Newey (1997). Let $\widetilde{B} := (b_1(\widehat{v}_1, w_1), \ldots, b_1(\widehat{v}_n, w_n))'$.

By (4.7),

$$\begin{aligned} & \left\| n\widehat{R}^{1/2}\left(\widehat{\rho}^{M_3} - \rho^{M_3}\right)\right\|^2 / 4 \\ \leq\ & \left(\widehat{B} - \widetilde{B}\right)' r\widehat{R}^{-1} r' \left(\widehat{B} - \widetilde{B}\right) + \left(\widetilde{B} - B\right)' r\widehat{R}^{-1} r' \left(\widetilde{B} - B\right) \\ & + \left(B - B^X\right)' r\widehat{R}^{-1} r' \left(B - B^X\right) + \left(B^X - R\rho^{M_3}\right)' r\widehat{R}^{-1} r' \left(B^X - r\rho^{M_3}\right), \end{aligned} \tag{A.7}$$

where $B := (b_1(v_1, w_1), \ldots, b_1(v_n, w_n))'$ and $B^X := (b(x_1), \ldots, b(x_n))'$. Hence, $\xi = B - B^X$. I analyze the RHS of (A.7) term by term.

By Lemma S.5 of Imbens and Newey (2009), $\left\| n^{-1}\sum_i \widehat{\overline{p}}_i \widehat{\overline{p}}_i' - I\right\| = o_p(1)$. Then,

$$\begin{aligned} & n^{-2}\left(\widehat{B} - \widetilde{B}\right)' r\widehat{R}^{-1} r' \left(\widehat{B} - \widetilde{B}\right) \leq Cn^{-1}\left(\widehat{B} - \widetilde{B}\right)'\left(\widehat{B} - \widetilde{B}\right) \\ = & \ Cn^{-1}\sum_i \left(\widehat{\overline{p}}_i'\left(\widehat{\alpha}^{M_2} - \alpha^{M_2}\right) + \left(\widehat{\overline{p}}_i'\alpha^{M_2} - b_1(\widehat{v}_i, w_i)\right)\right)^2 \\ \leq & \ C\left\|\widehat{\alpha}^{M_2} - \alpha^{M_2}\right\|^2 + C\sup_{s\in\mathcal{S}}\left\|\overline{p}^{M_2}(s)'\alpha^{M_2} - b_1(v, w)\right\|^2 = O_p\left(\Delta_{2n}^2\right), \end{aligned} \tag{A.8}$$

where the first inequality holds because $n^{-1} r\widehat{R}^{-1} r'$ is idempotent, the last inequality holds by the Cauchy–Schwarz inequality (CS) and $\left\| n^{-1}\sum_i \widehat{\overline{p}}_i \widehat{\overline{p}}_i' - I\right\| = o_p(1)$, and the last equality uses Lemma 1.

Next,

$$\begin{aligned} n^{-2}\left(\widetilde{B} - B\right)' r\widehat{R}^{-1} r' \left(\widetilde{B} - B\right) & \leq Cn^{-1}\sum_i \left(b_1(\widehat{v}_i, w_i) - b_1(v_i, w_i)\right)^2 \\ & \leq Cn^{-1}\sum_i \left(\widehat{v}_i - v_i\right)^2 = O_p\left(\Delta_{1n}^2\right), \end{aligned} \tag{A.9}$$

where the last inequality holds by the mean value theorem and Assumption 5 and the equality holds by Lemma 1 and Markov's inequality.

Finally, for the last two terms on the right-hand side of (A.7),

$$\begin{aligned} & n^{-2}\mathbb{E}\left[\left(B - B^X\right)' r\widehat{R}^{-1} r' \left(B - B^X\right)\Big| \mathbf{X}\right] \\ = & \ n^{-2}\mathrm{tr}\left\{\mathbb{E}\left[\xi' r\widehat{R}^{-1} r'\xi\Big| \mathbf{X}\right]\right\} = n^{-2}\mathrm{tr}\left\{\mathbb{E}\left[\xi\xi' | \mathbf{X}\right] r\widehat{R}^{-1} r'\right\} \end{aligned}$$

$$\leq n^{-2}\text{tr}\left\{CIr\widehat{R}^{-1}r'\right\} = Cn^{-1}\text{tr}\left\{\widehat{R}^{-1}\widehat{R}\right\} = Cn^{-1}M_3,$$

and

$$n^{-2}\left(B^X - R\rho^{M_3}\right)' r\widehat{R}^{-1}r'\left(B^X - R\rho^{M_3}\right) \leq n^{-1}\left\|B^X - R\rho^{M_3}\right\|^2 = O_p\left(M_3^{-2d_3/j_3}\right). \quad \text{(A.10)}$$

By $\lambda_{\min}\left(\widehat{R}\right) \geq C > 0$ and conditional Markov's (CM) inequality (i.e., $\mathbb{E}\left[|Y_n|\,|\, Z_n\right] = O_p(r_n)$ implies $Y_n = O_p(r_n)$),

$$\left\|\widehat{\rho}^{M_3} - \rho^{M_3}\right\|^2 = O_p\left(\Delta_{2n}^2 + n^{-1}M_3 + M_3^{-2d_3/j_3}\right) =: O_p\left(\Delta_{3n}^2\right), \quad \text{(A.11)}$$

which yields

$$\begin{aligned}\int\left\|\widehat{b}(x) - b(x)\right\|^2 dF(x) &\leq \int\left(r^{M_3}(x)'\left(\widehat{\rho}^{M_3} - \rho^{M_3}\right) + \left(r^{M_3}(x)'\rho^{M_3} - b(x)\right)\right)^2 dF(x)\\ &\leq 2\left\|\widehat{\rho}^{M_3} - \rho^{M_3}\right\|^2 + 2\sup_{x\in\mathcal{X}}\left|b(x) - r^{M_3}(x)'\rho^{M_3}\right|^2 = O_p\left(\Delta_{3n}^2\right),\end{aligned}$$

and

$$\begin{aligned}\sup_{x\in\mathcal{X}}\left\|\widehat{b}(x) - b(x)\right\| &\leq \sup_{x\in\mathcal{X}}\left\|r^{M_3}(x)\right\|\left\|\widehat{\rho}^{M_3} - \rho^{M_3}\right\| + \sup_{x\in\mathcal{X}}\left|b(x) - r^{M_3}(x)'\rho^{M_3}\right|\\ &= O_p\left(\zeta(M_3)\Delta_{3n}\right).\end{aligned}$$

□

***Proof of Lemma 2.*** Define

$$\begin{aligned}\Omega_1 &:= \overline{p}^{M_2}(v,w)' P^{-1}(\Sigma + \Sigma_1) P^{-1}\overline{p}^{M_2}(v,w),\ \Sigma := \mathbb{E}p_i p_i' u_i^2, \quad \text{(A.12)}\\ \Sigma_1 &:= \mathbb{E}\overline{\mu}_i^I\overline{\mu}_i^{I\prime},\ p_i := p^{M_2}(S_i),\ q_i := q^{M_1}(X_i, Z_i, W_i),\\ \overline{\mu}_i^I &:= \mathbb{E}\left[G_V(S_j)\tau_V(V_j)p_j q_j' Q^{-1}q_i v_{ji}\middle|\mathcal{I}_i\right],\ u_i := Y_i - G(S_i),\\ G_V(S_j) &:= \partial G(s)/\partial v|_{s=S_j},\ \text{and}\ v_{ji} := \mathbb{1}\{x_i \leq x_j\} - F(x_j|\, z_i, w_i).\end{aligned}$$

and

$$\begin{aligned}\widehat{\Omega}_1 &:= \overline{p}^{M_2}(v,w)'\widehat{P}^{-1}\left(\widehat{\Sigma} + \widehat{\Sigma}_1\right)\widehat{P}^{-1}\overline{p}^{M_2}(v,w), \quad \text{(A.13)}\\ \widehat{\Sigma} &:= n^{-1}\sum_{i=1}^n \widehat{p}_i\widehat{p}_i'\left(y_i - \widehat{G}(\widehat{s}_i)\right)^2,\ \widehat{\Sigma}_1 := n^{-1}\sum_{i=1}^n \widehat{\overline{\mu}}_i^I\widehat{\overline{\mu}}_i^{I\prime},\\ \widehat{\overline{\mu}}_i^I &:= n^{-1}\sum_{j=1}^n \widehat{G}_V(\widehat{s}_j)\widehat{p}_j q_j'\widehat{Q}^{-1}q_i\widehat{v}_{ji},\ \text{and}\ \widehat{v}_{ji} := \left(\mathbb{1}\{x_i \leq x_j\} - \widehat{F}(x_j|\, z_i, w_i)\right).\end{aligned}$$

IN02 have proved asymptotic normality for known and scalar-valued functionals of $G(s)$.

I apply their results to $c'\widehat{b}_1(v,w)$ for any constant vector $c'c = 1$ and obtain

$$\begin{aligned} c'\sqrt{n}\Omega_1^{-1/2}\left(\widehat{b}_1(v,w) - b_1(v,w)\right) &\xrightarrow{d} N(0,1) \text{ and} \\ (c'\Omega_1 c)^{-1}\left[c'\left(\widehat{\Omega}_1 - \Omega_1\right)c\right] &\xrightarrow{p} 0. \end{aligned} \tag{A.14}$$

By (A.14) and Assumption 7(f),

$$c'\left(c_{1n}\widehat{\Omega}_1 - c_{1n}\Omega_1\right)c \xrightarrow{p} 0, \tag{A.15}$$

which implies

$$c_{1n}\widehat{\Omega}_1 \xrightarrow{p} \overline{\Omega}_1. \tag{A.16}$$

Then,

$$\begin{aligned} &\sqrt{n}\widehat{\Omega}_1^{-1/2}\left(\widehat{b}_1(v,w) - b_1(v,w)\right) \\ =\ &\left(c_{1n}\widehat{\Omega}_1\right)^{-1/2}(c_{1n}\Omega_1)^{1/2}\sqrt{n}\Omega_1^{-1/2}\left(\widehat{b}_1(v,w) - b_1(v,w)\right) \\ \xrightarrow{d}\ &\overline{\Omega}_1^{-1/2}\overline{\Omega}_1^{1/2}N(0,I) =_d N(0,I), \end{aligned} \tag{A.17}$$

where the convergence holds by (A.14), the Cramér–Wold device, (A.16), Assumption 7, and the continuous mapping theorem. □

***Proof of Theorem 3.*** Similarly to the proof of Lemma 2, by the Cramér–Wold device it suffices to consider the case when $b(x)$ is a scalar. First, I derive the influence functions for $\widehat{b}(x)$ that correctly account for the estimation errors from each step and prove its asymptotic normality. Then, I show consistency for the estimator of the variance of $\widehat{b}(x)$. I write $r^{M_3}(x)$ as $r(x)$ when there is no confusion. The proof relies on the results from IN02, and I point out differences.

Define

$$\begin{aligned} \Omega_2 &\coloneqq \Omega_{21} + \Omega_{22}, \text{ where} \\ \Omega_{21} &\coloneqq \mathbb{E}\left(A_1 P^{-1} p_i u_i\right)\left(A_1 P^{-1} p_i u_i\right)', \\ \Omega_{22} &\coloneqq \mathbb{E}\left(A_1 P^{-1}\overline{\mu}_i^I - r(x)'\left(\overline{\mu}_i^{II} + r_i\xi_i\right)\right)\left(A_1 P^{-1}\overline{\mu}_i^I - r(x)'\left(\overline{\mu}_i^{II} + r_i\xi_i\right)\right)', \\ A_1 &\coloneqq r(x)'\,\mathbb{E}r_i\overline{p}_i', \\ \overline{\mu}_i^{II} &\coloneqq \mathbb{E}\left[r_j b_{1,V}(V_j,W_j)\,q_j' Q^{-1} q_i v_{ji}\middle|\mathcal{I}_i\right], \text{ and} \\ b_{1,V}(V_j,W_j) &\coloneqq \partial b_1(v,w)/\partial v\big|_{v=V_j,w=W_j}. \end{aligned} \tag{A.18}$$

Let $F := \Omega_2^{-1/2}$, which is well-defined because $\Omega_2 = \Omega_{21} + \Omega_{22}$ and

$$\begin{aligned}\Omega_{21} &= A_1 P^{-1} \left(\mathbb{E} p_i p_i' u_i^2\right) P^{-1} A_1' \\ &= A_1 P^{-1} \left(\mathbb{E}\left[p_i p_i' \mathbb{E}\left(u_i^2 \middle| X_i, Z_i, W_i\right)\right]\right) P^{-1} A_1' \\ &\geq C A_1 P^{-1} A_1' = C r(x)' \left(\mathbb{E} r_i \overline{p}_i'\right) P^{-1} \left(\mathbb{E} \overline{p}_i r_i'\right) r(x) > 0,\end{aligned} \tag{A.19}$$

where the first inequality holds by Assumption 7(c) and the last inequality holds by Assumption 8(a).

Define functionals:

$$\begin{aligned}\widehat{a}\left(\widehat{b}_1, \widehat{V}\right) &:= \widehat{\mathbb{E}}\left[\widehat{b}_1\left(\widehat{V}, W\right) \middle| X = x\right] = \widehat{b}(x), \\ \widehat{a}\left(b_1, \widehat{V}\right) &:= \widehat{\mathbb{E}}\left[b_1\left(\widehat{V}, W\right) \middle| X = x\right], \\ \widehat{a}(b_1, V) &:= \widehat{\mathbb{E}}\left[b_1(V, W) \middle| X = x\right], \text{ and} \\ a(b_1, V) &= \mathbb{E}\left[b_1(V, W) \middle| X = x\right] = b(x).\end{aligned}$$

I expand

$$\begin{aligned}&\sqrt{n} F\left(\widehat{a}\left(\widehat{b}_1, \widehat{V}\right) - a(b_1, V)\right) \\ &= \sqrt{n} F\left(\widehat{a}\left(\widehat{b}_1, \widehat{V}\right) - \widehat{a}\left(b_1, \widehat{V}\right) + \widehat{a}\left(b_1, \widehat{V}\right) - \widehat{a}(b_1, V) + \widehat{a}(b_1, V) - a(b_1, V)\right) \\ &= n^{-1/2} \sum_i \left(\psi_{1i} + \psi_{2i} + \psi_{3i}\right) + o_p(1),\end{aligned} \tag{A.20}$$

and show that

$$\psi_{1i} = H_1\left(p_i u_i - \overline{\mu}_i^I\right), \ \psi_{2i} = H_2 \overline{\mu}_i^{II}, \text{ and } \psi_{3i} = H_2 r_i \xi_i, \tag{A.21}$$

where $H_1 := F A_1 P^{-1}$, $A_1 = r(x)' R^{-1} \mathbb{E} r_i \overline{p}_i' = r(x)' \mathbb{E} r_i \overline{p}_i'$, $r_i := r(x_i)$, $P := \mathbb{E} p_i p_i'$, $p_i := p^{M_2}(X_i, V_i, W_i)$, $u_i := y_i - G(s_i)$, $\overline{\mu}_i^I := \mathbb{E}\left[G_V(S_j) \tau_V(V_j) p_j q_j' Q^{-1} q_i v_{ji} \middle| \mathcal{I}_i\right]$, $G_V(S_j) := \partial G(s)/\partial v|_{s=S_j}$, $v_{ji} := \mathbb{1}\{x_i \leq x_j\} - F(x_j | z_i, w_i)$, $H_2 := F A_2 R^{-1} = F A_2$, $A_2 := r(x)'$, $\overline{\mu}_i^{II} := \mathbb{E}\left[r_j b_{1,V}(V_j, W_j) q_j' Q^{-1} q_i v_{ji} \middle| \mathcal{I}_i\right]$, $b_{1,V}(V_j, W_j) = \partial b_1(v, w)/\partial v|_{v=V_j}$, and $\xi_i = b_1(v_i, w_i) - b(x_i)$.

Let $\widehat{H}_1 := F \widehat{A}_1 \widehat{P}^{-1}$, $\widehat{A}_1 := r(x)' \widehat{R}^{-1} n^{-1} \sum_{i=1}^n r_i \widehat{\overline{p}}_i'$, $\widehat{R} := n^{-1} \sum_{i=1}^n r_i r_i'$, $\widehat{\overline{p}}_i := \overline{p}(\widehat{s}_i)$, $\widehat{H}_2 := F A_2 \widehat{R}^{-1}$, $G := (G(s_1), \ldots, G(s_n))'$, and $\widetilde{G} := (G(\widehat{s}_1), \ldots, G(\widehat{s}_n))'$. First, for $\psi_{1i}$,

$$\begin{aligned}&\sqrt{n} F\left(\widehat{a}\left(\widehat{b}_1, \widehat{V}\right) - \widehat{a}\left(b_1, \widehat{V}\right)\right) \\ &= n^{-1/2} F r(x)' \widehat{R}^{-1} r'\left(\widehat{B} - \widetilde{B}\right)\end{aligned}$$

$$
\begin{aligned}
&= n^{-1/2} Fr\left(x\right)' \widehat{R}^{-1} r' \left(n^{-1} \widehat{\overline{p}} \widehat{P}^{-1} \widehat{\overline{p}}' Y - \widetilde{B}\right) \\
&= n^{-1/2} Fr\left(x\right)' \widehat{R}^{-1} r' \left[n^{-1} \widehat{\overline{p}} \widehat{P}^{-1} \widehat{\overline{p}}' \left(Y - G + G - \widetilde{G} + \widetilde{G} - \widehat{\overline{p}\alpha}^{M_2}\right) + \left(\widehat{\overline{p}\alpha}^{M_2} - \widetilde{B}\right)\right] \\
&= n^{-1/2} \sum_i \widehat{H}_1 \widehat{p}_i \left[u_i - \left(G\left(\widehat{s}_i\right) - G\left(s_i\right)\right)\right] + n^{-1/2} \widehat{H}_1 \widehat{\overline{p}}' \left(\widetilde{G} - \widehat{\overline{p}\alpha}^{M_2}\right) \\
&\quad + n^{-1/2} \widehat{H}_2 r' \left(\widehat{\overline{p}\alpha}^{M_2} - \widetilde{B}\right) \\
&=: D_{11} + D_{12} + D_{13}. \qquad \text{(A.22)}
\end{aligned}
$$

I show $D_{11} = n^{-1/2} \sum_i \psi_{1i} + o_p(1)$, $D_{12} = o_p(1)$, and $D_{13} = o_p(1)$.

The proof of

$$
D_{11} = n^{-1/2} \sum_i \psi_{1i} + o_p(1) \qquad \text{(A.23)}
$$

is analogous to that of Lemma B7 and B8 of IN02, except that I need to establish $\left\|\widehat{H}_1 - H_1\right\| = o_p(1)$ and $\left\|\widehat{H}_2 - H_2\right\| = o_p(1)$. To prove these two claims, notice that

$$
\|H_1\| = O(1) \text{ and } \|H_2\| = O(1), \qquad \text{(A.24)}
$$

because $\|H_1\|^2 \leq CA_1A_1'/\Omega_2 \leq C$ and $\|H_2\|^2 = A_2A_2'/\Omega_2 \leq CA_1A_1'/\Omega_2 \leq C$. In addition, $\left\|\widehat{P} - P\right\| = o_p(1)$, $\left\|\widehat{R} - I\right\| = o_p(1)$, and $\left\|n^{-1} \sum_i r_i \widehat{\overline{p}}_i' - \mathbb{E} r_i \overline{p}_i'\right\| = o_p(1)$ by Lemma A3 of IN02 and Newey (1997). Furthermore, $\lambda_{\min}(P) \geq C > 0$ and $\|P\| = O(1)$ by Assumption 5. By Slutsky's theorem, $\left\|\widehat{R}^{-1} - I\right\| = o_p(1)$ and $\left\|\widehat{P}^{-1} - P^{-1}\right\| = o_p(1)$. By the CS inequality and Lemma A3 of IN02,

$$
\begin{aligned}
\left\|n^{-1} \sum_i r_i \left(\overline{p}_i - \widehat{\overline{p}}_i\right)'\right\|^2 &\leq n^{-1} \sum_i \|r_i\|^2 \times n^{-1} \sum_i \left\|\widehat{\overline{p}}_i - \overline{p}_i\right\|^2 \\
&= O_p\left(M_3 \zeta_1(m_2)^2 \Delta_{1n}^2\right) = o_p(1). \qquad \text{(A.25)}
\end{aligned}
$$

Therefore, by the triangle inequality with probability approaching one,

$$
\begin{aligned}
&\left\|\widehat{H}_1 - H_1\right\|^2 \\
= &\left\|F\widehat{A}_1\widehat{P}^{-1} - FA_1P^{-1}\right\|^2 \\
\leq &\, 2\left\|F\left(\widehat{A}_1 - A_1\right)\widehat{P}^{-1}\right\|^2 + 2\left\|FA_1\left(\widehat{P}^{-1} - P^{-1}\right)\right\|^2 \\
= &\, 2\left\|F\left(r(x)'(I + o_p(1))\left(\mathbb{E} r_i \overline{p}_i' + o_p(1)\right) - r(x)' \mathbb{E} r_i \overline{p}_i'\right)\widehat{P}^{-1}\right\|^2 \\
&\quad + 2\left\|FA_1P^{-1}\left(P - \widehat{P}\right)\widehat{P}^{-1}\right\|^2 \\
\leq &\, \|H_2\|^2 o_p(1) + \|H_1\|^2 o_p(1) = o_p(1), \qquad \text{(A.26)}
\end{aligned}
$$

and similarly $\left\|\widehat{H}_2 - H_2\right\| = o_p(1)$, which establish (A.23).

Next, recall that by Assumption 7

$$n^{-1}\left(\widetilde{G}-\widehat{p\alpha}^{M_2}\right)'\left(\widetilde{G}-\widehat{p\alpha}^{M_2}\right)=O_p\left(M_2^{-2d_4/j_4}\right) \text{ and}$$
$$n^{-1}\left(\widehat{p\alpha}^{M_2}-\widetilde{B}\right)'\left(\widehat{p\alpha}^{M_2}-\widetilde{B}\right)=O_p\left(M_2^{-2d_4/j_4}\right).$$

Therefore, for $D_{12}$, I have

$$\begin{aligned}\left|n^{-1/2}\widehat{H}_1\widehat{p}'\left(\widetilde{G}-\widehat{p\alpha}^{M_2}\right)\right|^2 &\leq n\left[\widehat{H}_1\widehat{P}\widehat{H}_1'\right]\left[n^{-1}\left(\widetilde{G}-\widehat{p\alpha}^{M_2}\right)'\left(\widetilde{G}-\widehat{p\alpha}^{M_2}\right)\right]\\ &\leq \left\|\widehat{H}_1\right\|^2 O_p\left(nM_2^{-2d_4/j_4}\right)=o_p(1).\end{aligned} \tag{A.27}$$

For $D_{13}$, similarly to (A.27), I have

$$\begin{aligned}\left|n^{-1/2}\widehat{H}_2 r'\left(\widehat{p\alpha}^{M_2}-\widetilde{B}\right)\right|^2 &\leq n\left[\widehat{H}_2\widehat{R}\widehat{H}_2\right]\left[n^{-1}\left(\widehat{p\alpha}^{M_2}-\widetilde{B}\right)'\left(\widehat{p\alpha}^{M_2}-\widetilde{B}\right)\right]\\ &\leq \left\|\widehat{H}_2\right\|^2 O_p\left(nM_2^{-2d_4/j_4}\right)=o_p(1).\end{aligned} \tag{A.28}$$

Combining the results for $D_{11}$, $D_{12,}$ and $D_{13}$, I have

$$\psi_{1i}=H_1\left(p_i u_i-\overline{\mu}_i^I\right). \tag{A.29}$$

To prove $\psi_{2i}=H_2\overline{\mu}_i^{II}$, by Taylor expansion it is true that

$$\begin{aligned}&\sqrt{n}F\left(\widehat{a}\left(b_1,\widehat{V}\right)-\widehat{a}\left(b_1,V\right)\right)\\ &=n^{-1/2}Fr(x)'\widehat{R}^{-1}r'\left(\widetilde{B}-B\right)\\ &=\widehat{H}_2 n^{-1/2}\sum_i r_i\left(b_1\left(\widehat{v}_i,w_i\right)-b_1\left(v_i,w_i\right)\right)\\ &=\widehat{H}_2 n^{-1/2}\sum_i r_i b_{1,V}\left(v_i,w_i\right)\left(\widehat{v}_i-v_i\right)+\widehat{H}_2 n^{-1/2}\sum_i r_i b_{1,VV}\left(\widetilde{v}_i,w_i\right)\left(\widehat{v}_i-v_i\right)^2/2\\ &=: D_{21}+D_{22},\end{aligned} \tag{A.30}$$

where $b_{1,VV}(v_i,w_i):=\partial^2 b_1(v,w)/\partial v^2|_{v=v_i,w=w_i}$ and $\widetilde{v}_i$ lies between $v_i$ and $\widehat{v}_i$. I prove $D_{21}=n^{-1/2}\sum_i H_2\overline{\mu}_i^{II}+o_p(1)$ and $D_{22}=o_p(1)$.

For $D_{21}$,

$$\begin{aligned}D_{21}&=\widehat{H}_2 n^{-1/2}\sum_i r_i b_{1,V}\left(v_i,w_i\right)\left(\widehat{v}_i-v_i\right)\\ &=H_2 n^{-1/2}\sum_i r_i b_{1,V}\left(v_i,w_i\right)\Delta_i^I+\left(\widehat{H}_2-H_2\right)n^{-1/2}\sum_i r_i b_{1,V}\left(v_i,w_i\right)\left(\widehat{v}_i-v_i\right)\end{aligned}$$

$$
\begin{aligned}
&+ H_2 n^{-1/2} \sum_i r_i b_{1,V}(v_i, w_i) \left(\Delta_i^{II} + \Delta_i^{III}\right) \\
=:\ & D_{211} + D_{212} + D_{213}, \qquad \text{(A.31)}
\end{aligned}
$$

where

$$
\begin{aligned}
&\delta_{ij} = F_{X_i|Z_j,W_j}(x_i|\, z_j, w_j) - q_j' \gamma^{M_1}(x_i),\ \Delta_i^I = q_i' \widehat{Q}^{-1} n^{-1} \sum_j q_j v_{ij}, \\
&\Delta_i^{II} = q_i' \widehat{Q}^{-1} n^{-1} \sum_j q_j \delta_{ij},\ \text{and } \Delta_i^{III} = -\delta_{ii}. \qquad \text{(A.32)}
\end{aligned}
$$

Following an argument identical to the proof of Lemma B7 in IN02,

$$
D_{211} = n^{-1/2} \sum_i H_2 \overline{\mu}_i^{II} + o_p(1). \qquad \text{(A.33)}
$$

For $D_{212}$,

$$
\begin{aligned}
|D_{212}|^2 &\leq Cn \left[\left(\widehat{H}_2 - H_2\right) \widehat{R} \left(\widehat{H}_2 - H_2\right)'\right] \left[n^{-1} \sum_i (\widehat{v}_i - v_i)^2\right] \\
&= O_p\left\{n \left(n^{-1} \zeta(M_3)^2 M_3\right) \Delta_{1n}^2\right\} = o_p(1). \qquad \text{(A.34)}
\end{aligned}
$$

For $D_{213}$,

$$
|D_{213}|^2 \leq Cn \left[H_2 \widehat{R} H_2'\right] \left[n^{-1} \sum_i \left(\left(\Delta_i^{II}\right)^2 + \left(\Delta_i^{III}\right)^2\right)\right] = O_p\left(n M_1^{1-2d_1/j_1}\right) = o_p(1), \qquad \text{(A.35)}
$$

where the first equality is established in the proof of Theorem 4 in IN02.

Next, for $D_{22}$,

$$
\begin{aligned}
|D_{22}| &\leq C\sqrt{n} \left\|\widehat{H}_2\right\| \sup_{x \in \mathcal{X}} \|r(x)\| \left|n^{-1} \sum_i (\widehat{v}_i - v_i)^2\right| \\
&= O_p\left(\sqrt{n} \zeta(M_3) \Delta_{1n}^2\right) = o_p(1). \qquad \text{(A.36)}
\end{aligned}
$$

Combining the results for $D_{21}$ and $D_{22}$,

$$
\sqrt{n} F\left(\widehat{a}\left(b_1, \widehat{V}\right) - \widehat{a}(b_1, V)\right) = n^{-1/2} \sum_i H_2 \overline{\mu}_i^{II} + o_p(1). \qquad \text{(A.37)}
$$

To show $\psi_{3i} = H_2 r_i \xi_i$, I expand

$$
\sqrt{n} F\left(\widehat{a}(b_1, V) - a(b_1, V)\right)
$$

$$
\begin{aligned}
&= n^{-1/2}\sum_i \widehat{H}_2 r_i b_1(v_i,w_i) - \sqrt{n}Fb(x) \\
&= n^{-1/2}\sum_i H_2 r_i\left(b_1(v_i,w_i) - b(x_i)\right) + n^{-1/2}\sum_i \left(\widehat{H}_2 - H_2\right) r_i\left(b_1(v_i,w_i) - b(x_i)\right) \\
&\quad + n^{-1/2}\sum_i \widehat{H}_2 r_i\left(b(x_i) - r_i'\rho^{M_3}\right) - \sqrt{n}F\left(b(x) - r(x)'\rho^{M_3}\right) \\
&=: D_{31} + D_{32} + D_{33} + D_{34}, \qquad \text{(A.38)}
\end{aligned}
$$

where I use the definition of $\widehat{H}_2$ and $\widehat{R}$ for the decomposition. Recall that $D_{31} = n^{-1/2}\sum_i H_2 r_i \xi_i$ by the definition of $\xi_i$. Thus, I only need to show $D_{32}$, $D_{33}$, and $D_{34}$ are all $o_p(1)$.

For $D_{32}$,

$$
\begin{aligned}
\mathbb{E}\left[\left|D_{32}\right|^2\middle|\mathbf{X}\right] &= n^{-1}\left(\widehat{H}_2 - H_2\right) r'\mathbb{E}\left[\xi\xi'\middle|\mathbf{X}\right] r\left(\widehat{H}_2 - H_2\right)' \\
&\leq C\left(\widehat{H}_2 - H_2\right)\widehat{R}\left(\widehat{H}_2 - H_2\right)' \\
&\leq C\left\|\widehat{H}_2 - H_2\right\|^2\left(1 + \left\|\widehat{R} - I\right\|\right) \\
&= O_p\left\{\left\|\widehat{H}_2 - H_2\right\|^2\right\} \\
&= O_p\left(n^{-1}\zeta(M_3)^2 M_3\right) = o_p(1), \qquad \text{(A.39)}
\end{aligned}
$$

where the first inequality holds by Assumption 6(c) and the fact that $\widehat{H}_2$ and $r$ are functions of $X$ only, the second equality holds by $\left\|\widehat{R} - I\right\| = o_p(1)$, and the third equality follows similarly as in equation (A.1) and (A.6) of Newey (1997). Therefore, $D_{32} = o_p(1)$ by the CM inequality.

For $D_{33}$, by the CS inequality,

$$
\begin{aligned}
\left|D_{33}\right|^2 &\leq n\left(\widehat{H}_2\widehat{R}\widehat{H}_2'\right) n^{-1}\sum_i\left(b(x_i) - r_i'\rho^{M_3}\right)^2 \\
&= O_p\left(nM_3^{-2d_3/j_3}\right) = o_p(1), \qquad \text{(A.40)}
\end{aligned}
$$

where the first equality holds by Assumption 6(a).

Finally, for $D_{34}$,

$$
\left|D_{34}\right|^2 = nF^2\left(b(x) - r(x)'\rho^{M_3}\right)^2 = O_p\left(nM_3^{-2d_3/j_3}\right) = o_p(1). \qquad \text{(A.41)}
$$

Combining the results for $D_{31}$, $D_{32}$, $D_{33}$, and $D_{34}$, I have

$$
\sqrt{n}F\left(\widehat{a}(b_1,V) - a(b_1,V)\right) = n^{-1/2}\sum_i H_2 r_i \xi_i + o_p(1). \qquad \text{(A.42)}
$$

In sum, I have proved

$$\sqrt{n}F\left(\widehat{a}\left(\widehat{b}_1,\widehat{V}\right)-a\left(b_1,V\right)\right)=n^{-1/2}\sum_i\left(\psi_{1i}+\psi_{2i}+\psi_{3i}\right)+o_p\left(1\right), \tag{A.43}$$

where

$$\psi_{1i}=H_1\left(p_iu_i-\overline{\mu}_i^I\right),\ \psi_{2i}=H_2\overline{\mu}_i^{II},\ \text{and }\psi_{3i}=H_2r_i\xi_i. \tag{A.44}$$

Furthermore, observe that

$$H_1p_iu_i\perp\left(H_1\overline{\mu}_i^I,H_2\overline{\mu}_i^{II},H_2r_i\xi_i\right) \tag{A.45}$$

because $\mathbb{E}\left[u_i|\,X_i,V_i,W_i\right]=0$.

Let $\Psi_{in}=n^{-1/2}\left(\psi_{1i}+\psi_{2i}+\psi_{3i}\right)$. It is clear that $\mathbb{E}\Psi_{in}=0$ and $\mathbb{V}\left(\Psi_{in}\right)=n^{-1}$. For any $\epsilon>0$, under Assumptions 7 and 8,

$$\begin{aligned}&n\mathbb{E}\left[\mathbb{1}\left\{|\Psi_{in}|>\epsilon\right\}\Psi_{in}^2\right]\\ \le\;& n\epsilon^2\mathbb{E}\left[\mathbb{1}\left\{|\Psi_{in}|>\epsilon\right\}\left(\Psi_{in}/\epsilon\right)^4\right]\le n\epsilon^{-2}\mathbb{E}\Psi_{in}^4\\ \le\;& Cn^{-1}\mathbb{E}\left[\left(H_1p_iu_i\right)^4+\left(H_1\overline{\mu}_i^I\right)^4+\left(H_2\overline{\mu}_i^{II}\right)^4+\left(H_2r_i\xi_i\right)^4\right]\\ \le\;& Cn^{-1}\left(\zeta\left(m_2\right)^2M_2+\zeta\left(m_2\right)^4\zeta\left(M_1\right)^4M_1+\zeta\left(M_3\right)^4\zeta\left(M_1\right)^4M_1+\zeta\left(M_3\right)^2M_3\right)\to0,\end{aligned} \tag{A.46}$$

where the last inequality holds by Lemma B5 of IN02.

Then, by the Lindeberg–Feller CLT,

$$\sqrt{n}\Omega_2^{-1/2}\left(\widehat{b}\left(x\right)-b\left(x\right)\right)\xrightarrow{d}N\left(0,1\right). \tag{A.47}$$

To construct a feasible confidence interval, one needs a consistent estimator of the covariance matrix $\Omega_2$. Define

$$\begin{aligned}\widehat{\Omega}_2&:=\widehat{\Omega}_{21}+\widehat{\Omega}_{22},\ \text{where}\\ \widehat{\Omega}_{21}&:=\widehat{A}_1\widehat{P}^{-1}\left(n^{-1}\sum_i\widehat{p}_i\widehat{p}_i'\widehat{u}_i^2\right)\widehat{P}^{-1}\widehat{A}_1',\\ \widehat{\Omega}_{22}&:=n^{-1}\sum_i\left(\widehat{A}_1\widehat{P}^{-1}\widehat{\overline{\mu}}_i^I-r\left(x\right)'\widehat{R}^{-1}\left(\widehat{\overline{\mu}}_i^{II}+r_i\widehat{\xi}_i\right)\right)\left(\widehat{A}_1\widehat{P}^{-1}\widehat{\overline{\mu}}_i^I-r\left(x\right)'\widehat{R}^{-1}\left(\widehat{\overline{\mu}}_i^{II}+r_i\widehat{\xi}_i\right)\right)',\\ \widehat{A}_1&:=r\left(x\right)'\widehat{R}^{-1}\left(n^{-1}\sum_i r_i\overline{p}\left(\widehat{s}_i\right)'\right),\\ \widehat{\overline{\mu}}_i^{II}&:=n^{-1}\sum_j r_j\widehat{b}_{1,V}\left(\widehat{v}_j,w_j\right)q_j'\widehat{Q}^{-1}q_i\widehat{v}_{ji},\ \text{and}\end{aligned}$$

$$\widehat{\xi}_i := \widehat{b}_1 (\widehat{v}_i, w_i) - \widehat{b}(x_i). \tag{A.48}$$

I show $\widehat{\Omega}_2/\Omega_2 - 1 \xrightarrow{p} 0$.

Recall that

$$\Omega_2 = \mathbb{E}\left(A_1 P^{-1} p_i u_i\right)^2 + \mathbb{E}\left(A_1 P^{-1}\overline{\mu}_i^I - A_2\left(\overline{\mu}_i^{II} + r_i\xi_i\right)\right)^2 = \Omega_{21} + \Omega_{22} \tag{A.49}$$

and

$$\widehat{\Omega}_2 = n^{-1}\sum_i \left(\widehat{A}_1 \widehat{P}^{-1}\widehat{p}_i\widehat{u}_i\right)^2 + n^{-1}\sum_i\left(\widehat{A}_1\widehat{P}^{-1}\widehat{\overline{\mu}}_i^I - \widehat{A}_2\widehat{R}^{-1}\left(\widehat{\overline{\mu}}_i^{II} + r_i\widehat{\xi}_i\right)\right)^2 =: \widehat{\Omega}_{21} + \widehat{\Omega}_{22}. \tag{A.50}$$

The proof of $\widehat{\Omega}_{21}/\Omega_2 - \Omega_{21}/\Omega_2 \xrightarrow{p} 0$ is almost identical to the proof of Lemma B10 of IN02, except that $\widehat{A}_1$ instead of $A_1$ appears in the definition of $\widehat{H}_1$. Nonetheless, I have shown $\left\|\widehat{H}_1 - H_1\right\| = o_p(1)$. Thus, the proof of Lemma B10 of IN02 directly applies.

For $\widehat{\Omega}_{22}$, I need to show

$$\begin{aligned} n^{-1}\sum_i\left(\widehat{H}_1\widehat{\overline{\mu}}_i^I - H_1\overline{\mu}_i^I\right)^2 &= o_p(1), \\ n^{-1}\sum_i\left(\widehat{H}_2\widehat{\overline{\mu}}_i^{II} - H_2\overline{\mu}_i^{II}\right)^2 &= o_p(1), \text{ and} \\ n^{-1}\sum_i\left(\widehat{H}_2 r_i\widehat{\xi}_i - H_2 r_i\xi_i\right)^2 &= o_p(1). \end{aligned} \tag{A.51}$$

The first two convergence results have been established in Lemma B9 of IN02. For the last one,

$$\begin{aligned} &\widehat{H}_2 r_i\widehat{\xi}_i - H_2 r_i \xi_i \\ &= \widehat{H}_2 r_i\left(\widehat{\xi}_i - \xi_i\right) + \left(\widehat{H}_2 - H_2\right) r_i\xi_i \\ &= \widehat{H}_2 r_i\left(\widehat{b}_1(\widehat{v}_i, w_i) - \widehat{b}(x_i) - b_1(v_i, w_i) + b(x_i)\right) + \left(\widehat{H}_2 - H_2\right) r_i\xi_i \\ &= \widehat{H}_2 r_i\left(\widehat{b}_1(\widehat{v}_i, w_i) - b_1(\widehat{v}_i, w_i)\right) + \widehat{H}_2 r_i\left(b_1(\widehat{v}_i, w_i) - b_1(v_i, w_i)\right) \\ &\quad + \widehat{H}_2 r_i\left(b(x_i) - \widehat{b}(x_i)\right) + \left(\widehat{H}_2 - H_2\right) r_i\xi_i \\ &=: D_{41i} + D_{42i} + D_{43i} + D_{44i}. \end{aligned} \tag{A.52}$$

For $D_{41}$,

$$n^{-1}\sum_i D_{41i}^2 \leq \left\|\widehat{H}_2\right\|^2 \sup_{x\in\mathcal{X}}\left\|r(x)\right\|^2 n^{-1}\sum_i\left(\widehat{b}_1(\widehat{v}_i, w_i) - b_1(\widehat{v}_i, w_i)\right)^2$$

$$\leq C\zeta(M_3)^2 n^{-1}\sum_i\left[\left(\widehat{\widetilde{p}}_i'\left(\widehat{\alpha}^{M_2}-\alpha^{M_2}\right)\right)^2+\left(\widehat{\widetilde{p}}_i'\alpha^{M_2}-b_1(\widehat{v}_i,w_i)\right)^2\right]$$
$$= O_p\left(\zeta(M_3)^2\Delta_{2n}^2\right)=o_p(1), \tag{A.53}$$

where the second inequality holds by $\left\|\widehat{H}_2\right\| = O_p(1)$ and Assumption 8(a) and the first equality holds by (A.8).

For $D_{42}$,

$$n^{-1}\sum_i D_{42i}^2 \leq \left\|\widehat{H}_2\right\|^2 \sup_{x\in\mathcal{X}}\|r(x)\|^2 n^{-1}\sum_i\left(b_1(\widehat{v}_i,w_i)-b_1(v_i,w_i)\right)^2$$
$$\leq C\zeta(M_3)^2 n^{-1}\sum_i(\widehat{v}_i-v_i)^2 = O_p\left(\zeta(M_3)^2\Delta_{1n}^2\right)=o_p(1), \tag{A.54}$$

where the first equality holds by Lemma 1.

The proof of $n^{-1}\sum_i D_{43i}^2 = o_p(1)$ is completely analogous to (A.53) and is omitted.

For $D_{44}$,

$$\mathbb{E}\left[n^{-1}\sum_i D_{44i}^2\middle|\mathbf{X}\right] = \left(\widehat{H}_2-H_2\right)n^{-1}\sum_i r_i r_i'\mathbb{E}\left(\xi_i^2\middle| X_i\right)\left(\widehat{H}_2-H_2\right)'$$
$$\leq C\left(\widehat{H}_2-H_2\right)\widehat{R}\left(\widehat{H}_2-H_2\right)'$$
$$\leq C\left\|\widehat{H}_2-H_2\right\|^2 = o_p(1), \tag{A.55}$$

where the first equality holds by the fact that both $\widehat{H}_2$ and $r_i$ are functions of $\mathbf{X}$, the first inequality holds by Assumption 6(c), and the last inequality uses $\left\|\widehat{R}-I\right\| = o_p(1)$. Then, by the CM inequality,

$$n^{-1}\sum_i D_{44i}^2 = o_p(1). \tag{A.56}$$

Combining the results for $D_{41}$, $D_{42}$, $D_{43}$, and $D_{44}$, I have

$$n^{-1}\sum_i\left(\widehat{H}_2 r_i\widehat{\xi}_i - H_2 r_i\xi_i\right)^2 = o_p(1). \tag{A.57}$$

Therefore, by (A.51),

$$n^{-1}\sum_i\left(\left(\widehat{H}_1\widehat{\overline{\mu}}_i^I-\widehat{H}_2\widehat{\overline{\mu}}_i^{II}-\widehat{H}_2 r_i\widehat{\xi}_i\right)-\left(H_1\overline{\mu}_i^I-H_2\overline{\mu}_i^{II}-H_2 r_i\xi_i\right)\right)^2$$
$$\leq 3n^{-1}\sum_i\left(\widehat{H}_1\widehat{\overline{\mu}}_i^I-H_1\overline{\mu}_i^I\right)^2+3n^{-1}\sum_i\left(\widehat{H}_2\widehat{\overline{\mu}}_i^{II}-H_2\overline{\mu}_i^{II}\right)^2$$
$$+3n^{-1}\sum_i\left(\widehat{H}_2 r_i\widehat{\xi}_i-H_2 r_i\xi_i\right)^2 = o_p(1). \tag{A.58}$$

Since $\mathbb{E}\left(H_1\overline{\mu}_i^I - H_2\overline{\mu}_i^{II} - H_2 r_i \xi_i\right)^2 = \Omega_{22}/\Omega_2 \leq 1$, by Markov's inequality and Lemma B6 of IN02,

$$\left|\widehat{\Omega}_{22}/\Omega_2 - n^{-1}\sum_i \left(H_1\overline{\mu}_i^I - H_2\overline{\mu}_i^{II} - H_2 r_i \xi_i\right)^2\right| = o_p(1). \tag{A.59}$$

By the law of large numbers,

$$\left|n^{-1}\sum_i \left(H_1\overline{\mu}_i^I - H_2\overline{\mu}_i^{II} - H_2 r_i \xi_i\right)^2 - \Omega_{22}/\Omega_2\right| = o_p(1). \tag{A.60}$$

Therefore, by the triangle inequality,

$$\widehat{\Omega}_{22}/\Omega_2 - \Omega_{22}/\Omega_2 = o_p(1). \tag{A.61}$$

Combining the results for $\widehat{\Omega}_{21}$ and $\widehat{\Omega}_{22}$, I obtain

$$\widehat{\Omega}_2/\Omega_2 - 1 \xrightarrow{p} 0. \tag{A.62}$$

Next, for the APE estimator $\widehat{\overline{b}}$, define $b_t^{\text{APE}} = \mathbb{E}\beta_{it}$ and thus $\overline{b} = T^{-1}\sum_{t=1}^T b_t^{\text{APE}}$. I first estimate $b_t^{\text{APE}}$ by setting $r^{M_3}(x) = 1$ in (4.7) for $\widehat{b}_t(x)$. Hence, the asymptotic normality proof for $\widehat{b}_t(x)$ directly applies to $\widehat{b}_t^{\text{APE}}$. In particular, by adapting to (A.43) with $r^{M_3}(x) = 1$, the influence function for $\widehat{b}_t^{\text{APE}}$ in period $t$ is

$$\psi_{it} = A_{3t}P_t^{-1}(p_{it}u_{it} + \overline{\mu}_{it}^I) - (\overline{\mu}_{it}^{III} + \xi_{it}), \tag{A.63}$$

where $A_{3t} = \mathbb{E}\overline{p}_{it}'$ and $\overline{\mu}_{it}^{III} = \mathbb{E}\left[b_{1t,V}\left(V_{jt}, W_j\right) q_{jt}' Q_t^{-1} q_{it} v_{ji,t} \middle| \mathcal{I}_{i,t}\right]$.

Let

$$\Omega_3 := \mathbb{V}\left(\frac{1}{T}\sum_{t=1}^T \psi_{it}\right) = \frac{1}{T^2}\sum_{s=1}^T\sum_{t=1}^T \text{Cov}(\psi_{it}, \psi_{is}). \tag{A.64}$$

Following the same argument as for $\widehat{b}(x)$, I have

$$\sqrt{n}\,\Omega_3^{-1/2}(\widehat{\overline{b}} - \overline{b}) = \sqrt{n}\,\Omega_3^{-1/2}\left[\frac{1}{T}\sum_{t=1}^T(\widehat{b}_t^{\text{APE}} - b_t^{\text{APE}})\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\Omega_3^{-1/2}\frac{1}{T}\sum_{t=1}^T\psi_{it}\right) + o_p(1) \tag{A.65}$$

and the asymptotic normality follows by the Lindeberg-Feller CLT:

$$\sqrt{n}\,\Omega_3^{-1/2}(\widehat{\overline{b}} - \overline{b}) \xrightarrow{d} N(0, I). \tag{A.66}$$

Finally, by (A.64), an estimator $\widehat{\Omega}_3$ for $\Omega_3$ is written as

$$\widehat{\Omega}_3 = \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} \left( \frac{1}{n} \sum_{i=1}^{n} \widehat{\psi}_{it} \widehat{\psi}_{is}' \right), \tag{A.67}$$

where $\widehat{\psi}_{it} = \widehat{A}_{3t} \widehat{P}_t^{-1} (\widehat{p}_{it} \widehat{u}_{it} + \widehat{\widehat{\mu}}_{it}^{I}) - (\widehat{\widehat{\mu}}_{it}^{III} + \widehat{\xi}_{it})$, $\widehat{A}_{3t} = n^{-1} \sum_i \widehat{\widehat{p}}_{it}'$, and $\widehat{\widehat{\mu}}_{it}^{III} = n^{-1} \sum_j \widehat{b}_{1t,V} (\widehat{v}_{jt}, w_j)\, q_{jt}' \widehat{Q}^{-1} q_{it} \widehat{v}_{ji,t}$. The consistency is established following the same argument as in deriving (A.62):

$$\widehat{\Omega}_3 / \Omega_3 - 1 \xrightarrow{p} 0. \tag{A.68}$$

□

*Remark* 5 (**Degree of Basis Functions**). The three degrees $M_1, M_2, M_3$ correspond to the series dimensions used to estimate $V_t(x, z, w)$, $G_t(x, v, w)$, and $b_t(x)$, respectively. For consistency I require $M_i \to \infty$ while $M_i/n \to 0$, so that approximation bias and sampling variance both vanish. Under Assumption 5, the first-step estimator satisfies

$$\text{MSE}(\widehat{V}_t) = O\left( \underbrace{M_1^{1-2d_1/j_1}}_{\text{bias}} + \underbrace{M_1/n}_{\text{variance}} \right),$$

where the extra factor $M_1$ in the bias term comes from the uniform-in-$x$ analysis with $\mathbf{1}\{X_{it} \leq x\}$ as the dependent variable (as in IN02). For steps 2–3 the analysis is analogous, with two additional points: (i) since the model implies $G_t$ is linear in $x$, I use the reduced basis $p^{M_2}(x, v, w) = x \otimes p^{m_2}(v, w)$, which reduces dimension ($M_2 = m_2 d_X$); and (ii) first-step errors propagate, so the error for $\widehat{b}_t(x)$ aggregates across steps

$$\text{MSE}(\widehat{b}_t(x)) = \left( M_1^{1-2d_1/j_1} + M_1/n \right) + \left( m_2^{-2d_2/j_2} + m_2/n \right) + \left( M_3^{-2d_3/j_3} + M_3/n \right).$$

Sup-norm results additionally impose standard bounds on basis functions and derivatives (Assumption 6(b)), satisfied by common choices such as splines and power series. Finally, Assumption 7 imposes "not too fast" growth on $M_i$ for asymptotic normality; for instance, for $\widehat{\widehat{b}}$ (which does not involve $M_3$), with splines it suffices that $M_1 = o(n^{1/7})$ and $M_2 = o(n^{1/7})$, which is feasible when $d_1/j_1 \geq 4$ and $d_2/j_2 \geq 4$.

## A.2 Nonparametric Justification of the Index Exclusion Assumption

Next, I introduce a proposition that supplies sufficient conditions for Assumption 2 by adapting the exchangeability condition of Altonji and Matzkin (2005).

**Proposition 1** (**Sufficient Conditions for Assumption 2**). *Suppose the following conditions hold:*

(a) $f_{\eta_i|A_i} = f_{\widetilde{\eta}_i\big|A_i}$ *for any permutation* $\widetilde{\eta}_i$ *of the* $\eta_i$ *vector in time,*

(b) $Z_{it} \perp (A_i, \eta_{it})$ *and the support of* $(X_{it}, Z_{it})$ *is compact, and,*

(c) $f_{A_i|\mathbf{X}_i,\mathbf{Z}_i}$ *is continuous in* $(\mathbf{X}_i, \mathbf{Z}_i)$.

*Then, Assumption 2 is satisfied.*

***Proof of Proposition 1.*** To begin with, I use the first condition $f_{\eta_i|A_i} = f_{\widetilde{\eta}_i\big|A_i}$ to establish the exchangeability condition

$$f_{A_i|\mathbf{X}_i,\mathbf{Z}_i} = f_{A_i|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{Z}}_i}, \tag{A.69}$$

where $\left(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{Z}}_i\right)$ is any permutation of $(\mathbf{X}_i, \mathbf{Z}_i)$ in time. It is worth emphasizing that (A.69), which will be proved in the following, is different from Assumption 2.3 of Altonji and Matzkin (2005). In particular, Altonji and Matzkin (2005) assume an exchangeability condition involving both $A$ and $\eta_t$, i.e., $f_{A_i,\eta_{it}|\mathbf{X}_i} = f_{A_i,\eta_{it}|\widetilde{\mathbf{X}}_i}$, which effectively rules out time-varying endogeneity through the random coefficients because it requires the density of $\eta_{it}$ given $X_{it} = x_{it}$ to be the same as that given $X_{it} = x_{is}$ for any $s \neq t$. Instead, I prove (A.69) under a more primitive assumption of $f_{\eta_i|A_i} = f_{\widetilde{\eta}_i\big|A_i}$ that is compatible with time-varying endogeneity through the random coefficients.

Without loss of generality, I prove (A.69) for $T = 2$ since any ordering of $\{1, \ldots, T\}$ for finite $T$ can be achieved via a finite number of pairwise permutations. Then, I prove that one can construct $W_i$ such that Assumption 2 holds. For simplicity of notation, I assume $X_{it}$ and $Z_{it}$ are both scalars and suppress $i$ subscripts in all variables. Thus, in this proof all subscripts of the variables denote the time period.

By condition (a) of Proposition 1,

$$f_{A,\eta_1,\eta_2}(a, h_1, h_2) = f_{A,\eta_1,\eta_2}(a, h_2, h_1). \tag{A.70}$$

Let $g^{-1}(X, Z, A)$ denote the inverse function of $g(Z, A, \eta)$ with respect to $\eta$, which exists by Assumption 1. Define $h_1 = g^{-1}(x_1, z_1, a)$ and $h_2 = g^{-1}(x_2, z_2, a)$. Calculate the determinants of the Jacobians as

$$D_1 := \left| \begin{matrix} \frac{\partial A}{\partial X_1} & \frac{\partial A}{\partial X_2} & \frac{\partial A}{\partial A} \\ \frac{\partial g^{-1}(X_1, Z_1, A)}{\partial X_1} & \frac{\partial g^{-1}(X_1, Z_1, A)}{\partial X_2} & \frac{\partial g^{-1}(X_1, Z_1, A)}{\partial A} \\ \frac{\partial g^{-1}(X_2, Z_2, A)}{\partial X_1} & \frac{\partial g^{-1}(X_2, Z_2, A)}{\partial X_2} & \frac{\partial g^{-1}(X_2, Z_2, A)}{\partial A} \end{matrix} \right|_{\substack{(X_1, X_2, Z_1, Z_2, A) \\ = (x_1, x_2, z_1, z_2, a)}}$$

$$= \left| \begin{matrix} 0 & 0 & 1 \\ \frac{\partial g^{-1}(X_1, Z_1, A)}{\partial X_1} & 0 & \frac{\partial g^{-1}(X_1, Z_1, A)}{\partial A} \\ 0 & \frac{\partial g^{-1}(X_2, Z_2, A)}{\partial X_2} & \frac{\partial g^{-1}(X_2, Z_2, A)}{\partial A} \end{matrix} \right|_{\substack{(X_1, X_2, Z_1, Z_2, A) \\ = (x_1, x_2, z_1, z_2, a)}}$$

$$= \partial g^{-1}(X, Z, A) / \partial X\Big|_{(X,Z,A)=(x_1,z_1,a)} \times \partial g^{-1}(X, Z, A) / \partial X\Big|_{(X,Z,A)=(x_2,z_2,a)},$$

and similarly

$$D_2 := \left| \begin{matrix} \frac{\partial g(Z_1, A, \eta_1)}{\partial A} & \frac{\partial g(Z_1, A, \eta_1)}{\partial \eta_1} & \frac{\partial g(Z_1, A, \eta_1)}{\partial \eta_2} \\ \frac{\partial g(Z_2, A, \eta_2)}{\partial A} & \frac{\partial g(Z_2, A, \eta_2)}{\partial \eta_1} & \frac{\partial g(Z_2, A, \eta_2)}{\partial \eta_2} \\ \frac{\partial A}{\partial A} & \frac{\partial A}{\partial \eta_1} & \frac{\partial A}{\partial \eta_2} \end{matrix} \right|_{\substack{(Z_1, Z_2, A, \eta_1, \eta_2) \\ = (z_2, z_1, a, h_2, h_1)}}$$

$$= \partial g(Z, A, \eta) / \partial \eta|_{(Z,A,\eta)=(z_2,a,h_2)} \times \partial g(Z, A, \eta) / \partial \eta|_{(Z,A,\eta)=(z_1,a,h_1)}.$$

Then,

$$\begin{aligned} & f_{X_1,X_2,A|Z_1,Z_2}(x_1, x_2, a|\, z_1, z_2) \\ &= f_{A,\eta_1,\eta_2|Z_1,Z_2}\left(a, g^{-1}(x_1, z_1, a), g^{-1}(x_2, z_2, a)\middle|\, z_1, z_2\right) |D_1| \\ &= f_{A,\eta_1,\eta_2|Z_1,Z_2}\left(a, g^{-1}(x_2, z_2, a), g^{-1}(x_1, z_1, a)\middle|\, z_2, z_1\right) |D_1| \\ &= f_{X_1,X_2,A|Z_1,Z_2}(x_2, x_1, a|\, z_2, z_1) |D_2 D_1| \\ &= f_{X_1,X_2,A|Z_1,Z_2}(x_2, x_1, a|\, z_2, z_1), \end{aligned} \tag{A.71}$$

where the first equality holds by the change of variables for $\eta_1$ and $\eta_2$, the second equality uses (A.70) and $Z \perp (A, \eta)$, the third equality holds by $X_1 = g(z_2, a, g^{-1}(x_2, z_2, a)) = x_2$ and $X_2 = g(z_1, a, g^{-1}(x_1, z_1, a)) = x_1$, and the last equality uses the fact that the product of derivatives of inverse functions equals one.

Given (A.71), I integrate $A$ out in (A.71) to obtain

$$f_{X_1,X_2|Z_1,Z_2}\left(x_1,x_2|\, z_1,z_2\right)=f_{X_1,X_2|Z_1,Z_2}\left(x_2,x_1|\, z_2,z_1\right), \tag{A.72}$$

which implies

$$\begin{aligned}
& f_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1,x_2,z_1,z_2\right)\\
&= f_{X_1,X_2,A|Z_1,Z_2}\left(x_1,x_2,a|\, z_1,z_2\right)/f_{X_1,X_2|Z_1,Z_2}\left(x_1,x_2|\, z_1,z_2\right)\\
&= f_{X_1,X_2,A|Z_1,Z_2}\left(x_2,x_1,a|\, z_2,z_1\right)/f_{X_1,X_2|Z_1,Z_2}\left(x_2,x_1|\, z_2,z_1\right)\\
&= f_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_2,x_1,z_2,z_1\right),
\end{aligned} \tag{A.73}$$

where the second equality uses (A.71) and (A.72).

The rest of the proof follows the same argument as in Section 2.2 of Altonji and Matzkin (2005). Specifically, I show that for any on-support $a$, the conditional density $f_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1,x_2,z_1,z_2\right)$ can be approximated arbitrarily closely by a function of the form $f_{A|W}\left(a|\, w\right)$. To construct $W$, I follow Weyl (1939) and let $\varphi_1\left(u\right)=\sum_{t=1}^{T}u_t$, $\varphi_2\left(u,v\right)=\sum_{t\neq s}^{T}u_tv_s$, ..., $\varphi_T\left(u,v,\ldots,w\right)=\sum_{t\neq s\neq\ldots\neq k}^{T}u_tv_s\ldots w_k$, where $u,v,\ldots,w$ are generic $T\times 1$ vectors. Then, I substitute each column vector of $\mathbf{D}_i=\left(\mathbf{X}_i,\mathbf{Z}_i\right)$ for each of the arguments $u,v,\ldots,w$ (repetitions included) to construct $W_i$. See Chapter II.3 of Weyl (1939) for details of the construction and its proof.

By condition (b) of Proposition 1, the supports of $X$ and $Z$ are compact. By condition (c) of Proposition 1, $f_{A|X_1,X_2,Z_1,Z_2}$ is continuous in $(X_1,X_2,Z_1,Z_2)$. Therefore, by the Stone-Weierstrass theorem, there exists a function $f^w_{A|X_1,X_2,Z_1,Z_2}$ that is a polynomial in $(X_1,X_2,Z_1,Z_2)$ over a compact set with the property that for any fixed $\delta$ that is arbitrarily close to 0,

$$\max_{x_1,x_2\in\mathcal{X},z_1,z_2\in\mathcal{Z}}\left|f_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1,x_2,z_1,z_2\right)-f^w_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1,x_2,z_1,z_2\right)\right|\leq\delta. \tag{A.74}$$

Let

$$\begin{aligned}
&\overline{f}_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1,x_2,z_1,z_2\right)\\
&:=\left[f_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1,x_2,z_1,z_2\right)+f_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_2,x_1,z_2,z_1\right)\right]/2!
\end{aligned}$$

denote the simple average of $f_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1,x_2,z_1,z_2\right)$ over all $T!$ (here $T=2$) unique permutations of $(x_t,z_t)$, and similarly for $\overline{f}^w_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1,x_2,z_1,z_2\right)$. By (A.73),

$$\overline{f}_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1,x_2,z_1,z_2\right)=f_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1,x_2,z_1,z_2\right).$$

By (A.73), (A.74), and the triangle inequality,

$$\begin{aligned}
&\left| f_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1, x_2, z_1, z_2\right) - \overline{f}^{w}_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1, x_2, z_1, z_2\right)\right| \\
=\ &\left| \overline{f}_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1, x_2, z_1, z_2\right) - \overline{f}^{w}_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1, x_2, z_1, z_2\right)\right| \\
\leq\ &T! \times (\delta/T!) = \delta. \qquad \text{(A.75)}
\end{aligned}$$

Since $f^w$ can be chosen to make $\delta$ arbitrarily small, equation (A.75) implies that $f_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1, x_2, z_1, z_2\right)$ can be approximated arbitrarily closely by a polynomial $\overline{f}^{w}$ that is symmetric in $(x_t, z_t)$ pairs for $t = 1, 2$. Furthermore, by the fundamental theorem of symmetric functions, for any $a$ on its support, $\overline{f}^{w}$ can be written as a polynomial function of the elementary symmetric functions $W$ of $((x_1, z_1), (x_2, z_2))$. I denote this function by $f_{A|W}\left(a|\, w\right)$ and obtain that $f_{A|X_1,X_2,Z_1,Z_2}\left(a|\, x_1, x_2, z_1, z_2\right)$ can be approximated arbitrarily closely by $f_{A|W}\left(a|\, W\right)$. Let $\delta \to 0$ in (A.75). Note that to let $\delta \to 0$ the degree of polynomials in the approximating function $f^w$ needs to increase. However, it does not affect the dimension of $W$ since by the fundamental theorem of symmetric functions, the elements of $W$ have a fixed degree of $T$. The key is that there are two different degrees of polynomials: one is the degree of the approximating functions $f^w$, and the other is the degree of the arguments of $W$.[4]

Then, for any $t \in \{1, ..., T\}$ and on-support $(x_t, z_t, a, w)$

$$f_{A|X_t,Z_t,W}\left(a|\, x_t, z_t, w\right) = f_{A|W}\left(a|\, w\right). \qquad \square$$

## A.3 Further Discussion of Identification Assumptions

When the exchangeability conditions are used to justify Assumption 2, I present a symmetry-based argument in the discussion of Assumption 4 that justifies Assumptions 2 and 4 simultaneously. I provide an example here to clarify this point. Suppose $T = 2$ and $(X_{it}, Z_{it}) \in \mathbb{R}^2$. By Proposition 1, when $W_i$ includes averages through time of the polynomials of $(X_{it}, Z_{it})$ up to the second degree, Assumption 2 holds. Suppose $(X_{i1}, X_{i2}, Z_{i1}, Z_{i2}) = (x_1, x_2, z_1, z_2)$ satisfies $\{V_{it} = v, W_i = w\}$ for $t = 1, 2$ where $x_1, x_2, z_1, z_2 \in \mathbb{R}^2$ and $x_1$ and $x_2$ are linearly independent. Then, by definition of $V_{it}$ and the symmetry of $W_i$ in $(X_{it}, Z_{it})$ through $t$, $(X_{i1}, X_{i2}, Z_{i1}, Z_{i2}) = (x_2, x_1, z_2, z_1)$ must also satisfy $V_{it} = v$ and $W_i = w$ for $t = 1, 2$. Therefore, given $V_{it} = v$ and $W_i = w$ there are two linearly independent vectors of $x_1$ and $x_2$ that $X_{it}$ can take, satisfying Assumption 4.

[4] For example, one may let the degree $P_1$ and $P_2$ of a polynomial function $h(x, y) = \sum_{i=1}^{P_1} (x + y)^i + \sum_{i=1}^{P_2} (xy)^i$ to increase while keeping the degree of its symmetric arguments (in this example, $W = (x+y, xy)$) fixed. See also footnotes 9 and 10 of Altonji and Matzkin (2005) for a detailed discussion.

However, Assumption 4′ requires more residual variation in $X_{it}$ given $(V_{it}, W_i)$ and can be restrictive if nonparametric justifications are used for Assumption 2. I summarize the proposal by Altonji and Matzkin (2005) to achieve such residual variation. Specifically, Altonji and Matzkin (2005) suggest to

(i) leverage unconditional variation in the exogenous regressors that do not enter $W_i$, and I provide more details at the end of this section;

(ii) restrict the conditional distribution of $A_i$ given $W_i$ to depend on not all but only a subset of coordinates of $W_i$, which is testable by comparing the fit of $\mathbb{E}[Y_{it}|\, X_{it}, V_{it}, W_i]$ with $W_i$ replaced by a proper subset of it;

(iii) assume $f_{A_i|W_i}$ to depend on $W_i$ only through a linear combination of the elements of $W_i$, i.e., $f_{A_i|W_i} = f_{A_i|\sum_{k=1}^{d_W} c_k W_{i,k}}$ where $W_i = (W_{i,1}, \dots, W_{i,d_W})'$ and the $c_k$s are unknown parameters, which implies that

$$\mathbb{E}[Y_{it}|\, X_{it}, V_{it}, W_i] = \mathbb{E}\left[Y_{it}|\, X_{it}, V_{it}, \sum_{k=1}^{d_W} c_k W_{i,k}\right]$$

and the $c_k$s can be estimated using Ichimura and Lee (1991);

(iv) impose a priori restrictions on model (2.1)–(2.2) such that $W_i$ enters $\mathbb{E}[Y_{it}|\, X_{it}, V_{it}, W_i]$ in a parametric way (e.g.,

$$\mathbb{E}[\beta_{it}|\, V_{it}, W_i] = \sum_{k=1}^{d_W} c_k W_{i,k} + \sum_{l=1}^{d_V} d_l V_{it,l}$$

when $d_X = 1$); and,

(v) directly impose functional restrictions on $\mathbb{E}[Y_{it}|\, X_{it}, V_{it}, W_i]$.

Finally, to see how the *unconditional* variations in the exogenous regressors help to satisfy the required residual variation in endogenous regressors, suppose $X_{it,1}$ is endogenous while $X_{it,-1}$ (all other coordinates of $X_{it}$ except $X_{it,1}$) are exogenous. Suppose the conditional support of $X_{it,1}$ given $(V_{it}, W_i)$ only contains a non-zero singleton, while the unconditional support of $X_{it,-1}$ contains a small ball of positive radius. Denote $b_1(V_{it}, W_i) = \mathbb{E}[\beta_{it}|\, V_{it}, W_i]$. Then, I can identify the first-order moments of $\beta_{it}$ by taking partial derivative with respect to $X_{it,-1}$ on both sides of

$$\mathbb{E}[Y_{it}|\, X_{it}, V_{it}, W_i] = X_{it,1} b_{1,1}(V_{it}, W_i) + X'_{it,-1} b_{1,-1}(V_{it}, W_i) \tag{A.76}$$

to identify

$$b_{1,-1}(V_{it}, W_i) = \partial \mathbb{E}[Y_{it}|\, X_{it}, V_{it}, W_i] / \partial X_{it,-1}. \tag{A.77}$$

Finally, identification of $b_{1,1}(V_{it}, W_i)$ can be obtained as

$$b_{1,1}(V_{it}, W_i) = \left(\mathbb{E}\left[Y_{it}\middle| X_{it}, V_{it}, W_i\right] - X'_{it,-1} b_{1,-1}(V_{it}, W_i)\right)/X_{it,1}.$$

Note that there are two conditions on $X_{it,-1}$ needed for the argument to work. First, $X_{it,-1}$ needs to be exogenous, so that it does not enter the construction of $V_{it}$ and $W_i$. Because of this, when taking partial derivative with respect to $X_{it,-1}$ in (A.76), it directly leads to (A.77). Second, $X_{it,-1}$ needs to be continuous vector to enable the derivative-based argument. When $X_{it,-1}$ is discrete, one may use arithmetic differencing technique to identify $b_{1,-1}(V_{it}, W_i)$.

# B Additional Empirical Results

I conduct several robustness checks of the empirical findings in Section 5.

**Measurement error**. To address measurement error that is typical in firm-level data, I go beyond the baseline cleaning (removing outliers, negative, and extreme values) and re-estimate the model using alternative deflators and excluding firms in the top and bottom 1% of total output. Table A.1 reports the results; the estimates largely remain in line with the baseline.

**Instrument validity**. I probe instrument validity by varying the geographic aggregation (provincial rather than finer units), using unweighted competitor averages, and adding city fixed effects. As shown in Table A.2, the main conclusion is unchanged. The specifications with city fixed effects yield noisier estimates, likely because many firms have few within-city competitors, making the within-city demeaning numerically less stable.

**Control-function specification**. I assess sensitivity to the control-function specification along two dimensions. First, I consider an alternative construction of $W$ that includes averages of both the instruments and the regressors. Second, I re-estimate the series control function using alternative bases: second-degree polynomials and third-degree cubic splines with a knot at the median. Table A.3 shows that the estimates are essentially unchanged. One systematic difference is that cubic splines tend to deliver larger point estimates and wider confidence intervals across industries—consistent with the simulation evidence in Table A.8—suggesting a curse-of-dimensionality cost when the basis becomes too rich.

**Comparison with classic methods.** I further compare the TERC estimates with standard OLS and a naive IV approach in Table A.4. Overall, the TERC method yields higher capital elasticities and lower labor elasticities than OLS in three sectors; in the remaining two sectors, the estimates are broadly similar. These patterns echo the findings of Olley and Pakes (1996), albeit in a different context and using a different dataset. By

Table A.1: Robustness to Measurement Error

| Textile | TERC | Alt Deflator | Non-Extreme Output |
|---|---|---|---|
| Capital Elasticity | 0.415 | 0.403 | 0.408 |
| 95% CI | [0.393, 0.441] | [0.379, 0.428] | [0.385, 0.436] |
| Labor Elasticity | 0.452 | 0.472 | 0.441 |
| 95% CI | [0.415, 0.485] | [0.436, 0.505] | [0.405, 0.475] |
| **Chemical** | **TERC** | **Alt Deflator** | **Non-Extreme Output** |
| Capital Elasticity | 0.463 | 0.420 | 0.463 |
| 95% CI | [0.430, 0.497] | [0.387, 0.453] | [0.429, 0.492] |
| Labor Elasticity | 0.311 | 0.439 | 0.293 |
| 95% CI | [0.263, 0.361] | [0.387, 0.489] | [0.245, 0.339] |
| **Nonmetallic Mineral** | **TERC** | **Alt Deflator** | **Non-Extreme Output** |
| Capital Elasticity | 0.371 | 0.377 | 0.361 |
| 95% CI | [0.332, 0.410] | [0.341, 0.417] | [0.324, 0.398] |
| Labor Elasticity | 0.311 | 0.314 | 0.291 |
| 95% CI | [0.255, 0.365] | [0.258, 0.365] | [0.238, 0.341] |
| **General Equipment** | **TERC** | **Alt Deflator** | **Non-Extreme Output** |
| Capital Elasticity | 0.433 | 0.434 | 0.422 |
| 95% CI | [0.402, 0.469] | [0.401, 0.474] | [0.393, 0.457] |
| Labor Elasticity | 0.521 | 0.526 | 0.522 |
| 95% CI | [0.479, 0.564] | [0.477, 0.569] | [0.479, 0.566] |
| **Transportation Equipment** | **TERC** | **Alt Deflator** | **Non-Extreme Output** |
| Capital Elasticity | 0.425 | 0.453 | 0.429 |
| 95% CI | [0.393, 0.459] | [0.415, 0.493] | [0.396, 0.463] |
| Labor Elasticity | 0.588 | 0.566 | 0.571 |
| 95% CI | [0.536, 0.644] | [0.508, 0.632] | [0.515, 0.631] |

*Notes:* "Alt Deflator" refers to the price deflators constructed by Brandt et al. (2014). "Non-Extreme Output" stands for the subsample that excludes the top and bottom 1% firms in terms of total output.

contrast, the naive IV estimates are economically implausible, underscoring the importance of accounting for potentially nonlinear, nonseparable heterogeneity through time-varying random coefficients.

Taken together, these exercises indicate that the results are robust to a wide range of alternative specifications and sample restrictions.

# C Simulation

In this section, I examine the finite-sample performance of the series estimators via a simulation study. A discussion of the data generating process (DGP) motivated by production

Table A.2: IV Validity Check

| Textile | TERC | Provincial IV | Unweighted IV | Location FE |
|---|---|---|---|---|
| Capital Elasticity | 0.415 | 0.393 | 0.414 | 0.410 |
| 95% CI | [0.393, 0.441] | [0.368, 0.417] | [0.390, 0.442] | [0.386, 0.436] |
| Labor Elasticity | 0.452 | 0.458 | 0.460 | 0.427 |
| 95% CI | [0.415, 0.485] | [0.423, 0.494] | [0.425, 0.494] | [0.392, 0.462] |
| **Chemical** | **TERC** | **Provincial IV** | **Unweighted IV** | **Location FE** |
| Capital Elasticity | 0.463 | 0.444 | 0.459 | 0.449 |
| 95% CI | [0.430, 0.497] | [0.411, 0.477] | [0.425, 0.492] | [0.409, 0.490] |
| Labor Elasticity | 0.311 | 0.321 | 0.303 | 0.360 |
| 95% CI | [0.263, 0.361] | [0.274, 0.371] | [0.257, 0.352] | [0.300, 0.411] |
| **Nonmetallic Mineral** | **TERC** | **Provincial IV** | **Unweighted IV** | **Location FE** |
| Capital Elasticity | 0.371 | 0.309 | 0.346 | 0.410 |
| 95% CI | [0.332, 0.410] | [0.258, 0.353] | [0.303, 0.387] | [0.370, 0.455] |
| Labor Elasticity | 0.311 | 0.289 | 0.321 | 0.381 |
| 95% CI | [0.255, 0.365] | [0.227, 0.345] | [0.263, 0.377] | [0.321, 0.429] |
| **General Equipment** | **TERC** | **Provincial IV** | **Unweighted IV** | **Location FE** |
| Capital Elasticity | 0.433 | 0.433 | 0.434 | 0.414 |
| 95% CI | [0.402, 0.469] | [0.400, 0.467] | [0.404, 0.471] | [0.378, 0.452] |
| Labor Elasticity | 0.521 | 0.535 | 0.520 | 0.572 |
| 95% CI | [0.479, 0.564] | [0.492, 0.581] | [0.479, 0.563] | [0.523, 0.618] |
| **Transportation Equipment** | **TERC** | **Provincial IV** | **Unweighted IV** | **Location FE** |
| Capital Elasticity | 0.425 | 0.408 | 0.424 | 0.382 |
| 95% CI | [0.393, 0.459] | [0.372, 0.440] | [0.393, 0.460] | [0.346, 0.418] |
| Labor Elasticity | 0.588 | 0.589 | 0.587 | 0.612 |
| 95% CI | [0.536, 0.644] | [0.538, 0.649] | [0.538, 0.645] | [0.558, 0.673] |

*Notes:* "Provincial IV" denotes the leave-one-out weighted input prices of the competitors in the same province (c.f., baseline IV constructed at city level), with the weights given by the firm's total debt (for interest rate) and number of employees (for wages). "Unweighted IV" stands for the leave-one-out unweighted input prices of the competitors in the same city. "Location FE" includes a city fixed effect in the three-step series regression.

Table A.3: Sensitivity to Control-Function Specification

| Textile | TERC | Alt $W$ | Alt Basis I | Alt Basis II |
|---|---|---|---|---|
| Capital Elasticity | 0.415 | 0.416 | 0.385 | 0.445 |
| 95% CI | [0.393, 0.441] | [0.394, 0.443] | [0.364, 0.407] | [0.423, 0.478] |
| Labor Elasticity | 0.452 | 0.448 | 0.438 | 0.486 |
| 95% CI | [0.415, 0.485] | [0.413, 0.482] | [0.406, 0.469] | [0.451, 0.528] |
| **Chemical** | **TERC** | **Alt $W$** | **Alt Basis I** | **Alt Basis II** |
| Capital Elasticity | 0.463 | 0.462 | 0.434 | 0.490 |
| 95% CI | [0.430, 0.497] | [0.429, 0.495] | [0.402, 0.464] | [0.450, 0.534] |
| Labor Elasticity | 0.311 | 0.312 | 0.250 | 0.349 |
| 95% CI | [0.263, 0.361] | [0.265, 0.362] | [0.206, 0.291] | [0.297, 0.406] |
| **Nonmetallic Mineral** | **TERC** | **Alt $W$** | **Alt Basis I** | **Alt Basis II** |
| Capital Elasticity | 0.371 | 0.357 | 0.334 | 0.370 |
| 95% CI | [0.332, 0.410] | [0.317, 0.396] | [0.298, 0.369] | [0.320, 0.415] |
| Labor Elasticity | 0.311 | 0.303 | 0.331 | 0.303 |
| 95% CI | [0.255, 0.365] | [0.245, 0.356] | [0.280, 0.386] | [0.239, 0.358] |
| **General Equipment** | **TERC** | **Alt $W$** | **Alt Basis I** | **Alt Basis II** |
| Capital Elasticity | 0.433 | 0.428 | 0.407 | 0.442 |
| 95% CI | [0.402, 0.469] | [0.398, 0.464] | [0.377, 0.441] | [0.408, 0.484] |
| Labor Elasticity | 0.521 | 0.513 | 0.496 | 0.530 |
| 95% CI | [0.479, 0.564] | [0.471, 0.556] | [0.455, 0.540] | [0.486, 0.579] |
| **Transportation Equipment** | **TERC** | **Alt $W$** | **Alt Basis I** | **Alt Basis II** |
| Capital Elasticity | 0.425 | 0.424 | 0.411 | 0.416 |
| 95% CI | [0.393, 0.459] | [0.393, 0.458] | [0.383, 0.445] | [0.376, 0.451] |
| Labor Elasticity | 0.588 | 0.586 | 0.525 | 0.616 |
| 95% CI | [0.536, 0.644] | [0.539, 0.644] | [0.471, 0.574] | [0.563, 0.678] |

*Notes:* "Alt $W$" means including the time average of both the IVs and regressors into the $W$ vector. "Alt Basis I" represents the second-degree polynomial basis function. "Alt Basis II" denotes the third-degree spline basis function.

Table A.4: Comparison with OLS and Naive IV Estimates

| Textile | TERC | OLS | Naive IV |
|---|---|---|---|
| Capital Elasticity | 0.415 | 0.379 | 1.309 |
| 95% CI | [0.393, 0.441] | [0.358, 0.400] | [0.889, 1.729] |
| Labor Elasticity | 0.452 | 0.509 | 0.266 |
| 95% CI | [0.415, 0.485] | [0.481, 0.537] | [-0.281, 0.813] |
| **Chemical** | **TERC** | **OLS** | **Naive IV** |
| Capital Elasticity | 0.463 | 0.452 | 0.555 |
| 95% CI | [0.430, 0.497] | [0.419, 0.484] | [-0.094, 1.204] |
| Labor Elasticity | 0.311 | 0.339 | -0.757 |
| 95% CI | [0.263, 0.361] | [0.291, 0.386] | [-1.302, -0.212] |
| **Nonmetallic Mineral** | **TERC** | **OLS** | **Naive IV** |
| Capital Elasticity | 0.371 | 0.382 | -0.670 |
| 95% CI | [0.332, 0.410] | [0.354, 0.411] | [-4.605, 3.265] |
| Labor Elasticity | 0.311 | 0.384 | -3.512 |
| 95% CI | [0.255, 0.365] | [0.340, 0.427] | [-9.599, 2.575] |
| **General Equipment** | **TERC** | **OLS** | **Naive IV** |
| Capital Elasticity | 0.433 | 0.389 | 0.951 |
| 95% CI | [0.402, 0.469] | [0.363, 0.416] | [0.765, 1.138] |
| Labor Elasticity | 0.521 | 0.501 | -0.619 |
| 95% CI | [0.479, 0.564] | [0.464, 0.538] | [-1.220, -0.019] |
| **Transportation Equipment** | **TERC** | **OLS** | **Naive IV** |
| Capital Elasticity | 0.425 | 0.442 | 0.723 |
| 95% CI | [0.393, 0.459] | [0.408, 0.475] | [0.523, 0.924] |
| Labor Elasticity | 0.588 | 0.560 | 0.508 |
| 95% CI | [0.536, 0.644] | [0.509, 0.612] | [-0.110, 1.127] |

*Notes:* "OLS" denotes the pooled OLS method with constant coefficients to estimate the output elasticities. "Naive IV" represents two-stage least squares method where the IVs are the same as in TERC estimates. CIs are obtained using standard Stata command.

function applications is first provided. Then, I present the baseline results of estimating the APE $\bar{b} \coloneqq T^{-1}\sum_{t=1}^{T}\mathbb{E}\beta_{it}$. Next, I show how the method performs for estimating the LAR $b_t(x) \coloneqq \mathbb{E}\left[\beta_{it}\middle| X_{it}=x\right]$. Finally, as robustness checks, I (i) vary the number of firms and periods, (ii) change the degree of basis functions used for series estimation, and (iii) include ex-post shocks $\widetilde{\epsilon}_{it}$ and $\upsilon_{it}$.

## C.1 DGP

The baseline revenue-based DGP is

$$Y_{it} = k_{it}\beta_{it,K} + l_{it}\beta_{it,L} + \omega_{it} + \widetilde{\epsilon}_{it},$$

where $\omega_{it}$, $\beta_{it,K}$, and $\beta_{it,L}$ are all functions of $(A_i, \varepsilon_{it}, \upsilon_{it})$, $k_{it}$ and $l_{it}$ are the natural logs of optimal capital and labor calculated from the solution to firm $i$'s profit maximization problem, and $Y_{it}$ is the natural log of value-added output measured in dollars. Let $\mathcal{U}[a,b]$ denote the uniform distribution over $[a,b]$. I draw $A_i \sim_{i.i.d.} \mathcal{U}[1,2]$ and $\varepsilon_{it} \sim_{i.i.d.} \mathcal{U}[1,2]$. I let true $\omega_{it} = \ln(A_i + \varepsilon_{it}/2 + 1)$, $\beta_{it,K} = (A_i + \varepsilon_{it})/10$, and $\beta_{it,L} = (A_i + \varepsilon_{it})/10$, and write $\beta_{it} = (\omega_{it}, \beta_{it,K}, \beta_{it,L})'$. I compute the true $\overline{\omega} \coloneqq T^{-1}\sum_{t=1}^{T}\mathbb{E}\omega_{it} = 1.1736$, $\bar{b}_K \coloneqq T^{-1}\sum_{t=1}^{T}\mathbb{E}\beta_{it,K} = .3$, and $\bar{b}_L \coloneqq T^{-1}\sum_{t=1}^{T}\mathbb{E}\beta_{it,L} = .3$. The range of the random coefficients are set such that the second-order condition for the profit maximization problem is satisfied. For the baseline results, I let ex-post shocks $\widetilde{\epsilon}_{it} = \upsilon_{it} = 0$ and investigate their impacts as robustness checks later. Suppose capital and labor choices are made separately (e.g., investment and hiring decisions made by different departments) based on ex-ante signals about $\varepsilon_{it}$ with noise $\lambda_{it,K}$ and $\lambda_{it,L}$: $\eta_{it,K} = \varepsilon_{it} + \lambda_{it,K}$ and $\eta_{it,L} = \varepsilon_{it} + \lambda_{it,L}$, where $\lambda_{it,K}$ and $\lambda_{it,L} \sim_{i.i.d.} \mathcal{U}[-.05, .05]$. Since $Y_{it}$ is measured in dollars which is the case with most real production datasets, the price of output $P_{it}$ is assumed to be 1. I draw each element of the IVs $Z_{it} = (r_{it}, w_{it})'$ from $\mathcal{U}[0, \ln 3]$ independent of each other and all other variables (in particular, $Z_{it} \perp (\varepsilon_{it}, \eta_{it})'$), and solve the firm's profit maximization problem to obtain

$$k_{it} = \frac{r_{it} - (A_i + \eta_{it,K})(r_{it} - w_{it})/10 - \ln((A_i + \eta_{it,K})/10) - \ln(A_i + \eta_{it,K}/2 + 1)}{(A_i + \eta_{it,K})/5 - 1},$$
$$l_{it} = \frac{w_{it} - (A_i + \eta_{it,L})(w_{it} - r_{it})/10 - \ln((A_i + \eta_{it,L})/10) - \ln(A_i + \eta_{it,L}/2 + 1)}{(A_i + \eta_{it,L})/5 - 1}.$$

Let $X_{it} = (k_{it}, l_{it})'$ denote the two endogenous choice variables. It is clear that $X_{it}$ is correlated with $\beta_{it}$ in each period via $(A_i, \eta_{it,K}, \eta_{it,L})$ in a nonseparable way. I use $n$, $T$, and $M$ to denote the total number of firms, periods, and simulations, respectively. The simulated

Table A.5: Performance of $\widehat{\overline{b}}$

| | Formula | $\widehat{\overline{\omega}}$ | $\widehat{\overline{b}}_K$ | $\widehat{\overline{b}}_L$ |
|---|---|---|---|---|
| Bias | $M^{-1}\sum_m \left(\widehat{\overline{b}}_d^{(m)} - \overline{b}_d\right) / \left\lvert\overline{b}_d\right\rvert$ | 2.81% | 3.22% | 3.24% |
| rMSE | $\sqrt{M^{-1}\sum_m \left(\widehat{\overline{b}}_d^{(m)} - \overline{b}_d\right)^2} / \left\lvert\overline{b}_d\right\rvert$ | 2.89% | 3.65% | 3.62% |
| MND | $M^{-1}\sum_m \left\lvert\widehat{\overline{b}}_d^{(m)} - \overline{b}_d\right\rvert / \left\lvert\overline{b}_d\right\rvert$ | 2.81% | 3.26% | 3.26% |

dataset consists of $\left\{X_{it}^{(m)}, Y_{it}^{(m)}, Z_{it}^{(m)}\right\}$ for $i = 1, \ldots, n$, $t = 1, \ldots, T$, and $m = 1, \ldots, M$, which are used to construct $\widehat{\overline{b}}^{(m)}$ and $\widehat{b}_t^{(m)}(x)$ to estimate $\overline{b}$ and $b_t(x)$, respectively, via the estimation procedure outlined in Subsection 4.1. I evaluate the performance of the estimators by their biases, root-mean squared errors (rMSE), and mean normed deviations (MND), with the explicit mathematical definitions provided in the tables below.

## C.2 Results

For the baseline results, I set $n = 1,000$ and $T = 2$, and use a second-degree polynomial spline basis with its knot at the median for all series estimation steps. I run $M = 1,000$ simulations. For each $i$, I construct $W_i$ as the mean over time of each coordinate of $X_{it}$. Following the theory, I set $V_{it} := \left(F_{k_{it}|Z_{it},W_i}\left(k_{it}\middle|\, Z_{it}, W_i\right), F_{l_{it}|Z_{it},W_i}\left(l_{it}\middle|\, Z_{it}, W_i\right)\right)'$. The performance of $\widehat{\overline{\omega}}$, $\widehat{\overline{b}}_K$, and $\widehat{\overline{b}}_L$ is summarized in Table A.5. For notational simplicity, I use $\sum_m$ for $\sum_{m=1}^M$ and $\sum_{i,t,m}$ for $\sum_{i=1}^n \sum_{t=1}^T \sum_{m=1}^M$.

The first row of Table A.5 reports the normalized bias of each coordinate of $\widehat{\overline{b}}$. The bias is reasonably small across all three coordinates, with a magnitude between 2.81% and 3.24% of the length of the corresponding coordinate of $\overline{b}$. The second row shows the normalized rMSE of each coordinate of $\widehat{\overline{b}}$. My method achieves low normalized rMSEs between 2.89% and 3.65% of the size of the corresponding coordinate $\overline{b}_d$ of $\overline{b}$. The last row presents the normalized MNDs of each coordinate of $\widehat{\overline{b}}$. Again, the method performs well with an MND between 2.81% and 3.26% of the size of the corresponding coordinate $\overline{b}_d$ of $\overline{b}$.

Next, I investigate the performance of $\widehat{b}_t(x)$, which is obtained following the estimation procedure outlined in Subsection 4.1. For the true $b_t(x)$, due to the complex dependence structure of $X_{it}$ on $\beta_{it}$, there is no analytical solution available. Therefore, for each $t$ I pool the observations across all $M$ simulations ($n \times M$ observations for each $t$) and approximate $b_t(x)$ by regressing the true $\beta_{it}$ on the same spline basis functions of $X_{it}$.

Table A.6: Performance of $\widehat{b}_t(x)$

| | Formula | $\widehat{\omega}_t(x)$ | $\widehat{b}_{t,K}(x)$ | $\widehat{b}_{t,L}(x)$ |
|---|---|---|---|---|
| rMSE | $\sqrt{(NTM)^{-1}\sum_{i,t,m}\left(\widehat{b}_{t,d}^{(m)}\left(x_{it}^{(m)}\right)-b_{t,d}\left(x_{it}^{(m)}\right)\right)^2}/\left\lvert\bar{b}_d\right\rvert$ | 3.36% | 5.57% | 5.59% |
| MND | $(NTM)^{-1}\sum_{i,t,m}\left\lvert\widehat{b}_{t,d}^{(m)}\left(x_{it}^{(m)}\right)-b_{t,d}\left(x_{it}^{(m)}\right)\right\rvert/\left\lvert\bar{b}_d\right\rvert$ | 2.87% | 4.36% | 4.37% |

Table A.6 summarizes the results. Note that by definition the bias of $\widehat{b}_t(x)$ is the same as that of $\widehat{\bar{b}}$, so I omit it here. The normalized rMSE of $\widehat{b}_t(x)$ is bigger than that of $\widehat{\bar{b}}$, with a magnitude between 3.36% and 5.59% of the size of the corresponding $\bar{b}_d$. The normalized MND follows a similar pattern. The performance of $\widehat{b}_t(x)$ for estimating $b_t(x)$ is not as good as that of $\widehat{\bar{b}}$ for $\bar{b}$, which is expected because (i) $b_t(x)$ is a function rather than a finite-dimensional vector and (ii) there is approximation error in calculating the true $b_t(x)$ via simulations.

To show how robust my method is in estimating $\bar{b}$ and $b_t(x)$, I conduct another set of exercises. I evaluate the performance of my TERC estimator using rMSE defined as

$$\sqrt{M^{-1}\sum_m\left\|\widehat{\bar{b}}^{(m)}-\bar{b}\right\|^2}/\left\|\bar{b}\right\|$$

for the vector $\bar{b}$, and

$$\sqrt{(NTM)^{-1}\sum_{i,t,m}\left\|\widehat{b}_t^{(m)}\left(x_{it}^{(m)}\right)-b_t\left(x_{it}^{(m)}\right)\right\|^2}/\left\|\bar{b}\right\|$$

for the vector $b_t(x)$. First, I vary $n$ and $T$, and present the results in Table A.7. As expected, a larger $n$ is good for overall performance. However, the magnitude in the improvement of performance is mild, possibly because with more agents I need more data to control for the increasing degree of heterogeneity. On the other hand, I find that the method performs reasonably well even with a small sample size of $n = 500$. Having a larger $T$ improves the performance, which is again expected as I can exploit more information from repeated observations of the same individual to better control for the fixed effect $A_i$.

Table A.7: Performance under Varying $n$ and $T$

| $n$ ($T = 2$) | rMSE: $\bar{b}$ | rMSE: $b_t(x)$ | $T$ ($n = 1,000$) | rMSE: $\bar{b}$ | rMSE: $b_t(x)$ |
|---|---|---|---|---|---|
| 500 | 3.25% | 4.23% | 2 | 2.99% | 3.69% |
| 1,000 | 2.99% | 3.69% | 3 | 2.74% | 3.38% |
| 2,000 | 2.87% | 3.43% | 4 | 2.66% | 3.27% |

Second, I vary the degree of the spline bases used to construct the series estimators for all steps. The knot is placed at the median for all specifications. Table A.8 contains the results. I find that increasing the degree of basis functions from one to two improves estimation accuracy significantly. When I increase the degree of basis functions from two to three, the performance deteriorates. When the degree becomes larger, there are more regressors in each step of regression, leading to a possible over-fitting problem. Based on this result, I use the second-degree splines in the empirical application. Note that one may also use the AIC criterion to select the degree of basis functions.

Table A.8: Performance under Varying Degree of Basis Functions

| Degree of Basis Functions | rMSE: $\bar{b}$ | rMSE: $b_t(x)$ |
|---|---|---|
| 1 | 5.27% | 6.41% |
| 2 | 2.99% | 3.69% |
| 3 | 4.82% | 6.40% |

Lastly, I examine how including the ex-post shocks $\tilde{\epsilon}_{it}$ and $\upsilon_{it}$ into the model affects the performance of my estimators. I draw $\tilde{\epsilon}_{it} \sim \mathcal{U}[-.25, .25]$, the ex-post shock to the main equation (2.1), independently of all the other variables. $\tilde{\epsilon}_{it}$ can also be interpreted as an ex-post shock to $\omega_{it}$. For $\upsilon_{it}$, since ex-post shock to $\omega_{it}$ has been considered by $\tilde{\epsilon}_{it}$, I draw $\upsilon_{it}^{s} \sim \mathcal{U}[-.1, .1]$ independently of all the other variables and use $\beta_{it,s}^{\text{new}} = \beta_{it,s} + \upsilon_{it,s}$ in (2.1) for $s \in \{K, L\}$. Results are presented in Table A.9. As expected, adding ex-post errors $\tilde{\epsilon}_{it}$ or $\upsilon_{it}$ negatively affects the performance of the estimators. When $\tilde{\epsilon}_{it}$ is included, the rMSE of $\widehat{\bar{b}}$ for estimating $\bar{b}$ increases from 2.99% to 3.64% and the rMSE of $\widehat{b}_t(x)$ for estimating $b_t(x)$ rises from 3.64% to 5.43%. The effect of including $\upsilon_{it}$ on the performance is similar.

Table A.9: Performance with and without Ex Post Shocks

| Add $\epsilon_{it}$ to $Y_{it}$? | rMSE: $\bar{b}$ | rMSE: $b_t(x)$ | Add $\upsilon_{it}$ to $\beta_{it}$? | rMSE: $\bar{b}$ | rMSE: $b_t(x)$ |
|---|---|---|---|---|---|
| No | 2.99% | 3.69% | No | 2.99% | 3.69% |
| Yes | 3.64% | 5.43% | Yes | 3.57% | 5.19% |