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

Ming Li†

November 20, 2024

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 identification 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 conditional 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.

Keywords: Correlated random coefficients, panel data, time-varying endogeneity, index exclusion, conditional control variable, average partial effect, local average response function.

*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, Âureo de Paula, Wayne Yuan 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 several anonymous referees for their suggestions, which greatly improved the paper. All remaining errors are my own.

†Department of Economics, 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 instance, it is possible that a firm with a high fixed effect (e.g., strong management skills) endogenously selects lower input levels during certain periods. Such behavior could arise if the firm has information about its time-varying output efficiency shock, in addition to its time-invariant management skill, when optimally determining input values.

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\(^1\) consists of two equations:

\[
\begin{align*}
Y_{it} &= X'_{it} \beta (A_i, \varepsilon_{it}), \\
X_{it} &= g (Z_{it}, A_i, \eta_{it}).
\end{align*}
\]

In (1.1), I model the vector of random coefficients \(\beta_{it} = \beta (A_i, \varepsilon_{it})\) as a vector of unknown and possibly nonlinear functions \(\beta (\cdot)\) 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, unknown and possibly nonlinear function \(g (\cdot)\) of the vector of exogenous instrumental variables (IV) \(Z_{it}\), fixed effect \(A_i\), and per-period information \(\eta_{it}\) about \(\varepsilon_{it}\). The motivation for (1.2) is that the agent \(i\) in period \(t\) optimally chooses values of her regressors \(X_{it}\) by solving

\(^1\)In the baseline model, I focus exclusively on endogenous regressors to address the issue of time-varying endogeneity through the random coefficients. The analysis remains largely unchanged when exogenous regressors are included in (1.1). I specify the necessary changes at the end of the proof of Theorem 1 in the Appendix.
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\(^2\), and any pair of the three variables can be correlated. The functional forms are potentially different across different coordinates of \(\beta(\cdot)\) (and also \(g(\cdot)\)).

As an analyst, I observe \(\{X_{it}, Y_{it}, Z_{it}\}\) for \(i = 1, ..., n\) and \(t = 1, ..., T\) and aim to identify the average partial effect \(E_{\beta_{it}}\) (APE, see Graham and Powell (2012)) and the local average response function \(E[\beta_{it} | X_{it}]\) (LAR, see Altonji and Matzkin (2005)).

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 \(X_{it}\). I present three empirical examples in Section 2 that demonstrate the validity of such correlation in real-life scenarios. The associated technical challenge is that I need to control for the time-varying endogeneity through the random coefficients when both unknown (to the analyst) \(A_i\) and \(\eta_{it}\) appear in \(g(\cdot)\) in a nonlinear and nonseparable way in (1.2). Furthermore, I do not specify the distributions of \(A_i, \varepsilon_{it}, \) or \(\eta_{it}\), nor the functional form of \(\beta(\cdot)\) or \(g(\cdot)\). In this sense, the TERC model is a fixed effect panel data model (Laage (2024)). It has many important applications in economics; e.g., heterogeneous Cobb-Douglas production function estimation, return to schooling estimation, labor supply estimation, Engel curve analysis, and demand analysis, among others (Blundell, MaCurdy, and Meghir (2007b), Blundell, Chen, and Kristensen (2007a), Wooldridge (2009), Chernozhukov, Hausman, and Newey (2019), Li and Sasaki (2024), Keiller, de Paula, and Van Reenen (2024)).

I introduce a new panel data-based identification strategy to identify the APE and LAR for the TERC model. The idea is to control for \(A_i\) and \(\eta_{it}\) via a sufficient statistic and conditional control variable, respectively, such that, conditional on these two controls, the residual variation in \(X_{it}\) is driven solely by the exogenous \(Z_{it}\). In other words, the residual variation in \(X_{it}\) is causal and thus can be exploited to identify the parameters of interest. Specifically, since the APE and LAR are generally

\(^2\)In the baseline model, I assume in Assumption 1 that each coordinate of \(g(\cdot)\) depends monotonically on a single coordinate of \(\eta_{it}\) to invoke the control function argument. As an extension, I discuss how to relax this assumption to allow the entire vector \(\eta_{it}\) to enter certain coordinates of \(g(\cdot)\) at the end of Section 3.
not identified if the distribution of $A_i \mid X_i$ is left unrestricted (Liu, Poirier, and Shiu (2024)), I first impose an index exclusion assumption that supplies a sufficient statistic $W_i$ for $A_i$. I present parametric and nonparametric justifications from the literature for this assumption. Next, I construct a feasible conditional cumulative distribution function (CDF) $V_{it} = F_{X_{it} \mid Z_{it}, W_i} (X_{it} \mid Z_{it}, W_i)$ to control for $\eta_{it}$. I show that, under the maintained assumptions, $V_{it} = F_{\eta_{it} \mid A_i, W_i} (\eta_{it} \mid A_i, W_i)$ and is thus strictly increasing in $\eta_{it}$ given $A_i$ and $W_i$. Finally, conditioning on $V_{it}$ and $W_i$ which effectively fixes both $\eta_{it}$ and $A_i$, by (1.2) the residual variation in $X_{it}$ is determined solely by $Z_{it}$, which allows me to identify the APE and LAR via (1.1). At the end of Section 3, I present three extensions: (i) allowing vector-valued $\eta_{it}$ to enter certain coordinates of $g(\cdot)$ function, (ii) identifying the second-order moments of $\beta_{it}$, and (iii) introducing ex-post shocks to $\beta_{it}$ and $Y_{it}$, as well as incorporating exogenous covariates into $X_{it}$.

The constructive identification analysis leads to easy-to-implement series estimators for the APE and LAR. I derive the convergence rates and establish the asymptotic normality of the proposed estimators. The inference results presented in this paper build on the literature on the asymptotic analysis of multi-step series estimators (Andrews (1991), Newey (1997), Imbens and Newey (2009)). The key departures from the existing literature include the fact that the object of interest is a partial mean process (Newey (1994)) of the derivative of the second-step estimator, and that the final step of the three-step estimator is an unknown but estimable functional of a conditional expectation function. Consequently, the estimation errors from each step must be properly accounted for to ensure accurate large sample properties.

As an empirical illustration, I estimate a Cobb-Douglas production function with heterogeneous output elasticities for Chinese manufacturing firms. My estimated APEs are broadly consistent with those derived from applying the classical methods with constant coefficients (Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, Caves, and Frazer (2015)) to the same data. I also estimate the LARs evaluated at the realized input levels for each firm and find significant variation across firms. To support the empirical results, I conduct simulations motivated by production function estimation applications to evaluate the finite-sample performance of

---

3 $V_{it}$ is referred to as a “conditional” control variable because, strictly speaking, it is not a one-to-one mapping of $\eta_{it}$ unless the unknown $A_i$ is also conditioned upon. To address this, I use the law of iterated expectations (LIE), first conditioning on $A_i$ in the inner expectation. The details of the identification analysis are provided in Section 3.

4 Note that, with constant coefficient models, the LAR is equivalent to the APE.
the estimators proposed in this paper. The results demonstrate that the proposed method performs well.

Next, I discuss the contributions and limitations of this paper. The first contribution concerns the features of the TERC model in (1.1)–(1.2). Allowing $X_{it}$ to be correlated with $\beta_{it}$ not only through $A_i$ but, more importantly, through $\eta_{it}$ significantly enhances the model’s flexibility and carries important empirical implications, as demonstrated by the examples discussed so far. Another novel aspect of the TERC model is the inclusion of $A_i$ in both (1.1) and (1.2). In particular, I argue that including $A_i$ in the first-step function $g(\cdot)$ better reflects the reality that $A_i$ is typically in the information set of agent $i$ when she optimally chooses the value of $X_{it}$, for instance, through repeated learning. Moreover, $A_i$ enters both (1.1) and (1.2) in a nonlinear and nonseparable way, which naturally arises from agent optimization behavior. Consequently, enabling these features enhances the model’s flexibility and broadens its applicability.

The second contribution is the proposal of a new panel data-based method to identify the APE and LAR of the TERC model. I construct a sufficient statistic $W_i$ to control for the time-invariant $A_i$ and a per-period conditional control variable $V_{it}$ to control for the time-varying shock $\eta_{it}$. I prove that, conditional on $V_{it}$ and $W_i$, the residual variation in $X_{it}$ is entirely attributable to the exogenous $Z_{it}$, thereby enabling the identification of the APE and LAR. The identification method proposed in this paper also contributes to the literature on control functions with multiple unobservables in nonseparable regression models (e.g., Blundell, Kristensen, and Matzkin (2013), Kasy (2014)).

The final contribution involves implementing the proposed method and characterizing its statistical properties. I develop three-step series estimators for the APE and LAR, establishing their convergence rates and asymptotic normality. The effectiveness of the method is demonstrated using a production dataset of Chinese manufacturing firms, revealing significant heterogeneity in capital and labor elasticities across firms. To further support the empirical findings, I conduct a simulation study inspired by applications in production function estimation, which produces satisfactory results.

I discuss two limitations of this paper. First, although both the unknown $A_i$ and $\eta_{it}$ determine the value of $X_{it}$, the baseline model imposes a restriction such that
each coordinate of \( g(\cdot) \) depends on a single coordinate of \( \eta_{it} \) (as well as \( Z_{it} \) and \( A_i \)). As noted by Imbens and Newey (2009), this restriction excludes nonseparable supply and demand models with one time-varying disturbance per equation. To relax this restriction, I propose two approaches under additional empirically motivated assumptions at the end of Section 3. Second, my method requires IVs. I provide a few examples of IVs for the applications discussed in this paper. However, depending on the specific application, identifying suitable IVs can still be challenging.

1.1 Related Literature

This paper contributes to the literature on CRC linear panel models. I highlight a few studies that are particularly relevant to this work. Chamberlain (1992) studies the regular identification of the first-order moments of the random coefficients when \( T > d_X \), where \( T \) is the number of periods and \( d_X \) is the dimension of regressors. Wooldridge (2005) derives conditions for consistency of the fixed-effects estimators of the APE for a class of linear and nonlinear panel data models with time-invariant random coefficients. The condition requires the random coefficients to be mean independent of the residuals from regressing the covariates on the aggregate time variables. Graham and Powell (2012) prove irregular identification when \( T = d_X \) and allow for more persistent processes in the regressors. They leverage different identifying content of the subpopulations of “movers” and “stayers” to identify the first-order moments of the time-varying random coefficients. Arellano and Bonhomme (2012) identify the first-order moments of the time-invariant random coefficients by using the within-group and between-group variation in the regressors. They rely on a mean independence assumption between the residuals and both the regressors and the random coefficients. They further identify the second-order moments and distribution functions of the random coefficients by exploiting assumptions on the time-dependence structure of the residuals. Graham, Hahn, Poirier, and Powell (2018) show how to identify quantiles of the random coefficients in a general linear quantile regression model. Masten (2018) considers a classical linear simultaneous equations model with random coefficients on the endogenous variables. By extending the Chamberlain-Mundlak approach for balanced panels, Wooldridge (2019) develops estimation strategies for CRC linear panel models that account for sample selection in unbalanced panels. Laage (2024) studies a time-invariant CRC linear panel data
model with additive scalar-valued fixed effects. She proposes a two-step identification approach, where in the first step she identifies nonparametrically the conditional expectation of the disturbances given the regressors and the control variables, and then in the second step she uses between-group variation to identify the APE.

The key difference between my paper and the CRC literature discussed above is that my model, \((1.1)-(1.2)\), features time-varying random coefficients that can be correlated with the regressors through both fixed effects and time-varying shocks. None of the papers, except Graham and Powell (2012), explicitly considers such correlation. Graham and Powell (2012) impose a time-stationarity assumption on the time-varying random shocks given the regressors for all periods and the fixed effect, and exploit the different identifying contents of the subpopulations for identification. The novel method proposed by Graham and Powell (2012) covers a wide range of empirical applications. However, their main time-stationarity assumption effectively rules out the time-varying endogeneity through the random coefficients, wherein \(X_{it}\) and \(\beta_{it}\) are correlated through both \(A_i\) and \(\eta_{it}\), which is the focus of this paper. Instead, I tackle this issue by proposing two controls for \(A_i\) and \(\eta_{it}\) and, subsequently, using the residual variation in \(X_{it}\) to identify the APE and LAR. It should be clarified that my work is not necessarily more general than that of Graham and Powell (2012), as the set of assumptions used by each is not nested within the other’s. Instead, I consider this paper an initial step toward addressing the important issue of time-varying endogeneity through the random coefficients, which existing methods have not yet fully resolved.

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

\[ \eta_{it} \sim d \eta_{it} | X_i, A_i, \text{ for } t \neq s. \] To see why it rules out time-varying endogeneity through the random coefficients, consider a simplified example in which the number of periods \(T = 2\) and the data generating process (DGP) of a scalar-valued \(X_{it}\) is \(X_{it} = Z_{it}^{\prime} \lambda + A_i + \eta_{it}\). Suppose one observes \(X_{i2} > X_{i1}\) and \(Z_{i2} = Z_{i1}\) in the data, which by the DGP of \(X_{it}\) implies \(\eta_{i2} > \eta_{i1}\) deterministically given \(X_i\) and \(A_i\). Consequently, the time-stationarity condition is violated when the time-varying endogeneity through the random coefficients is present.
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. My method requires the existence of a sufficient statistic $W_i$ for the time-invariant $A_i$ and a feasible control variable $V_{it}$ for the time-varying $\eta_{it}$. To justify the sufficiency of $W_i$ for $A_i$, I present 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 (2024) for a detailed discussion on the index sufficiency condition for the fixed effects in nonlinear semiparametric panel data models. It may be worth pointing out that the exchangeability condition used by Altonji and Matzkin (2005) concerns the conditional probability density function (PDF) of $A_i$ and $\eta_{it}$ given $X_i$. I show that it is not directly applicable to the TERC model in which $X_{it}$ depends on $\eta_{it}$ for each $t$. Instead, I impose a lower-level exchangeability condition involving only the unobservable $A_i$ and $\eta_{it}$ and use it to establish a sufficiency result for $A_i$ in Proposition 1.

To find the feasible control variable for $\eta_{it}$ in (1.2), I generalize the technique of Imbens and Newey (2009) in two nontrivial ways. First, because the unknown $A_i$ enters equation (1.2) together with $\eta_{it}$, the conditional CDF $F_{X_{it}|Z_{it}} (X_{it}|Z_{it})$ proposed by Imbens and Newey (2009) does not uniquely determine $\eta_{it}$ even if it is a scalar. Instead, I use the index exclusion condition on $A_i$ to construct a feasible conditional control variable $V_{it} = F_{X_{it}|Z_{it},W_i} (X_{it}|Z_{it},W_i)$ and show that it controls for $\eta_{it}$ given $A_i$ and $W_i$. Second, I allow $\eta_{it}$ to be vector-valued, enabling the model to capture the rich heterogeneity commonly observed in empirical applications. In the baseline model, I assume that each coordinate of $X_{it}$ depends upon a single coordinate of $\eta_{it}$. As an extension, I discuss how to relax this restriction when additional empirically motivated assumptions are imposed. Moreover, constructing $V_{it}$ requires IVs that satisfy specific exogeneity conditions. Murtazashvili and Wooldridge (2008) provide a set of conditions sufficient for the consistency of a general class of fixed effects IV estimators in a time-invariant individual-specific CRC panel data model. The model and method of this paper are different from theirs.

The asymptotic analysis in this paper builds on a series of foundational works. Andrews (1991) investigates the asymptotic properties of series estimators for nonparametric and semiparametric regression models. His results are applicable to a wide variety of estimands, including derivatives and integrals of the regression function.
I use his results to cover vector-valued functionals of regression functions. Newey (1997) also studies series estimators and provides conditions for proving convergence rates and asymptotic normality for the estimators of conditional expectations. Newey, Powell, and Vella (1999) propose a two-step nonparametric estimator for a triangular simultaneous equation model with a separable first-step equation. They derive asymptotic normality for their two-step estimator with the first-step estimator plugged in. Imbens and Newey (2009) also analyze a triangular simultaneous equation model, but with a nonseparable first-step equation. They obtain mean-squared convergence rates for the first-step estimator and prove asymptotic normality (in their 2002 working paper version) for known functionals of the conditional expectation of the outcome variable given the regressors and control variables. I extend their analysis to establish asymptotic normality for unknown but estimable functionals of conditional expectation functions.

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 of all theorems, Proposition 1, and a simulation study are included in Appendix A, B, and C, respectively.

2 Model

Let \( i \in \{1, ..., n\} \) index \( n \) agents and \( t \in \{1, ..., T\} \) index finite \( T \geq 2 \) periods. Let \( d_X \) be the dimension of \( X_{it} \) and similarly for \( Z_{it}, A_i, \varepsilon_{it}, \) and \( \eta_{it} \). Consider the following baseline TERC model

\[
Y_{it} = X_{it}' \beta \left( A_i, \varepsilon_{it} \right), \quad (2.1)
\]

\[
X_{it} = g \left( Z_{it}, A_i, \eta_{it} \right), \quad (2.2)
\]

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 a vector-valued, unknown, and potentially nonlinear
function $\beta(\cdot)$ of $A_i$ and $\varepsilon_{it}$. Here, $A_i$ represents a fixed effect and $\varepsilon_{it}$ drives the time-varying behavior of $\beta_{it}$, both of which can be either scalar- or vector-valued. Furthermore, $A_i$ and $\varepsilon_{it}$ can be arbitrarily correlated.

- $\eta_{it} \in \mathbb{R}^{d_{\eta}}$ is a continuously distributed and time-varying random vector that influences the value of $X_{it}$. $\eta_{it}$ is allowed to be correlated with both $A_i$ and $\varepsilon_{it}$. Additionally, different coordinates of $\eta_{it}$ can also be correlated.

- $X_{it} \in \mathbb{R}^{d_X}$ is a vector of choice variables for agent $i$ in period $t$. $Y_{it} \in \mathbb{R}$ denotes the scalar-valued outcome variable, and $Z_{it} \in \mathbb{R}^{d_Z}$ represents a vector of IVs.

- $g(\cdot)$ is an unknown vector-valued function of $Z_{it}$, $A_i$, and $\eta_{it}$ that determines the value of each coordinate of $X_{it}$.

Denote $X_i = (X_{i1}, \ldots, X_{iT})'$, $Y_i = (Y_{i1}, \ldots, Y_{iT})'$, $Z_i = (Z_{i1}, \ldots, Z_{iT})'$, $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$, and $\eta_i = (\eta_{i1}, \ldots, \eta_{iT})'$. Denote superscript $(l)$ the $l$th coordinate of a vector. I use $\sim_d$ to indicate that two random variables are identically distributed, $F$ for CDF, $f$ for PDF, $E$ for expectation, $V$ for variance, $I$ for $d_X \times d_X$ identity matrix, $\xrightarrow{p}$ for convergence in probability, $\xrightarrow{d}$ for convergence in distribution, and $\xrightarrow{a.s.}$ for almost sure convergence. I use upper case letters for random variables and lower case letters for their values.

To clarify the model’s information structure, first nature draws $(A_i, \varepsilon_{it})$ for each agent and period in the economy. Then, agent $i$ in period $t$ chooses her endogenous $X_{it}$ by solving an optimization problem (e.g., profit maximization) after observing $A_i$, some signal $\eta_{it}$ about $\varepsilon_{it}$, and exogenously determined IVs $Z_{it}$, leading to (2.2). Finally, agent $i$ obtains its outcome $Y_{it}$ via (2.1). As an analyst, I observe $\{Y_{it}, X_{it}, Z_{it}\}$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$ and aim to identify $E\beta_{it}$ (APE) and $E[\beta_{it} | X_{it}]$ (LAR). I illustrate the information structure in Figure 1.

---

For clarity of argument, in the baseline model I assume all coordinates of $X_{it}$ to be endogenous in the sense that they can be correlated with $\beta_{it}$ through both $A_i$ and $\eta_{it}$. In fact, the proposed method in this paper can benefit from additional exogenous regressors in (2.1). I discuss it in relation to Assumption 4 and note the changes needed when additional exogenous regressors are included in (2.1) at the end of the proof of Theorem 1.
Model (2.1)–(2.2) appears in various economic applications. I describe three of them here.

Example 1 (Heterogeneous 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 $\beta_{it} = \beta (A_i, \varepsilon_{it}) \in \mathbb{R}^2$ are modeled as a two-dimensional function of the firm fixed effect $A_i$ (e.g., manager ability) and a time-varying random shock $\varepsilon_{it}$ (e.g., realized per-period technology shock). When deciding its input values $X_{it} \in \mathbb{R}^2$ for 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 some exogenously determined input prices $Z_{it}$, such as interest rates and wages, and that the cost function is given by $c (x, z)$. Then, firm $i$ chooses input values $X_{it}$ by maximizing its expected profit with the knowledge of $(Z_{it}, A_i, \eta_{it})$, i.e.,

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

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

Example 2 (Labor Supply Estimation). Suppose agent $i$ has a linear labor supply function of the form (2.1), where $Y_{it}$ represents the number of hours worked and $X_{it}$ includes the endogenous hourly wage along with other exogenous demographic variables. The coordinate of $\beta_{it}$ corresponding to the wage is the key object of interest, as it quantifies how labor supply responds to variations in wage rate over time. Given some exogenous IVs $Z_{it}$ (e.g., county minimum wage and non-labor income), individual capability $A_i$, and expected health shocks $\eta_{it}$ (e.g., the likelihood of contracting Covid-19 at work), agent $i$ chooses a job with the wage that maximizes her expected utility, resulting in (2.2).

Example 3 (Engel Curve Estimation). Suppose the budget share of gasoline $Y_{it}$ for household $i$ at time $t$ is a function of gas price and total expenditure represented in (2.1). Here, $\beta_{it} \in \mathbb{R}^2$ is modeled as a vector of unknown functions of the household fixed effect $A_i$ and a wealth shock $\varepsilon_{it}$, and captures the elasticity of gasoline demand with respect to total expenditure and gas price, respectively. Given $A_i$, a signal $\eta_{it}$ about the wealth shock $\varepsilon_{it}$, and an IV $Z_{it}$ (e.g., head of household’s gross income), household $i$ searches for a gas price and plans its total expenditure budget by maximizing its expected utility, leading to (2.2).
In these examples, the time-varying correlation between $X_{it}$ and $\beta_{it}$ highlights the significance and importance of the time-varying endogeneity through the random coefficients, which naturally arises due to agent optimization behavior. Numerous classic papers (Manski (1987), Altonji and Matzkin (2005), Graham and Powell (2012), Chernozhukov, Fernández-Val, Hahn, and Newey (2013)) have addressed such endogeneity by allowing for the correlation between $X_{it}$ and $A_i$. For instance, in production function estimation, a firm with a good management team is typically associated with larger input levels to leverage its managerial skill. This type of correlation has been addressed in the aforementioned papers. However, it carries important empirical implication to allow $X_{it}$ and $\beta_{it}$ to be correlated through factors beyond $A_i$. For instance, firms with a good management team may strategically select low levels of inputs for certain periods, which can not be explained only by the correlation between $X_{it}$ and $A_i$. To account for such behavior, it is useful to introduce an additional time-varying shock $\eta_{it}$ that contains important information about $\beta_{it}$ and affects the value of $X_{it}$.

Allowing $A_i$ to enter both the first-step equation (2.2) and the outcome equation (2.1) is another novel feature of this paper, distinguishing it from classic triangular simultaneous equations models (e.g., Newey, Powell, and Vella (1999), Imbens and Newey (2009)). In these models, the assumption is typically made on the first-step equation that determines the values of regressors that there is only one unobservable scalar-valued random variable $\eta_{it}$ that is correlated with both $A_i$ and $\varepsilon_{it}$, i.e., $X_{it} = g(Z_{it}, \eta_{it})$ instead of $X_{it} = g(Z_{it}, A_i, \eta_{it})$. Under regularity conditions, they propose to control for $\eta_{it}$ via either the regression residual from regressing $X_{it}$ on $Z_{it}$ (e.g., Newey, Powell, and Vella (1999)) or the conditional CDF $F_{X_{it}|Z_{it}}(X_{it}|Z_{it})$ which equals $F_{\eta_{it}}(\eta_{it})$ (e.g., Imbens and Newey (2009)). However, having $A_i$ appear in (2.2) renders the residual and CDF-based methods inapplicable because the one-to-one relationship between the control variable and $\eta_{it}$ no longer holds. Additionally, my model features a nonseparable first-step equation (2.2) in which $A_i$ enters $g(\cdot)$ potentially nonlinearly. The nonseparability and nonlinearity arise naturally from agent optimization behavior as illustrated in the empirical examples. The nonlinearity precludes the use of common demeaning or first-differencing techniques to eliminate $A_i$ (e.g., Laage (2024)). Instead, I rely on an index exclusion condition to deal with the fixed effects (e.g., Liu, Poirier, and Shiu (2024)).
3 Identification

3.1 Identification Idea

The main challenge to identify $E \beta_{it}$ and $E [\beta_{it} | X_{it}]$ comes from the correlation between $X_{it}$ and $\beta_{it}$ in model (2.1)–(2.2) through their dependence on the pairwise correlated random variables $A_i$, $\varepsilon_{it}$, and $\eta_{it}$. Suppose I follow the standard identification argument for linear models by taking the conditional expectation of both sides of (2.1) given $X_{it}$:

$$E [Y_{it} | X_{it}] = X_{it}' E [\beta_{it} | X_{it}].$$

Then, because $E [\beta_{it} | X_{it}]$ is a function of $X_{it}$, the textbook methods to identify $E \beta_{it}$ and $E [\beta_{it} | X_{it}]$ fail.

To see why, first I cannot identify $E [\beta_{it} | X_{it}]$ by multiplying both sides of (3.1) by $X_{it}$, taking outer expectation with respect to $X_{it}$, and finally inverting $E [X_{it} X_{it}']$:

$$E [X_{it} Y_{it}] = E \{X_{it} X_{it}' E [\beta_{it} | X_{it}]\},$$

since $X_{it} X_{it}'$ and $E [\beta_{it} | X_{it}]$, a function of $X_{it}$, are not separable inside the expectation on the right hand side (RHS) of (3.2).

Second, the perturbation-based identification method in linear models does not go through either. This method perturbs $X_{it}$ on both sides of (3.1):

$$\partial_E [Y_{it} | X_{it}] / \partial X_{it} = E [\beta_{it} | X_{it}] + X_{it}' \partial E [\beta_{it} | X_{it}] / \partial X_{it}. \quad (3.3)$$

Again, because $E [\beta_{it} | X_{it}]$ is a function of $X_{it}$, by the chain rule there is an additional term color-coded in red showing up on the RHS of (3.3), invalidating the perturbation-based method.

The key identification idea of this paper is to address the time-varying correlation between $X_{it}$ and $\beta_{it}$ using two control variables, $W_{it}$ and $V_{it}$. Conditional on $W_{it}$ and $V_{it}$, the residual variation in $X_{it}$ is determined solely by the exogenous $Z_{it}$. First, since the APE and LAR are generally not identified if the distribution of $A_i|X_i$ is left unrestricted (Liu, Poirier, and Shiu (2024)), I exploit an index exclusion condition, $A_i|X_{it}, Z_{it}, W_{it}) \sim_d A_i|W_{it}$, to control for $A_i$, where the existence and construction of $W_{it}$ are supported by the established results in the literature (Mundlak...
Then, for $\eta_{it}$ I propose a feasible conditional control variable $V_{it} = F_{X_{it}|Z_{it},W_i}(X_{it}|Z_{it},W_i)$ and prove that $V_{it} = F_{\eta_{it}|A_i,W_i}(\eta_{it}|A_i,W_i)$, the RHS of which is a one-to-one function of $\eta_{it}$ given $A_i$ and $W_i$. Since $A_i$ is unknown, I apply the LIE to $E[\beta_{it}|X_{it},V_{it},W_i]$ by further conditioning on the unknown $A_i$ in the inner conditional expectation to fix $\eta_{it}$ via $V_{it}$, such that conditional on $A_i$, $V_{it}$, and $W_i$, the residual variation in $X_{it}$ is entirely driven by the exogenous $Z_{it}$. This allows me to exclude $X_{it}$ from the inner conditional expectation. Next, I exclude $X_{it}$ from the conditioning set of the outer conditional expectation using the index exclusion condition of $W_i$ for $A_i$, i.e., $A_i|(X_{it},Z_{it},W_i) \sim d A_i|W_i$. Hence, I have

$$E[\beta_{it}|X_{it},V_{it},W_i] = E[E[\beta_{it}|X_{it},A_i,V_{it},W_i]|X_{it},V_{it},W_i] = E[\beta_{it}|V_{it},W_i], \quad (3.4)$$

which by (2.1) implies

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

Given sufficient variation in $X_{it}$ conditional on $V_{it}$ and $W_i$, I identify $E[\beta_{it}|V_{it},W_i]$ by either

$$E[\beta_{it}|V_{it},W_i] = E\left[X_{it}X_{it}'|V_{it},W_i\right]^{-1}E\left[X_{it}Y_{it}|V_{it},W_i\right] \quad (3.5)$$

or

$$E[\beta_{it}|V_{it},W_i] = \partial E\left[Y_{it}|X_{it},V_{it},W_i\right]/\partial X_{it}. \quad (3.6)$$

Consequently, I use the LIE to identify

$$E\beta_{it} = E[E[\beta_{it}|V_{it},W_i]] \quad \text{and} \quad E[\beta_{it}|X_{it}] = E[E[\beta_{it}|V_{it},W_i]|X_{it}]. \quad (3.7)$$

### 3.2 Assumptions and Main Identification Result

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 (Model).** Each coordinate of $g(Z_{it},A_i,\eta_{it})$ depends on one coordinate of $\eta_{it}$ strictly monotonically.

In the baseline model, I consider the case where each *endogenous* coordinate of $X_{it}$ depends on only one coordinate of $\eta_{it}$ strictly monotonically. When $d_{\eta} = 1$ and $g(\cdot)$
is additive in $\eta_{it}$ (e.g., Newey, Powell, and Vella (1999)), this assumption is satisfied. When $d_X = d_\eta = 1$ and $g(\cdot)$ is nonseparable in $\eta_{it}$, Assumption 1 is similar to the monotonicity condition imposed in equation (2) of Imbens and Newey (2009), except that I further have $A_i$ in (2.2). This is particularly relevant for empirical applications such as labor supply estimation where the focus in these models is usually on one endogenous regressor of wage. When $d_X > 1$ and $d_\eta = 1$ as in the literature of production function estimation (Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, Caves, and Frazer (2015)), Assumption 1 can be relaxed to require just one coordinate of endogenous $X_{it}$ to depend on $\eta_{it}$ strictly monotonically, which is also imposed in the aforementioned papers.

Assumption 1 requires justification when $d_X > 1$, $d_\eta > 1$, and each endogenous coordinate of $X_{it}$ admits a different coordinate of $\eta_{it}$. For example, it is possible that firms have different departments making hiring and investment decisions separately. Such decisions are based on different signals from the labor market and financial market. Therefore, it requires different coordinates of $\eta_{it}$ to enter the corresponding coordinate of $X_{it}$. Additionally, it is natural to allow firms to take advantage of positive shocks from labor market and financial market by hiring more employees and making more investments, leading to coordinate-wise strict monotonicity of $g(\cdot)$ in $\eta_{it}$. As a result, Assumption 1 is satisfied in this context.

It is worth highlighting that Assumption 1 does not restrict how different coordinates of $\eta_{it}$ are correlated. Thus, without loss of generality, I can redefine $\eta_{it}$ such that $d_\eta = d_X$ and each coordinate of $g(\cdot)$ depends on the corresponding coordinate of $\eta_{it}$. Moreover, Assumption 1 does not restrict how $A_i$ and $\eta_{it}$ are correlated, which could be important to empirical applications. Still, Assumption 1 restricts how the vector of $\eta_{it}$ enters the $g(\cdot)$ function that determines the value of $X_{it}$, which is a limitation of the paper. In Section 3.3, I discuss two other specifications to allow multiple coordinates of $\eta_{it}$ to enter one coordinate of $g(\cdot)$ when additional timing assumptions are considered plausible by practitioners and show how to adapt my method to these alternative specifications.

**Assumption 2 (Index Exclusion).** Let $W_i := W(X_i, Z_i)$, where $W : \mathbb{R}^{T \times (d_X + d_Z)} \mapsto \mathbb{R}^{d_W}$ is known. Suppose $A_i \mid (X_{it}, Z_{it}, W_i) \sim_d A_i \mid W_i$ for any fixed $t$.

Assumption 2 is an index exclusion restriction that requires the conditional distribution of $A_i \mid (X_{it}, Z_{it}, W_i)$ to depend only on $W_i$ which is a known function of $X_i$ and
It is similar to Assumption 2.1 of Altonji and Matzkin (2005) and Assumption A2 of Liu, Poirier, and Shiu (2024). A similar condition is also imposed by Wooldridge (2019). Liu, Poirier, and Shiu (2024) prove a non-identification result when the distribution of $A_i | X_i$ is left unrestricted in a nonlinear panel data model and suggest that restrictions on $A_i$ are generally required to identify the APE and LAR. Given $W_i$, the distribution of $A_i$ no longer depends on $X_{it}$ and $Z_{it}$ for any fixed $t$. In other words, $W_i$ is sufficient for $A_i$. Assumption 2 is used to (i) construct the conditional control variable for $\eta_{it}$, which I discuss in detail in relation to Assumption 3; and (ii) to control for $A_i$ when using the LIE argument to exclude $X_{it}$ from the conditioning set of $E[\beta_{it} | X_{it}, V_{it}, W_i]$ in (3.4), which I present in the proof of Theorem 1.

As pointed out by Liu, Poirier, and Shiu (2024), there are several ways to justify Assumption 2 from the literature. I follow their discussion to provide three justifications of Assumption 2. The first two methods are essentially parametric, resulting in a simpler expression for $W_i$, while the last one is nonparametric, leading to a more complex expression for $W_i$. This distinction highlights the trade-off between simplicity and generality. Moreover, I provide lower-level conditions for a nonparametric exchangeability condition and show that it is sufficient for Assumption 2 in Proposition 1.

First, Assumption 2 can be justified by generalizing equation (2.4) of Mundlak (1978) to be $A_i = h \left( T^{-1} \sum_{t=1}^{T} X_{it}, \nu_i \right)$, where $\nu_i \perp (X_{it}, Z_{it})$ for any $t$. Then, Assumption 2 is satisfied by taking $W_i = T^{-1} \sum_{t=1}^{T} X_{it}$. 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 $\nu_i$. Notice that there is no need to specify the functional form of $h(\cdot)$ nor the distribution of $\nu_i$.

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 the random vector $(X_{it}, Z_{it})' \in \mathbb{R}^2$ conditional on $A_i$ is a two-dimensional normal random vector i.i.d. through time with known covariance\(^7\) and $Z_i \perp A_i$, then $W_i = \left( \sum_{t=1}^{T} X_{it}, \sum_{t=1}^{T} Z_{it} \right)'$ satisfies Assumption 2. To see this, notice that the density of $(X_{it}, Z_{it}) | A_i$ depend on $A_i$ only through the mean of $X_{it}$. Hence, by the sufficient statistic for bi-variate normal random vectors with known covariance matrix (see

---

\(^7\)The analysis can be extended to cover the case when the covariance matrix is unknown. The required change is to expand $W_i$ to include the averages of the second-order polynomials of $X_{it}$ and $Z_{it}$ through time. See Chapter 7 of Hogg, McKean, and Craig (2019) for more details.

16
Chapter 7 of Hogg, McKean, and Craig (2019), \( W_i = \left( \sum_{t=1}^{T} X_{it}, \sum_{t=1}^{T} Z_{it} \right) \) suffices for Assumption 2. Furthermore, sufficient conditions for the analysis above are (i) \( \eta_t | A_i \) is a normal random vector i.i.d. through time with known variance, (ii) \( Z_t \perp (A_i, \eta_t) \), and (iii) \( g(\cdot) \) is linear in its arguments. Then, \( W_i = \left( \sum_{t=1}^{T} X_{it}, \sum_{t=1}^{T} Z_{it} \right) \) satisfies Assumption 2.8

Third, the nonparametric exchangeability condition (Assumption 2.3) from Altonji and Matzkin (2005) can also be adapted to justify Assumption 2. A similar idea is also used by Wooldridge (2019). Notice that the original exchangeability condition of Altonji and Matzkin (2005) is incompatible with the time-varying endogeneity through the random coefficients since it concerns the density of \( \eta_t \) given \( X_i \), and I present how to adapt it to the TERC model below. To grasp the idea, suppose (i) \( f_{\eta_t | A_i} = f_{\bar{\eta}_t | A_i} \) where \( \bar{\eta}_i \) is any permutation of the vector of \( \eta_i \) in time, (ii) \( Z_{it} \perp (\eta_{it}, A_i) \), and (iii) \( f_{A_i | \eta_i, Z_i} \) is continuous in \((X_i, Z_i)\). Then, combined with Assumption 1, it is straightforward to prove that \( f_{A_i | \eta_i, Z_i} = f_{A_i | \bar{X}_i, \bar{Z}_i} \) where \( \bar{X}_i \) and \( \bar{Z}_i \) are permutations of \( X_i \) and \( Z_i \) in time, respectively. Then, following Altonji and Matzkin (2005), I use the Weierstrass approximation theorem and the fundamental theorem of symmetric functions to obtain that the vector \( W_i \) of elementary symmetric functions of the paired elements of \( \{(X_{i1}, Z_{i1}), \ldots, (X_{iT}, Z_{iT})\} \) satisfies Assumption 2. I provide the details of the above analysis in the proof of Proposition 1. To see an example of \( W_i \) under the nonparametric exchangeability condition, suppose \( T = 2 \) and both \( X_{it} \) and \( Z_{it} \) are scalars, and define \( D_{it} := (X_{it}, Z_{it})' \). Then, \( W_i = \left( \sum_{t=1}^{T} D_{it}, \sum_{t=1}^{T} D_{it} D_{it}' \right) \) satisfies Assumption 2.10 Notice that parametric as-

\[ f_{(x_1, z_1, x_2, z_2) | A_i} (x_1, x_2; z_1, z_2) = f_{(h_1, h_2) | a} f_{z_1, z_2} (z_1, z_2) \]

Using the formula of density of normal distribution for \( f_{\eta_t | A} (h_t | a) \) and applying the Fisher-Neyman factorization theorem yield the desired result.

9To be precise, by \( f_{\eta_t | A_i} = f_{\bar{\eta}_t | A_i} \) I mean \( f_{\eta_t | A_i} (h_1, \ldots, h_{iT} | a_i) = f_{\bar{\eta}_t | A_i} (h_1, \ldots, h_{iT} | a_i) \) for any \( (h_1, \ldots, h_{iT}) \) belonging to the conditional support of \( \eta_i \) given \( A_i = a_i \) where \( (t_1, \ldots, t_T) \) is any permutation of \( (1, \ldots, T) \). The definition is similar for \( f_{A_i | \eta_i, Z_i} = f_{A_i | \bar{X}_i, \bar{Z}_i} \).

10In the concrete example discussed above, the maximum order of the polynomials in the construction of \( W_i \) depends on \( T \). In general, let \( \varphi_1 (u) = \sum_{t=1}^{T} u_t, \varphi_2 (u, v) = \sum_{t \neq \hat{t}} u_t v_{\hat{t}}, \ldots, \varphi_T (u, \ldots, w) = \sum_{t \neq \hat{t}, \ldots, \hat{k}} u_{t} v_{\hat{t}} \ldots w_{\hat{k}}, \) where \( u, v, \ldots, w \) are generic \( T \times 1 \) vectors. Then, I substitute each column vector of \( D_i = (X_i, Z_i) \) for each of the arguments \( u, v, \ldots, w \) (repetitions included) to
sumptions on the distribution of \( A_i \) are not needed for this result.

Once the fixed effect \( A_i \) is controlled for by \( W_i \), I still need to control for the time-varying \( \eta_{it} \) that enters the \( g ( \cdot ) \) function and is correlated with \( \beta_{it} \) via its correlation with both \( A_i \) and \( \varepsilon_{it} \). When \( A_i \) is not present in (2.2), I may follow the method of Imbens and Newey (2009) to use the conditional CDF \( V_{it}^{\text{im09}} := F_{X_{it}|Z_{it}} (X_{it} | Z_{it}) \) as a control for \( \eta_{it} \). This is true because Theorem 1 of Imbens and Newey (2009) proves that \( V_{it}^{\text{im09}} = F_{\eta_{it}} (\eta_{it}) \), the RHS of which is a one-to-one mapping of \( \eta_{it} \) when \( d_X = d_\eta = 1 \) provided (i) \( Z_{it} \perp (\varepsilon_{it}, \eta_{it}) \), (ii) \( g ( \cdot ) \) is strictly monotonic in \( \eta_{it} \), and (iii) \( F_{\eta_{it}} (\cdot) \) is a strictly increasing function. Therefore, conditional on \( V_{it}^{\text{im09}} \), \( X_{it} \) and \( \varepsilon_{it} \) are independent. The argument above can be generalized to the setting where \( d_X > 1 \) by letting

\[
V_{it}^{\text{im09}} = \left( F_{X_{it}^{(1)}|Z_{it}} (X_{it}^{(1)} | Z_{it}) , \ldots , F_{X_{it}^{(d_X)}|Z_{it}} (X_{it}^{(d_X)} | Z_{it}) \right) \right)^{'}
\]

\[
= \left( F_{\eta_{it}^{(1)}} (\eta_{it}^{(1)}) , \ldots , F_{\eta_{it}^{(d_X)}} (\eta_{it}^{(d_X)}) \right) \right)^{'}
\]

to control for the whole vector of \( \eta_{it} \) coordinate by coordinate.

The challenge in applying their method to the TERC model arises from the existence of \( A_i \) in (2.2), which is justified in many applications as agents usually know their \( A_i \) (e.g., via repeated learning) when choosing the value of \( X_{it} \). When \( A_i \) is present in (2.2) and \( d_X = 1 \), one may want to use \( F_{X_{it}|Z_{it},A_i} (X_{it} | Z_{it}, A_i) \) as the control variable for \( \eta_{it} \), following a similar argument as in Imbens and Newey (2009). However, this approach is infeasible because \( A_i \) is unknown. The idea of this paper is to replace \( A_i \) with \( W_i \) in the construction of \( V_{it} = F_{X_{it}|Z_{it},W_i} (X_{it} | Z_{it}, W_i) \) and show that it controls for \( \eta_{it} \) under a conditional independence assumption on \( Z_{it} \) and a monotonicity condition on the conditional CDF \( F_{\eta_{it}|A_i,W_i} (\eta_{it} | A_i, W_i) \) in \( \eta_{it} \). I summarize them in the next assumption.

**Assumption 3 (Control for \( \eta_{it} \)).** Suppose the following conditions are satisfied:

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

construct \( W_i \). Such a \( W_i \) satisfies Assumption 2. See Chapter II.3 of Weyl (1939) for details of the construction and its proof.
(b) \( F^{(l)}_{\eta_{it} \mid A_i, W_i} \left( \eta_{it}^{(l)} \mid A_i, W_i \right) \) is strictly increasing in \( \eta_{it}^{(l)} \) for every realization of \((A_i, W_i)\) on its support and each \( l = 1, \ldots, d_X \).

Assumption 3(a) concerns the vector of exogenous IVs \( Z_{it} \). Since \( W_i \) can be viewed as summarizing all the time-invariant information about \( A_i \) in the data, this assumption, loosely speaking, requires \( Z_{it} \perp (\varepsilon_{it}, \eta_{it}) \mid A_i \), which is already implied by the standard exogeneity assumption of \( Z_{it} \perp (A_i, \varepsilon_{it}, \eta_{it}) \). For instance, when \( d_X = 1 \) and suppose one follows Mundlak (1978) to assume \( A_i = \left( \sum_{t=1}^{T} X_{it} \right) \alpha + \nu_i = W_i \alpha + \nu_i \), where \( \nu_i \) is independent of everything else and \( \alpha \neq 0 \). Then, if \( Z_{it} \perp (A_i, \varepsilon_{it}, \eta_{it}, \nu_i) \), I have \( Z_{it} \perp (\eta_{it}, \varepsilon_{it}) \mid (A_i, W_i) \) by the law of conditional independence because \( W_i \) is a function of \( A_i \) and \( \nu_i \).\(^{11}\) Assumption 3(a) is used to ensure that \( Z_{it} \) can be excluded from the conditioning set of \( F^{(l)}_{\eta_{it} \mid Z_{it}, A_i, W_i} \left( \eta_{it}^{(l)} \mid Z_{it}, A_i, W_i \right) \) for each \( l = 1, \ldots, d_X \).

Assumption 3(b) is similar to condition (ii) of Theorem 1 in Imbens and Newey (2009). It is mild since it concerns a conditional CDF, which is typically strictly increasing for continuous random variables. Assumption 3(b) guarantees that fixing \((A_i, V_{it}, W_i)\) uniquely determines the value of \( \eta_{it} \).

Assumption 3 allows me to construct 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 \( V_{it} = F_{X_{it} \mid Z_{it}, W_i} \left( X_{it} \mid Z_{it}, W_i \right) \). Then,

\[
V_{it} = F_{X_{it} \mid Z_{it}, A_i, W_i} \left( X_{it} \mid Z_{it}, A_i, W_i \right) = F_{\eta_{it} \mid Z_{it}, A_i, W_i} \left( \eta_{it} \mid Z_{it}, A_i, W_i \right)
\]

where the first equality holds by Assumption 2 which allows me to add \( A_i \) to the conditioning set, 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). Notice that \( V_{it} \) is a one-to-one mapping of \( \eta_{it} \) only when \( A_i \) is also conditioned upon. In the proof of Theorem 1,

\(^{11}\)I provide another parametric example where \( Z_{it} \perp (A_i, \varepsilon_{it}, \eta_{it}) \) implies \( Z_{it} \perp (\varepsilon_{it}, \eta_{it}) \mid (A_i, W_i) \). For illustration purpose, suppress subscript \( i \) and suppose \( T = 2, d_X = d_Z = 1 \), and \( W = (X_1 + X_2, Z_1 + Z_2) \) satisfies Assumption 2 (Liu, Poirier, and Shiu (2024)). Furthermore, suppose \( X_t = Z_t \gamma + A + \eta_t \) for \( t = 1, 2 \), and \((Z_1, Z_2, \varepsilon_1, \varepsilon_2, \eta_1, \eta_2) \) \( \sim N(\mu, \Sigma) \) where \( \Sigma = \begin{bmatrix} \Sigma_1 & 0_{2 \times 4} \\ 0_{4 \times 2} & \Sigma_2 \end{bmatrix} \) and \( \Sigma_1 \) and \( \Sigma_2 \) are arbitrary positive definite \( 2 \times 2 \) and \( 4 \times 4 \) matrices, respectively. Then, formula for the conditional density function of normally distributed random vectors shows that \((Z_t, \varepsilon_t, \eta_t) \mid (A, W) \) is also normally distributed with diagonal covariance matrix for \( t = 1, 2 \), satisfying Assumption 3(a).
I use the LIE on $\mathbb{E} \left[ \beta_{it} | X_{it}, V_{it}, W_{it} \right]$ where I first condition on $A_i$ to control for $\eta_{it}$ via $V_{it}$ in the inner expectation step. Then, I deal with the unknown $A_i$ with $W_{i}$ in the outer expectation step to prove (3.4). Finally, I use the residual variation in $X_{it}$, guaranteed by the next assumption, to identify the APE and LAR via (3.5) or (3.6).

The last assumption concerns the residual variation in $X_{it}$ given $V_{it}$ and $W_{i}$.

**Assumption 4 (Residual Variation in $X_{it}$).** There are at least $d_X$ linearly independent points in the support of $X_{it}$ a.s. given $V_{it}$ and $W_{it}$.

Assumption 4 guarantees that $\mathbb{E} \left[ X_{it}X_{it}' | V_{it}, W_{it} \right]$ in (3.5) is invertible\(^{12}\), thus the APE and LAR can be identified by (3.5) and the LIE. As discussed by Imbens and Newey (2009), requiring residual variation in $X_{it}$ conditioning on $V_{it}$ is generally not restrictive. Essentially, it requires $Z_{it}$ to move $X_{it}$ sufficiently well such that the set $\{ X_{it} : V_{it} = v \text{ for } v \in \text{Supp} (V_{it}) \}$ is not a singleton. Therefore, it is similar to the relevance condition in the classical IV literature.\(^{13}\) The restriction on the support of $X_{it}$ mainly comes from conditioning on $W_{i}$. When parametric justifications for Assumption 2 are used (e.g., Mundlak 1978, Arkhangelsky and Imbens 2022, Liu, Poirier, and Shiu 2024) so that $W_{i}$ only includes averages of $X_{it}$ and/or $Z_{it}$ through time, Assumption 4 is not restrictive. However, when nonparametric justifications for Assumption 2 (e.g., the exchangeability condition of Altonji and Matzkin 2005) are involved, Assumption 4 can be restrictive and generally requires that $T$ is larger than $d_X$.\(^{14}\) It is worth pointing out that I only need to use the *endogenous* coordinates of $X_{it}$ to construct $V_{it}$ and $W_{i}$. Thus, when some coordinates of $X_{it}$ are *exogenous* in the sense that they are independent of $(A_i, \eta_{it}, \varepsilon_{it})$, the *unconditional* variation in the exogenous coordinates of $X_{it}$ can be exploited to satisfy Assumption 4. See footnote 16 below for more details on this point.

\(^{12}\)This is true because for any random vector $X \in \mathbb{R}^{d_X}$, $\mathbb{E} \left[ XX' \right]$ is singular if and only if there exists a non-random vector $a \in \mathbb{R}^{d_X} \setminus \{0\}$ such that $a'X = 0$ almost surely.

\(^{13}\)I thank an anonymous referee for pointing this out.

\(^{14}\)One way to justify Assumption 4 when $W_{i}$ includes averages through time of the polynomials of $(X_{it}, Z_{it})$ up to the $T^{th}$ order (Altonji and Matzkin 2005) is to use the symmetry through time in the solution to $\{ X_i : V_{it} = v, W_{i} = w \}$. To see a concrete example, suppose $T = 3$, $X_{it} \in \mathbb{R}^{2}$, $Z_{it} \in \mathbb{R}^{2}$, and $D_{it} := \left( X_{it}', Z_{it}' \right)' \in \mathbb{R}^{4}$. If $D_{i} = (D_{1i}, D_{2i}, D_{3i}) = (a, b, c)$ satisfies $V_{it} = v$ for $t = 1, 2$ and $W_{i} = w$ where all $a, b, c \in \mathbb{R}^{4}$ and the first two coordinates of $a$ and $b$ are linearly independent, by the symmetry of $W_{i}$ in $(X_{it}', Z_{it}')$ through $t$, $(b, a, c)$ must also satisfy $V_{it} = v$ and $W_{i} = w$ for $t = 1, 2$. Therefore, given $V_{i1} = v$ and $W_{i} = w$ there are two linearly independent vectors of $(a^{(1)}, a^{(2)})$ and $(b^{(1)}, b^{(2)})$ that $X_{i1}$ can take, satisfying Assumption 4.
Next, I present another assumption on the residual variation of $X_{it}$ given $V_{it}$ and $W_i$ to facilitate the derivative-based argument (3.6).

**Assumption 4’ (More Variation in $X_{it}$).** The support of $X_{it}$ given $V_{it}$ and $W_i$ contains some ball of positive radius a.s. with respect to $(V_{it}, W_i)$.

Assumption 4’ is similar to Assumption 2.2 of Altonji and Matzkin (2005) and is stronger than Assumption 4. When parametric justifications (e.g., Mundlak (1978), Arkhangelsky and Imbens (2022), Liu, Poirier, and Shiu (2024)) are used for Assumption 2 such that $W_i$ contains only mean of $X_{it}$ and/or $Z_{it}$ through time, Assumption 4’ is generally not restrictive. However, when nonparametric justifications (e.g., Altonji and Matzkin (2005)) are used for Assumption 2, Assumption 4’ typically requires ex-ante information about $W_i$ (e.g., the existence of exogenous regressors in (2.1)). This is true because in this case $W_i$ includes averages of each endogenous coordinate of $X_{it}$, $Z_{it}$, and their interaction terms up to the $T^{th}$ order through time, conditioning on which there may not be enough residual variation in $X_{it}$. To address this issue, when nonparametric justifications for Assumption 2 are used, it is advisable to follow the recommendations by Altonji and Matzkin (2005) to ensure continuous residual variation in $X_{it}$ given $V_{it}$ and $W_i$.\(^{15}\)^\(^{16}\)

---

\(^{15}\)Specifically, Altonji and Matzkin (2005) suggest to (i) leverage unconditional variation in the exogenous regressors that do not enter $W_i$, which I provide details in footnote 16; (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 $E[Y_{it}|X_{it}, V_{it}, W_i]$ with $W_i$ replaced by a proper subset of it; (iii) assume $A_i|W_i$ to depend on $W_i$ only through a linear combination of the elements of $W_i$, i.e., $f_{A_i|W_i} = f_{A_i|\sum_{k=1}^{dW} c_k W_i^{(k)}}$ where $W_i = \left(W_i^{(1)}, ..., W_i^{(dW)}\right)$ and the $c_k$’s are unknown parameters, which implies that $E[Y_{it}|X_{it}, V_{it}, W_i] = E[Y_{it}|X_{it}, V_{it}, \sum_{k=1}^{dW} c_k W_i^{(k)}]$ and the estimation of $c_k$’s can proceed by using the method of Ichimura and Lee (1991); (iv) impose a priori restrictions on model (2.1)-(2.2) such that $W_i$ enters $E[Y_{it}|X_{it}, V_{it}, W_i]$ in a parametric way (e.g., $E[\beta_{it}|V_{it}, W_i] = \sum_{k=1}^{dW} c_k W_i^{(k)} + \sum_{l=1}^{dV} d_l V_l^{(l)}$ when $d_X = 1$); and, (v) directly impose restrictions on $E[Y_{it}|X_{it}, V_{it}, W_i]$.

\(^{16}\)To see a concrete example how the unconditional variations in the exogenous regressors help to satisfy Assumption 4’, suppose $X_{it}^{(1)}$ is endogenous while $X_{it}^{(-1)}$, all other coordinates of $X_{it}$ except $X_{it}^{(1)}$, are exogenous (independent of $X_{it}^{(1)}$ and $\beta_{it}$). 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 some small ball of positive radius. Denote $b_1(V_{it}, W_i) = E[\beta_{it}|V_{it}, W_i]$. 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

$$E\left[Y_{it}|X_{it}^{(1)}, X_{it}^{(-1)}, V_{it}, W_i\right] = X_{it}^{(1)} b_1^{(1)}(V_{it}, W_i) + X_{it}^{(-1)} b_1^{(-1)}(V_{it}, W_i)$$
Assumption 4' allows the straightforward perturbation-based method (3.6) to identify the APE and LAR without requiring the computation of the inverse of $E \left[ X_{it} X_{it}' \mid V_{it}, W_i \right]$. 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. Accordingly, I impose it for the estimation and inference results in the next section.

I now state my main identification theorem.

**Theorem 1 (Identification).** If Assumptions 1–3 and either Assumption 4 or 4' are satisfied, then $E \beta_{it}$ and $E [\beta_{it} \mid X_{it}]$ are identified.

### 3.3 Extensions

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

In my baseline model, I assume that each coordinate of $X_{it}$ depends on only one coordinate of $\eta_{it}$ (as well as $Z_{it}$ and $A_i$). Here, I discuss two ways to include vector-valued $\eta_{it}$ into coordinates of $X_{it}$ under additional assumptions motivated by empirical research. To clearly present the idea, I consider the case where $d_X = d_\eta = 2$.

First, timing assumptions on the choice of regressors can be leveraged to allow vector-valued $\eta_{it}$ to enter certain coordinates of $X_{it}$. For instance, in production function applications, choices of later inputs such as capital typically do not depend on random shocks to earlier chosen inputs such as labor and/or material after conditioning on the level of the earlier chose inputs.\footnote{Appendix 1 of the 2006 working paper version of Ackerberg, Caves, and Frazer (2015) discusses a similar idea. I thank an anonymous referee for this suggestion.} If this timing assumption is imposed, I modify model (2.1)–(2.2) to be

$$
Y_{it} = X_{it}' \beta (A_i, \varepsilon_{it}), \\
X_{it}^{(1)} = g^{(1)} \left( Z_{it}, A_i, \eta_{it}^{(1)} \right), \text{ and } X_{it}^{(2)} = g^{(2)} \left( X_{it}^{(1)}, Z_{it}, A_i, \eta_{it}^{(2)} \right).
$$

**17**

To identify $b_{1}^{(-1)} (V_{it}, W_i)$. Finally, identification of $b_{1}^{(1)} (V_{it}, W_i)$ can be obtained as

$$
b_{1}^{(1)} (V_{it}, W_i) = \left( E \left[ Y_{it} \mid X_{it}^{(1)}, X_{it}^{(-1)}, V_{it}, W_i \right] - X_{it}^{(-1)} b_{1}^{(-1)} (V_{it}, W_i) \right) / X_{it}^{(1)}.
$$

To identify $b_{1}^{(-1)} (V_{it}, W_i)$. Finally, identification of $b_{1}^{(1)} (V_{it}, W_i)$ can be obtained as

$$
b_{1}^{(1)} (V_{it}, W_i) = \left( E \left[ Y_{it} \mid X_{it}^{(1)}, X_{it}^{(-1)}, V_{it}, W_i \right] - X_{it}^{(-1)} b_{1}^{(-1)} (V_{it}, W_i) \right) / X_{it}^{(1)}.
$$

\[17\] Appendix 1 of the 2006 working paper version of Ackerberg, Caves, and Frazer (2015) discusses a similar idea. I thank an anonymous referee for this suggestion.
Note that now $X_{it}^{(2)}$ depends on the whole vector of $\eta_{it}$ if $X_{it}^{(1)}$ is substituted into $g^{(2)}(\cdot)$. I propose to use $V_{it}^{(2)} := F_{X_{it}^{(2)}}(Z_{it}, W_{i}) (X_{it}^{(2)} | X_{it}^{(1)}, Z_{it}, W_{i})$ to control for $\eta_{it}^{(2)}$ first. Then, I use $F_{X_{it}^{(1)}}(Z_{it}, W_{i}) (X_{it}^{(1)} | Z_{it}, W_{i})$ to control for $\eta_{it}^{(1)}$. The rest of the identification analysis follows as in this paper.\(^{18}\)

Second, the interaction between different coordinates of $X_{it}$ also provides useful information about certain coordinates of $\eta_{it}$. One way to exploit such information is to assume that while both coordinates of $\eta_{it}$ enter functions that determine each coordinate of $X_{it}$, the ratio of $X_{it}^{(1)}/X_{it}^{(2)}$ may depend on only one coordinate of $\eta_{it}$, e.g., $\eta_{it}^{(2)}$ (and $Z_{it}$ and $A_{i}$). For example, in production function applications, the interpretation of $\eta_{it}$ could be that $\eta_{it}^{(1)}$ is the Hicks neutral productivity and $\eta_{it}^{(2)}$ is the labor-augmenting technology. Although both material and labor choices depend on both $\eta_{it}^{(1)}$ and $\eta_{it}^{(2)}$, it is plausible that *material per worker* depends only on the labor-augmenting technology $\eta_{it}^{(2)}$. See Proposition 2.1 of Demirer (2022) for more discussions about this assumption. Given this assumption, I can proceed to control for $\eta_{it}^{(2)}$ first by $V_{it}^{(2)} := F_{X_{it}^{(2)}}(Z_{it}, W_{i}) (X_{it}^{(1)}/X_{it}^{(2)} | Z_{it}, W_{i})$, and then for $\eta_{it}^{(1)}$ by $V_{it}^{(1)} := F_{X_{it}^{(1)}}(Z_{it}, W_{i}) (X_{it}^{(1)} | Z_{it}, V_{it}^{(2)}, W_{i})$. The rest of the identification analysis follows as in this paper.\(^{19}\)

---

\(^{18}\) Specifically, for this method to work, I maintain Assumptions 2, 3(a), and 4 (or 4') and change Assumption 1 to be that $g^{(1)}(Z_{it}, A_{i}, \eta_{it}^{(1)})$ and $g^{(2)}(X_{it}^{(1)}, Z_{it}, A_{i}, \eta_{it}^{(2)})$ are strictly monotonic in $\eta_{it}^{(1)}$ and $\eta_{it}^{(2)}$, respectively. I also modify Assumption 3(b) to be that $F_{\eta_{it}^{(1)}}(A_{i}, W_{i}) (\eta_{it}^{(1)} | A_{i}, W_{i})$ and $F_{\eta_{it}^{(2)}}(X_{it}^{(1)}, A_{i}, W_{i}) (\eta_{it}^{(2)} | X_{it}^{(1)}, A_{i}, W_{i})$ are strictly increasing in $\eta_{it}^{(1)}$ and $\eta_{it}^{(2)}$, respectively.

\(^{19}\) Specifically, suppose $X_{it}^{(1)} = g^{(1)}(Z_{it}, A_{i}, \eta_{it}^{(1)}, \eta_{it}^{(2)})$ and $X_{it}^{(1)}/X_{it}^{(2)} = g^{(2)}(Z_{it}, A_{i}, \eta_{it}^{(2)})$. I maintain Assumptions 2, 3(a), and 4. I modify Assumption 1 to be that $g^{(1)}(Z_{it}, A_{i}, \eta_{it}^{(1)}, \eta_{it}^{(2)})$ and $g^{(2)}(Z_{it}, A_{i}, \eta_{it}^{(2)})$ are strictly monotonic in $\eta_{it}^{(1)}$ and $\eta_{it}^{(2)}$, respectively. I further change Assumption 3(b) to be $F_{\eta_{it}^{(1)}}(\eta_{it}^{(1)} | A_{i}, W_{i}) (\eta_{it}^{(1)} | A_{i}, W_{i})$ and $F_{\eta_{it}^{(2)}}(\eta_{it}^{(2)} | A_{i}, W_{i}) (\eta_{it}^{(2)} | A_{i}, W_{i})$ are strictly increasing in $\eta_{it}^{(1)}$ and $\eta_{it}^{(2)}$, respectively. Then, I have

$$V_{it}^{(2)} := F_{X_{it}^{(1)}/X_{it}^{(2)}}(Z_{it}, W_{i}) (X_{it}^{(1)}/X_{it}^{(2)} | Z_{it}, W_{i})$$

$$= F_{X_{it}^{(1)}/X_{it}^{(2)}}(Z_{it}, A_{i}, W_{i}) (X_{it}^{(1)}/X_{it}^{(2)} | Z_{it}, A_{i}, W_{i}) = F_{\eta_{it}^{(2)}}(A_{i}, W_{i}) (\eta_{it}^{(2)} | A_{i}, W_{i})$$.
Higher-Order Moments of $\beta_{it}$

For clarity of exposition, suppose the regressors on the RHS of (2.1) include constant 1 and a scalar-valued $X_{it}$. It is straightforward to generalize the idea to cover $d_X$-dimensional $X_{it}$ for $d_X \geq 2$. With a slight abuse of notation, let $(\beta_{it}, \omega_{it}) \in \mathbb{R}^2$ where $\beta_{it}$ is the random coefficient for $X_{it}$ and $\omega_{it}$ is for constant 1. Since the residual variation in $X_{it}$ given $V_{it}$ and $W_i$ is driven only by exogenous $Z_{it}$, (3.4) holds for any measurable function of $\beta_{it}$. Thus, I have

$$E\left[ \beta_{it}^2 | X_{it}, V_{it}, W_i \right] = E \left[ \beta_{it}^2 | V_{it}, W_i \right],$$

and similarly for $E \left[ \omega_{it}^2 | X_{it}, V_{it}, W_i \right]$ and $E \left[ \omega_{it} \beta_{it} | X_{it}, V_{it}, W_i \right]$. Taking the conditional expectation of both sides of (2.1) squared given $(X_{it}, V_{it}, W_i)$ yields

$$E \left[ Y_{it}^2 | X_{it}, V_{it}, W_i \right] = X_{it}^2 E \left[ \beta_{it}^2 | V_{it}, W_i \right] + 2 X_{it} E \left[ \beta_{it} \omega_{it} | V_{it}, W_i \right] + E \left[ \omega_{it}^2 | V_{it}, W_i \right].$$

(3.8)

Then, I have

$$E \left[ \beta_{it}^2 | V_{it}, W_i \right] = \left( \partial^2 E \left[ Y_{it}^2 | X_{it}, V_{it}, W_i \right] / \partial X_{it}^2 \right) / 2,$$

(3.9)

$$E \left[ \beta_{it} \omega_{it} | V_{it}, W_i \right] = \left( \partial E \left[ Y_{it}^2 | X_{it}, V_{it}, W_i \right] / \partial X_{it} - 2 X_{it} E \left[ \beta_{it}^2 | V_{it}, W_i \right] \right) / 2,$$

and

$$E \left[ \omega_{it}^2 | V_{it}, W_i \right] = E \left[ Y_{it}^2 | X_{it}, V_{it}, W_i \right] - X_{it}^2 E \left[ \beta_{it}^2 | V_{it}, W_i \right] - 2 X_{it} E \left[ \beta_{it} \omega_{it} | V_{it}, W_i \right],$$

which identify the conditional (given $X_{it}$) and unconditional second-order moments of $\beta_{it}$ and $\omega_{it}$ by the LIE.

The above technique can also be used to identify the intertemporal correlations of the random coefficients. For example, $E \left[ \beta_{it} \beta_{is} | X_{it}, X_{is}, V_{it}, V_{is}, W_i \right]$ can be identified from $E \left[ Y_{it} Y_{is} | X_{it}, X_{is}, V_{it}, V_{is}, W_i \right]$ for any $t \neq s$ following an argument similar to

and

$$V_{it}^{(1)} := F_{X_{it}^{(2)} | Z_{it}, V_{it}^{(2)}, W_i} \left( X_{it}^{(1)} | Z_{it}, V_{it}^{(2)}, W_i \right)$$

$$= F_{X_{it}^{(1)} | Z_{it}, V_{it}^{(2)}, A_i, W_i} \left( X_{it}^{(1)} | Z_{it}, V_{it}^{(2)}, A_i, W_i \right) = F_{\eta_{it}^{(1)} | \eta_{it}^{(2)}, A_i, W_i} \left( \eta_{it}^{(1)} | \eta_{it}^{(2)}, A_i, W_i \right),$$

where the second equality holds because by Assumption (2), $A_i \perp (X_{it}, Z_{it}) | W_i$ and that $V_{it}^{(2)}$ is a function of $(X_{it}, Z_{it}, W_i)$ and the last equality holds because $(V_{it}^{(1)}, A_i, W_i)$ uniquely determines $\eta_{it}^{(2)}$. Therefore, fixing $(X_{it}, A_i, V_{it}, W_i)$ is equivalent to fixing $(X_{it}, A_i, \eta_{it}, W_i)$. The rest of the argument follows as in the proof of Theorem 1.
(3.8)–(3.9).

Exogenous Shocks and Covariates

The identification argument of this paper goes through when I further include ex-post shocks \( v_{it} \) to \( \beta_{it} \) (i.e., \( \beta_{it} = \beta (A_i, \varepsilon_{it}, v_{it}) \)) and \( \epsilon_{it} \) to \( Y_{it} \) (i.e., \( Y_{it} = X_{it}' \beta_{it} + \epsilon_{it} \)), where both \( v_{it} \) and \( \epsilon_{it} \) are independent of all other variables with \( \mathbb{E} \epsilon_{it} = 0 \). Note that due to the inclusion of \( v_{it} \), \( \mathbb{E}[\beta_{it} | X_{it}] \) becomes a possibly time-varying function of \( X_{it} \) as the distribution of \( v_{it} \) may depend on time \( t \). It does not affect my identification analysis because \( v_{it} \) is independent of all other variables. I present the changes required to the proof due to the inclusion of \( v_{it} \) into \( \beta_{it} \) at the end of the proof of Theorem 1. I follow the more general specification \( \beta_{it} = \beta (A_i, \varepsilon_{it}, v_{it}) \) in the estimation and examine its impact via simulations.

As for \( \epsilon_{it} \), since it is independent of all other variables, has mean zero, and enters (2.1) in an additive way, equations (3.4)–(3.7) hold as before. Thus, the identification of \( \mathbb{E} \beta_{it} \) and \( \mathbb{E} [\beta_{it} | X_{it}] \) is not affected. However, it has an impact on the identification of higher-order moments of \( \beta_{it} \). For example, in (3.9) \( \mathbb{E} [\omega_{it}^2 | V_{it}, W_i] \) is not identified if \( \epsilon_{it} \) is included with \( \mathbb{E} \epsilon_{it}^2 > 0 \). One way to handle this issue is by following Arellano and Bonhomme (2012) to impose and exploit additional assumptions on how \( \epsilon_{it} \) evolves over time (e.g., \( \epsilon_{it} \) follows an ARMA(p,q) process).

Finally, when exogenous covariates \( Z_{1,it} \) are available and included in \( X_{it} \), my identification analysis actually benefits from it because Assumptions 1–3 concern only endogenous covariates \( U_{it} \) of \( X_{it} \). Thus, I replace \( X_{it} \) by \( U_{it} \) in Assumptions 1–3 and in the construction of \( V_{it} \). Furthermore, as discussed in footnote 16, the unconditional variation of \( Z_{1,it} \) makes Assumption 4 and 4′ easier to be satisfied. The details on how to adapt the current analysis to include \( Z_{1,it} \) are provided at the end of the proof of Theorem 1.

4 Estimation and Large Sample Theory

In this section, I show how to estimate \( \mathbb{E} \beta_{it} \) and \( \mathbb{E} [\beta_{it} | X_{it}] \) via three-step series estimators based on the constructive identification argument. Next, I establish the
convergence rates of the proposed estimators. Finally, I prove that the estimators are asymptotically normal and provide consistent estimators for their asymptotic covariance matrices.

4.1 Estimation

As discussed in Extension 3 of Section 3, I include the ex-post shock \( v_{it} \) in \( \beta_{it} \), resulting in \( \mathbb{E} [ \beta_{it} | X_{it}, V_{it}, W_i ] =: b_{it} (V_{it}, W_i) \) and \( \mathbb{E} [ \beta_{it} | X_{it} ] =: b_t (X_{it}) \). The parameters of interest are

\[
\bar{b} := \mathbb{E} \beta_{it} \quad \text{and} \quad b_t (x) := \mathbb{E} [ \beta_{it} | X_{it} = x ]. \tag{4.1}
\]

\( \bar{b} \) is the APE over the unconditional distribution of \((A_i, \varepsilon_{it}, \nu_{it})\). \( b_t (x) \) is the LAR function for a subpopulation characterized by \( X_{it} = x \) in period \( t \). Both objects are useful for answering policy-related questions. For example, plugging realizations \( x_{it} \) of \( X_{it} \) into \( b_t (x) \) provides a fine approximation to \( \beta_{it} \) for agent \( i \) in time \( t \). Therefore, distributional properties of \( \beta_{it} \) may be inferred from the distribution of \( b_t (x_{it}) \).

I propose to estimate the parameters in (4.1) with three-step series estimators. For clarity of exposition, I let \( d_X = 1 \), and highlight the changes required for \( d_X > 1 \) when necessary. To fix ideas, I follow Liu, Poirier, and Shiu (2024) 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 \( V_t (x, z, w) := F_{X_{it}|Z_{it},W_i} (x | z, w) \) by regressing \( 1 \{ X_{it} \leq x \} \) on the basis functions \( q^{M1} (\cdot) \) of \((Z_{it}, W_i)\) with trimming function \( \tau (\cdot) \):

\[
\hat{V}_t (x, z, w) = \tau \left( \hat{F}_{X_{it}|Z_{it},W_i} (x | z, w) \right) = \tau \left( q^{M1} (z, w) \left( n^{-1} \hat{Q}_t^{-1} \sum_{j=1}^n q_{jt} 1 \{ x_{jt} \leq x \} \right) \right)
\]

\[
= \tau \left( q^{M1} (z, w) \hat{\gamma}_{M1}^t (x) \right), \tag{4.2}
\]

where \( \hat{Q}_t := n^{-1} \sum_{i=1}^n q_{it} q_{it}' \) and \( q_{it} := q^{M1} (z_{it}, w_i) \). Examples of \( q^{M1} (\cdot) \) include power series and spline functions. When \( d_X > 1 \), I can regress \( 1 \{ X_{it}^{(l)} \leq x^{(l)} \} \) on the basis functions \( q^{M1} (\cdot) \) of \((Z_{it}, W_i)\) with trimming function \( \tau (\cdot) \) for each \( l \) and obtain \( \hat{V}_t (x, z, w) = \left( \hat{V}_t^{(1)} (x, z, w), ..., \hat{V}_t^{(d_X)} (x, z, w) \right) \). I highlight two properties of \( \hat{V}_t (x, z, w) \). First, the regression coefficient \( \hat{\gamma}_{M1}^t (x) \) in (4.2) depends on \( x \) because

\footnote{\( M_1 \) (also \( M_2, m_2, \) and \( M_3 \) in this section) is a function of \( n \) that goes to infinity slowly as \( n \to \infty \).}
the dependent variable \( \mathbb{1} \{ X_{it} \leq x \} \) is a function of \( x \). This fact causes the convergence rate of \( \hat{V}_t (x, z, w) \) 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 \( \tau (x) = 1 \{ x > 0 \} \cdot \min (x, 1) \).

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

\[
\hat{G}_t (s) = p^{M_2} (s)^\prime n^{-1} \hat{P}_t^{-1} \hat{P}_t y_t =: p^{M_2} (s)^\prime \hat{\alpha}_t^{M_2},
\]

where \( \hat{P}_t := n^{-1} \sum_{i=1}^n \hat{p}_{it}\hat{p}_{it}^\prime, \ 
\hat{p}_{it} := p^{M_2} (x_{it}, \hat{v}_{it}, w_i), \ 
\hat{P}_t := (\hat{p}_{it}, \ldots, \hat{p}_{nt})^\prime, \) and \( y_t := (y_{it}, \ldots, y_{nt})^\prime \). Following Newey, Powell, and Vella (1999), I let \( p^{M_2} (s) = x \otimes p^{m_2} (v, w) \) by exploiting the index structure of (2.1), which enables a faster convergence rate of \( \hat{G}_t (s) \) to \( G_t (s) \).

Finally, I exploit the index structure of (2.1) again to estimate \( b_{it} (v, w) := \mathbb{E} [ \beta_{it} | V_{it} = v, W_i = w ] \). When (3.5) is used, I estimate \( b_{it} (v, w) \) by

\[
\hat{b}_{it} (v, w) = \left( \hat{\mathbb{E}} \left[ X_{it} X_{it}' | \hat{V}_{it} = v, W_i = w \right] \right)^{-1} \hat{\mathbb{E}} \left[ X_{it} Y_{it} | \hat{V}_{it} = v, W_i = w \right],
\]

where \( \hat{\mathbb{E}} \left[ X_{it} Y_{it} | \hat{V}_{it}, W_i \right] \) is obtained by regressing each coordinate of \( X_{it} Y_{it} \) on the basis functions of \( (\hat{V}_{it}, W_i) \) and similarly for \( \hat{\mathbb{E}} \left[ X_{it} X_{it}' | \hat{V}_{it}, W_i \right] \). When (3.6) is used, I estimate \( b_{it} (v, w) \) by

\[
\hat{b}_{it} (v, w) = \partial \hat{G}_t (s) / \partial x = (I_{d_X} \otimes p^{m_2} (v, w))^\prime \hat{\alpha}_t^{M_2} =: \overline{p}^{M_2} (s)^\prime \hat{\alpha}_t^{M_2}, \tag{4.3}
\]

where the second equality holds by the chain rule. I follow (4.3) in what follows for its simplicity in deriving the asymptotic properties.

To estimate \( \overline{b} \) and \( b_t (x) \), by the LIE I regress \( \hat{b}_{it} (\hat{V}_{it}, W_i) \) on the basis function \( r^{M_3} (\cdot) \) of constant one and \( X_{it} \), respectively:

\[
\hat{b} = (nT)^{-1} \sum_{t=1}^T \sum_{i=1}^n \hat{b}_{it} (\hat{v}_{it}, w_i), \text{ and } \hat{b}_t (x) = r^{M_3} (x)^\prime n^{-1} \hat{R}_t^{-1} \hat{r}_t \hat{B}_t =: r^{M_3} (x)^\prime \hat{\rho}_t^{M_3}, \tag{4.4}
\]
where \( \hat{R}_t := n^{-1} \sum_{i=1}^n r_{it}r_{it}' \), \( r_{it} := r^{M_3}(x_{it}) \), \( r_t := (r_{1t}, ..., r_{nt})' \), and \( \hat{B}_t := (\hat{b}_{1t}(\hat{v}_{1t}, w_1), ..., \hat{b}_{nt}(\hat{v}_{nt}, w_n))' \).

### 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 for brevity of exposition when there is no confusion. Let \( p_i := p^{M_2}(X_i, V_i, W_i) \) and \( P := \mathbb{E}p_ip_i' \). As Imbens and Newey (2009) have studied the convergence rates for the series estimators of conditional CDF (e.g., \( F_{X|Z} \)) and conditional expectation functions (e.g., \( \mathbb{E}[Y|X,V] \)), I adapt to their analysis to prove convergence rates for \( \hat{V}(x, z, w) \) and \( \hat{G}(s) \). I impose the following assumption.

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

(a) \( (X_i, Z_i, A_i, \varepsilon_i, \eta_i) \) is i.i.d. across \( i \).

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

(c) \( p^{m_2}(v, w) = p^{m_{2v}}(v) \otimes p^{m_{2w}}(w) \) and there exist constants \( C, \theta > 0 \) such that \( \lambda_{\min}(P) \geq C \) and \( \inf_{w \in W} f_{V,W}(v, w) \geq C \lfloor v(1-v) \rfloor^\theta \).

(d) \( G(s) \) is continuously differentiable of order \( d_2 > 1 \) on \( S \subset \mathbb{R}^{j_2} \).

(e) \( V(Y|X, Z, W) \) is bounded.

Let \( \zeta(m_2) := m_2^\theta m_2 \) and \( \zeta_1(m_2) := m_2^\theta + 2 m_2 \). With Assumption 5 in position, the next lemma follows directly from Theorem 12 of Imbens and Newey (2009).

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

\[
\mathbb{E}\left[n^{-1} \sum_{i=1}^n (\hat{V}_i - V_i)^2\right] = O\left(n^{-1} M_1 + M_1^{1-2d_1/j_1}\right) =: O\left(\Delta_{1n}^2\right),
\]

\[
\int [\hat{G}(s) - G(s)]^2 dF(s) = O_p\left(\Delta_{1n}^2 + n^{-1} m_2 + m_2^{-2d_2/j_2}\right) =: O_p\left(\Delta_{2n}^2\right), \text{ and}
\]

\[
\sup_{s \in S} \left|\hat{G}(s) - G(s)\right| = O_p\left(\zeta(m_2) \Delta_{2n}\right).
\]
Lemma 1 states that the mean-square convergence rate for $\hat{G}(s)$ is the sum of the first-step rate $\Delta_{1n}^2$, the variance term $n^{-1}m_2$, and the squared bias term $m_2^{-2d_3/j_2}$. $d_{1/j_1}$ 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 $\hat{V}(x,z,w)$ and $\hat{G}(s)$, respectively; see Assumption 3 and 5 of Imbens and Newey (2009) for the details. Note that even though the order of the basis function for the second-step estimation is $M_2$, by the index structure of (2.1) $M_2 = d_X \times m_2$ and $d_X$ is finite. Thus, the effective order that matters for the convergence rate is $m_2$.

Let $\xi_i := b_1(V_i,W_i) - b(X_i)$ and $\xi := (\xi_1,\ldots,\xi_n)'$. To derive the convergence rates of $\hat{b}$ and $\tilde{b}(x)$, 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 $\lambda_{\text{min}}(\mathbb{E} \tilde{r}^{M_3}(X_i) \tilde{r}^{M_3}(X_i)') \geq C$ and $\sup_{x \in \mathcal{X}} \|\tilde{r}^{M_3}(x)\| \leq C\zeta(M_3)$.

(c) $\mathbb{E} \left[ \xi \xi' | X \right]$ 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 $\hat{b}_1(v,w)$ and $\hat{G}(s)$ share the same series regression coefficient $\hat{\alpha}^{M_2}$, the convergence rate of $\hat{b}_1(v,w)$ to $b_1(v,w)$ is the same as that of $\hat{G}(s)$ to $G(s)$. I use this result to prove the convergence rates of $\hat{b}$ and $\tilde{b}(x)$, both of which are unknown but estimable functionals of $\hat{G}(s)$.

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

$$\left\| \hat{b} - \hat{b} \right\|^2 = O_p \left( \Delta_{2n}^2 \right),$$

$$\int \left\| \hat{b}(x) - b(x) \right\|^2 dF(x) = O_p \left( n^{-1}M_3 + M_3^{-2d_3/j_3} + \Delta_{2n}^2 \right) =: O_p \left( \Delta_{3n}^2 \right),$$

$$\sup_{x \in \mathcal{X}} \left\| \hat{b}(x) - b(x) \right\| = O_p \left( \zeta(M_3) \Delta_{3n} \right).$$

The 2002 working paper version of Imbens and Newey (2009) (henceforth IN02) has obtained asymptotic normality for estimators of known and scalar-valued linear
functions of $G(s)$. I adapt to their results to analyze vector-valued functionals of $G(s)$ via a Cramér–Wold device and prove asymptotic normality for $\widehat{b}_1(v,w)$. With slight abuse of notation, I take $\mathbf{p}^{M_2}(s)$ and $\mathbf{r}^{M_2}(x)$ that satisfy Assumptions 5 and 6 as $\mathbf{p}^{M_2}(s)$ and $r^{M_2}(x)$ 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 $\tilde{\mathbf{q}}^{M_1}(z,w) := N_q q^{M_1}(z,w)$ satisfies $\lambda_{\min}(\mathbb{E}\mathbf{q}^{M_1}(Z_i, W_i) \tilde{\mathbf{q}}^{M_1}(Z_i, W_i)^\prime) \geq C$ and $\sup_{(z,w)\in \mathbb{Z} \times \mathbb{W}} \|\tilde{\mathbf{q}}^{M_1}(z,w)\| \leq C\zeta(M_1)$.

(b) $G(s)$ is twice continuously differentiable with bounded first and second derivatives. For functional $a(\cdot)$ 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 $\tilde{\alpha}^{M_2}$ such that $\mathbb{E}\delta(S_i)^2 < \infty$, $a\left(p^{M_2}_m\right) = \mathbb{E}\delta(S_i) p^{M_2}_m(S_i)$ for all $m = 1, \ldots, M_2$, $a(G) = \mathbb{E}\delta(S_i) G(S_i)$, and $\mathbb{E}\left(\delta(S_i) - p^{M_2}(S_i) \tilde{\alpha}^{M_2}\right)^2 \to 0$; or (ii) for some $\tilde{\alpha}^{M_2}$, $\mathbb{E}\left[p^{M_2}(S_i) \tilde{\alpha}^{M_2}\right]^2 \to 0$ and $a\left(p^{M_2}(\cdot) \tilde{\alpha}^{M_2}\right)$ is bounded away from zero as $M_2 \to \infty$.

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

(d) $n M_1^{1-2d_1/j_1}, n M_2^{-2d_2/j_2}, n M_3^{-2d_3/j_3}, n^{-1} M_1^2 M_2^2 \zeta_1(M_2)^2, n^{-1} M_1 M_3 \zeta_1(M_2)^2, n^{-1} M_1 M_3 \zeta_2(M_3)^2, n^{-1} M_1 \zeta_2(M_3)^2, n^{-1} M_1 \zeta_3(M_1)^4 \zeta(M_2)^2, n^{-1} M_1 \zeta_3(M_1)^4 \zeta(M_3)^4, n^{-1} M_1 \zeta_3^2(M_1)^4 \zeta(M_2)^4, n^{-1} M_1 \zeta_3^2(M_1)^4 \zeta(M_3)^4$ and $n^{-1} M_1 \zeta_3^2(M_1)^4 \zeta(M_2)^4$ are all $o(1)$.

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

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

Also, $n M_2^{-2d_4/j_4}$ and $M_1 M_2^{-2d_4/j_4} \zeta_4(M_2)^2$ are $o(1)$.

(f) (Assumption $J(iii)$ of Andrews (1991)) For a bounded sequence of constants $\{c_{n^2} : n \geq 1\}$ and constant $pd$ matrix $\Omega_1$, it is true that $c_{n^2} \Omega_1 \overset{p}{\to} \Omega_1$, where $\Omega_1$ is defined in (A.8).

Assumptions 7(a)–(e) are also imposed by IN02 and are regularity conditions required for the asymptotic normality of $\widehat{b}_1(v,w)$. Assumption 7(f) concerns the
asymptotic covariance matrix of $\hat{b}_1(v,w)$ 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 $\hat{b}_1(v,w)$ converge at the same speed, which is a mild assumption in my setting because ex-ante I do not distinguish any coordinate of $\beta_\nu$ from the others.

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

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

where $\hat{\Omega}_1$ is defined in (A.9).

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 $\tilde{b}$ and $b(x)$ because they are unknown functionals of $G(s)$. To explain, notice that by the LIE

$$\tilde{b} = E\left[\partial G(S)/\partial X\right] \text{ and } b(x) = E\left[\partial G(S)/\partial X|X=x\right],$$

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 (4.5) and correctly account for the additional estimation bias in the asymptotic analysis.

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

(a) $E[\mathbf{r}'_i \mathbf{r}_i]$ has full column rank.

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

(c) For a sequence of bounded constants $\{c_{2n} : n \geq 1\}$ and some constant pd matrix $\Omega_2$, $c_{2n}\Omega_2 \xrightarrow{p} \Omega_2$, where $\Omega_2$ is defined in (A.12).

Assumption 8(a) is needed to show that the asymptotic covariance matrix $\Omega_2$ of $n^{1/2}(\hat{b}(x) - b(x))$ is positive definite. 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 $G(s)$.
Theorem 3 (Asymptotic Normality of $\hat{b}$ and $\hat{b}(x)$). If the conditions of Lemma 2 and Assumption 8 are satisfied, then
\[ n^{1/2}\hat{\Omega}_2^{-1/2} \left( \hat{b}(x) - b(x) \right) \xrightarrow{d} N(0, I), \]
where $\hat{\Omega}_2$ is defined in (A.15).

Furthermore, if $\mathbb{E} \| b_1(v, w) - \bar{b} \|^4 < \infty$, then
\[ n^{1/2}\hat{\Omega}_3^{-1/2} \left( \hat{b} - \bar{b} \right) \xrightarrow{d} N(0, I), \]
where $\hat{\Omega}_3$ is defined in (A.18).

5 Empirical Illustration

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 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. Results show that there are significant across-firm variations in the output elasticities. In Appendix C, I conduct a simulation study motivated by production function applications to support the findings of this illustration.

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), Brandt, Van Biesebroeck, Wang, and Zhang (2017)).
Table 1: Summary Statistics

| Variables                          | N   | mean  | sd   | min   | max   |
|------------------------------------|-----|-------|------|-------|-------|
| \( y_{it} = \ln(\text{value-added output}) \) | 46,268 | 9.558 | 1.338 | 2.236 | 16.965 |
| \( k_{it} = \ln(\text{capital}) \)       | 46,268 | 9.187 | 1.567 | 0.982 | 16.835 |
| \( l_{it} = \ln(\text{labor}) \)       | 46,268 | 5.093 | 1.082 | 2.079 | 11.972 |
| \( r_{it} = \ln(\text{real interest rate}) \) | 46,268 | 0.560 | 1.086 | -7.436 | 4.605 |
| \( w_{it} = \ln(\text{real wage}) \)       | 46,268 | 2.553 | 0.511 | -0.278 | 6.115 |
| Year                                | 4   | -     | -    | 2004  | 2007  |
| Firm ID                            | 11,567 | -     | -    | -     | -     |
| Industry Code                      | 5   | -     | -    | -     | -     |

I focus on the five largest 2-digit sectors in terms of the number of firms between 2004 and 2007.\(^{22,23}\) 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 11,567 firms over four years across five sectors. Summary statistics of the key variables are presented in Table 1.

The value-added production function I consider is

\[
y_{it} = k_{it}^\beta_k + l_{it}^\beta_l + \omega_{it} + \epsilon_{it},
\]

\[
\beta_{it}^K = \beta (A_i, \varepsilon_{it}, \upsilon_{it}), \quad \beta_{it}^L = \beta (A_i, \varepsilon_{it}, \upsilon_{it}), \quad \omega_{it} = \omega (A_i, \varepsilon_{it}, \upsilon_{it}),
\]

\[
k_{it} = g^K (Z_{it}, A_i, \eta_{it}^K), \quad l_{it} = g^L (Z_{it}, A_i, \eta_{it}^L).
\]

(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 \( \varepsilon_{it} \).

It is worth noting that the output is measured by the total revenue in dollars, rather than by the physical quantity produced in units, due to the absence of indi-

\(^{22}\)There are 3140 textile firms, 2186 chemical firms, 1804 nonmetallic minerals firms, 2869 general equipment firms, and 1568 transportation equipment firms in the final dataset. I order the industries by their 2-digit Chinese Industry Classification codes.

\(^{23}\)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 the financial crisis in the early 2000s, and (iii) use the most recent data.
individual output prices in the data. When firms operate in distinct imperfectly competitive output markets, this may cause problems (Klette and Griliches (1996)). I use input prices $Z_{it} := (r_{it}, w_{it})'$ as IVs for $X_{it}$. For estimation, I take $W_i$ to be the mean through time of each coordinate of $X_{it}$. I estimate each coordinate of $V_{it} := (F_{it|Z_{it}, W_i}, F_{lt|Z_{it}, W_i})'$ by regressing $1 (k \leq k_{it})$ and $1 (l \leq l_{it})$ on the second-degree polynomial spline basis of $(Z_{it}, W_i)$ with its knot at the median, respectively. Next, I estimate $G_{it} := \mathbb{E} [y_{it} | X_{it}, V_{it}, W_i]$ by regressing $y_{it}$ on the same spline basis of $(X_{it}, \hat{V}_{it}, W_i)$. Estimation of $b_{it} (V_{it}, W_i) := \mathbb{E} [ \beta_{it} | V_{it}, W_i]$ is then obtained by taking the partial derivative of $\hat{G}_{it} (X_{it}, \hat{V}_{it}, W_i)$ with respect to $X_{it}$. Finally, I estimate $\tilde{b} := \mathbb{E} \beta_{it}$ by averaging $\hat{b}_{it} (\hat{V}_{it}, W_i)$ over $i$ and $t$ and $b_t (X_{it}) := \mathbb{E} [ \beta_{it} | X_{it}]$ by regressing $\hat{b}_{it} (\hat{V}_{it}, W_i)$ on the second-degree polynomial spline basis of $X_{it}$ with its knot at the median.

5.2 Results

First, I summarize the results for estimating the average output elasticities defined as $\hat{b} := \left( \hat{b}^K, \hat{b}^L \right)'$ in Table 2. I compare my estimated $\hat{b}$'s (TERC) with those derived by applying the methods of Olley and Pakes (1996) (OP), Levinsohn and Petrin (2003) (LP), and Ackerberg, Caves, and Frazer (2015) (ACF) to the same data set. Standard errors (se) and 95% confidence intervals (CI) are also included. My estimated average capital elasticities are within the range of $[0.379, 0.504]$ across the five sectors, which are consistent with the results obtained from applying OP, LP, and ACF to the same dataset. The estimated average labor elasticities present more discrepancies between methods. My estimates of $\hat{b}^L$'s lie between $[0.329, 0.460]$ across the five sectors, similar to OP’s results. LP seem to derive small $\hat{b}^L$'s for all sectors, while ACF generate larger $\hat{b}^L$'s for the general equipment sector and transportation equipment sector. My 95% CI’s for $\hat{b}$'s are reasonably tight across

---

24 Real wages are very likely to be exogenous because the labor market in the manufacturing sectors of China is close to a perfectly competitive market in which firms face a relatively flat supply curve. Real interest rate may be considered as a valid IV because its fluctuation is mostly driven by monetary policy set by the central bank of China.

25 I use the Stata commands prodest (Rovigatti and Mollisi (2018)) for OP and LP and acfест (Manjón and Manez (2016)) for ACF for point estimates of $\hat{b}$ and 95% CI. See Keiller, de Paula, and Van Reenen (2024) for a discussion of the implementation of the ACF method. For TERC, since $\hat{b}$ has slower than $n^{1/2}$ rate by Theorem 2, I follow Politis, Romano, and Wolf (1999) to randomly subsample $b = [4n^{3/4}]$ firms from each sector without replacement for 1,000 times.
|                         | OP   | LP   | ACF   | TERC  |
|-------------------------|------|------|-------|-------|
| **Capital Elasticity**  | 0.359| 0.252| 0.300 | 0.387 |
| se                      | (0.030) | (0.023) | (0.037) | (0.031) |
| 95% CI                  | [0.300, 0.417] | [0.206, 0.297] | [0.227, 0.373] | [0.334, 0.455] |
| **Labor Elasticity**    | 0.470| 0.175| 0.567 | 0.375 |
| se                      | (0.014) | (0.011) | (0.065) | (0.023) |
| 95% CI                  | [0.442, 0.498] | [0.154, 0.196] | [0.440, 0.695] | [0.327, 0.418] |
| **Chemical**            |      |      |       |       |
| **Capital Elasticity**  | 0.294| 0.288| 0.344 | 0.380 |
| se                      | (0.036) | (0.031) | (0.096) | (0.037) |
| 95% CI                  | [0.223, 0.365] | [0.228, 0.348] | [0.155, 0.533] | [0.306, 0.451] |
| **Labor Elasticity**    | 0.296| 0.113| 0.378 | 0.329 |
| se                      | (0.022) | (0.015) | (0.213) | (0.026) |
| 95% CI                  | [0.253, 0.339] | [0.084, 0.143] | [-0.039, 0.795] | [0.276, 0.377] |
| **Nonmetallic Mineral**|      |      |       |       |
| **Capital Elasticity**  | 0.697| 0.311| 0.236 | 0.379 |
| se                      | (0.101) | (0.024) | (0.101) | (0.037) |
| 95% CI                  | [0.499, 0.895] | [0.263, 0.358] | [0.038, 0.434] | [0.314, 0.451] |
| **Labor Elasticity**    | 0.353| 0.071| 0.601 | 0.459 |
| se                      | (0.021) | (0.010) | (0.384) | (0.027) |
| 95% CI                  | [0.311, 0.394] | [0.051, 0.091] | [-0.151, 1.354] | [0.404, 0.514] |
| **General Equipment**   |      |      |       |       |
| **Capital Elasticity**  | 0.416| 0.246| 0.176 | 0.504 |
| se                      | (0.076) | (0.023) | (0.082) | (0.033) |
| 95% CI                  | [0.267, 0.565] | [0.202, 0.291] | [0.016, 0.337] | [0.451, 0.578] |
| **Labor Elasticity**    | 0.444| 0.071| 0.927 | 0.359 |
| se                      | (0.019) | (0.009) | (0.149) | (0.020) |
| 95% CI                  | [0.406, 0.482] | [0.053, 0.089] | [0.635, 1.218] | [0.320, 0.397] |
| **Transportation Equipment** |      |      |       |       |
| **Capital Elasticity**  | 0.523| 0.281| 0.217 | 0.383 |
| se                      | (0.064) | (0.033) | (0.084) | (0.035) |
| 95% CI                  | [0.397, 0.649] | [0.216, 0.346] | [0.052, 0.381] | [0.327, 0.469] |
| **Labor Elasticity**    | 0.523| 0.137| 1.042 | 0.460 |
| se                      | (0.025) | (0.019) | (0.138) | (0.028) |
| 95% CI                  | [0.473, 0.573] | [0.100, 0.173] | [0.771, 1.313] | [0.407, 0.517] |
all sectors. It is well documented in the literature that output elasticities in Cobb-Douglas production function estimation usually lie within $[0, 1]$. Hsieh and Klenow (2009) also calculate that roughly half of output is distributed to capital, according to Chinese input-output tables and national accounts. My estimates are consistent with the empirical evidence in the literature.

Next, I investigate the distribution of the conditional means of the output elasticities given the regressors. Policymakers could use the information contained in the conditional means to formulate more accurate sector specific policies, as these conditional means provide a fine approximation to the average elasticities for each subgroup of firms with a certain level of capital and labor. Moreover, distributional properties about the true output elasticities can be inferred from the distribution of the conditional means of the output elasticities. For example, by the law of total variance, the variance of the conditional means of the output elasticities provides a lower bound for the variance of the true output elasticities. Demirer (2022) argues that the heterogeneity in output elasticities is largely explained by across-firm variation. To investigate into this explanation, for each firm I calculate the averages of $\hat{\beta}^K_i (X_{it})$ and $\hat{\beta}^L_i (X_{it})$ through time and denote them by $\bar{\beta}^K_i$ and $\bar{\beta}^L_i$, respectively. Then, I plot the histograms of $\bar{\beta}^K_i$ and $\bar{\beta}^L_i$ across firms for each of the five sectors.

In Figure 2, I present the histograms of $\bar{\beta}^K_i$ and $\bar{\beta}^L_i$ for the textile sector and indicate the corresponding $\hat{b}$’s labeled by “Mean” on the same graph. The left subplot of Figure 2 shows the histogram of $\bar{\beta}^K_i$. All of its probability mass lies between zero and one.
and one, with its mode at 0.4. Over 90% of the probability mass of $\hat{\beta}_K$ lies between 0.3 and 0.5. The right subplot of Figure 2 presents the histogram of $\hat{\beta}_L$ across firms for the textile sector. Again, the majority of its probability mass lies between zero and one, with its mode at 0.38. The distributions of $\hat{\beta}_K$ and $\hat{\beta}_L$ are concentrated around the mean, suggesting that firms in the textile industry tend to be homogeneous in terms of capital and labor efficiency.

In Figure 3, I present the histograms of $\hat{\beta}_K$ and $\hat{\beta}_L$ for the other four sectors. First, the majority of $\hat{\beta}_K$ and $\hat{\beta}_L$ across all sectors lie between zero and one, which is consistent with empirical evidence. Second, I find significant across-firm variations in both capital and labor elasticities measured by $\hat{\beta}_K$ and $\hat{\beta}_L$ within each of the five sectors. However, the degree of heterogeneity in these elasticities differs. For example, in the left subplot of Figure 3, there is a larger variance in $\hat{\beta}_K$ in the nonmetallic mineral sector than in the chemical sector, which implies that the firms in the chemical industry may be more homogeneous in capital efficiency than those in the nonmetallic mineral sector.

6 Conclusion

In this paper, I propose a new TERC model in which the regressors are correlated with the random coefficients not only through a fixed effect but, more importantly, through a time-varying random shock—a feature that aligns with agent optimization behavior in many empirical applications. I construct feasible control variables for
both the fixed effect and the random shock, and subsequently utilize the residual variation in the regressors given the control variables to identify the APE and LAR. I provide three-step series estimators and establish their convergence rates and asymptotic normality. The empirical exercise of this paper reveals significant variation in the output elasticities across manufacturing firms in China.

I propose two directions for future research. First, beyond the first- and second-order moments analyzed in this paper, policymakers may also be interested in the density functions of the random coefficients. One possible, albeit laborious, approach is to identify the moments of all orders of the random coefficients by induction, which would uniquely determine their distribution (Stoyanov (2000)). Second, it remains an open question whether aspects of my method could be extended to dynamic linear or nonlinear panel data models, such as those discussed by Marx, Tamer, and Tang (2024) and Liu, Poirier, and Shiu (2024). Such an extension would likely require additional structure on how the lagged dependent variable correlates 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, 307–45.

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.

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., X. CHEN, AND D. KRISTENSEN (2007a): “Semi-nonparametric IV estimation of shape-invariant Engel curves,” *Econometrica*, 75, 1613–1669.

BLUNDELL, R., D. KRISTENSEN, AND R. L. MATZKIN (2013): “Control functions and simultaneous equations methods,” *American Economic Review*, 103, 563–69.

BLUNDELL, R., T. MACURDY, AND C. MEGHIR (2007b): “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.

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.

CHERNOZHUKOV, V., J. A. HAUSMAN, AND W. K. NEWEY (2019): “Demand analysis with many prices,” Working Paper.

DEMIRER, M. (2022): “Production function estimation with factor-augmenting technology: an application to markups,” Working Paper.

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.

HAUSMAN, J. A. AND W. E. TAYLOR (1981): “Panel data and unobservable individual effects,” Econometrica, 1377–1398.

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.

KASY, M. (2014): “Instrumental variables with unrestricted heterogeneity and continuous treatment,” Review of Economic Studies, 81, 1614–1636.

KEILLER, A. N., Á. DE PAULA, AND J. VAN REENEN (2024): “Production function estimation using subjective expectations data,” Working Paper.

KLETTE, T. J. AND Z. GRILICHES (1996): “The inconsistency of common scale estimators when output prices are unobserved and endogenous,” Journal of Applied Econometrics, 11, 343–361.

LAAGE, L. (2024): “A correlated random coefficient panel model with time-varying endogeneity,” Journal of Econometrics, forthcoming.

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 (2024): “Identification and estimation of partial effects in nonlinear semiparametric panel models,” Journal of Econometrics, forthcoming.
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.

NEWEY, W. K. (1994): “The asymptotic variance of semiparametric estimators,” *Econometrica*, 62, 1349–1382.

——— (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.

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

POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*, Springer.
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.

WANG, Y., P. C. PHILLIPS, AND L. SU (2024): “Panel data models with time-varying latent group structures,” *Journal of Econometrics*, forthcoming.

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.
A 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 $\mathbb{E}[\beta_{it} X_{it}, V_{it}, W_i]$ does not depend on $X_{it}$ via the LIE. Finally, I identify $\mathbb{E}_\beta$ and $\mathbb{E}_I$ by leveraging the residual variation in $X_{it}$ given $V_{it}$ and $W_i$. I show how the inclusion of exogenous shocks $\nu_{it}$ and $\epsilon_{it}$ 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 $X_{it}$.\(^{26}\) By Assumption 2, I have $A_i \perp (X_{it}, Z_{it})|W_i$, which implies $X_{it} \perp A_i|(Z_{it}, W_i)$. Thus, for each $l \in \{1, ..., d_X\}$ and any on-support $(x^{(l)}, z, a, w)$, I have

$$F_{X_{it}^{(l)}}|Z_{it}, W_i \left(x^{(l)} \mid z, w\right) = F_{X_{it}^{(l)}}|Z_{it}, A_i, W_i \left(x^{(l)} \mid z, a, w\right)$$

$$= \mathbb{P} \left(g^{(l)}(z, a, \eta_{it}^{(l)}) \leq x^{(l)} \mid Z_{it} = z, A_i = a, W_i = w\right)$$

$$= \mathbb{P} \left(\eta_{it}^{(l)} \leq g^{(l)^{-1}}(x^{(l)}, z, a) \mid A_i = a, W_i = w\right)$$

$$= F_{\eta_{it}^{(l)}}|A_i, W_i \left(g^{(l)^{-1}}(x^{(l)}, z, a) \mid a, w\right),$$

where the first equality holds by $X_{it} \perp A_i|(Z_{it}, W_i)$, the second uses (2.2), the third holds by Assumptions 1 and 3(a), and the last holds by definition of the conditional CDF of $\eta_{it}^{(l)}$ given $(A_i, W_i)$. By (2.2), the random variable $\eta_{it}^{(l)} = g^{(l)^{-1}}(X_{it}^{(l)}, Z_{it}, A_i)$ for each $l$, so that plugging in gives

$$V_{it}^{(l)} := F_{X_{it}^{(l)}}|Z_{it}, W_i \left(X_{it}^{(l)} \mid Z_{it}, W_i\right) = F_{\eta_{it}^{(l)}}|A_i, W_i \left(\eta_{it}^{(l)} \mid A_i, W_i\right), \quad (A.1)$$

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

Next, I have

$$\mathbb{E}[\beta_{it} X_{it}, A_i, V_{it}, W_i]$$

\(^{26}\)This is without loss of generality because under Assumption 1, I can always redefine $\bar{\eta}$ to be the vector that collects the coordinates of $\eta$ that enters each coordinate of $g(\cdot)$ function.
\[= \mathbb{E} \left[ \beta \left( A_i, \varepsilon_{it} \right) | g \left( Z_{it}, A_i, \eta_{it} \right), A_i, V_{it}, W_i \right] \]
\[= \mathbb{E} \left[ \beta \left( A_i, \varepsilon_{it} \right) | A_i, V_{it}, W_i \right] \]
\[=: b_2 \left( A_i, V_{it}, W_i \right), \tag{A.2} \]
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.1), 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 \(\mathbb{E} \left[ \beta_{it} | X_{it}, A_i, V_{it}, W_i \right] \) and write it out as \(b_2 \left( A_i, V_{it}, W_i \right)\).

By (A.2), I have
\[
\mathbb{E} \left[ \beta_{it} | X_{it}, V_{it}, W_i \right] \]
\[= \mathbb{E} \left[ b_2 \left( A_i, V_{it}, W_i \right) | X_{it}, V_{it}, W_i \right] \]
\[= \mathbb{E} \left[ \mathbb{E} \left[ b_2 \left( A_i, V_{it}, W_i \right) | X_{it}, Z_{it}, V_{it}, W_i \right] | X_{it}, V_{it}, W_i \right] \]
\[= \mathbb{E} \left[ \mathbb{E} \left[ b_2 \left( A_i, V_{it}, W_i \right) | X_{it}, Z_{it}, W_i \right] | X_{it}, V_{it}, W_i \right] \]
\[= \mathbb{E} \left[ \int b_2 \left( a, V_{it}, W_i \right) f_{A_i | X_{it}, Z_{it}, W_i} \left( a | X_{it}, Z_{it}, W_i \right) \mu \left( da \right) | X_{it}, V_{it}, W_i \right] \]
\[= \mathbb{E} \left[ \int b_2 \left( a, V_{it}, W_i \right) f_{A_i | W_i} \left( a | W_i \right) \mu \left( da \right) | X_{it}, V_{it}, W_i \right] \]
\[=: b_1 \left( V_{it}, W_i \right), \tag{A.3} \]
where the first and second equalities hold by the LIE, 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.3), I have
\[
\mathbb{E} \left[ Y_{it} | X_{it}, V_{it}, W_i \right] = X_{it}'b_1 \left( V_{it}, W_i \right). \tag{A.4} \]
When Assumption 4 is satisfied, I pre-multiply both sides of (A.4) by \(X_{it}\) and take conditional expectation of both sides conditioning on \(V_{it}\) and \(W_i\):
\[
\mathbb{E} \left[ \mathbb{E} \left[ X_{it}Y_{it} | X_{it}, V_{it}, W_i \right] | V_{it}, W_i \right] = \mathbb{E} \left[ X_{it}X_{it}' | V_{it}, W_i \right] b_1 \left( V_{it}, W_i \right), \]
which identifies \(b_1 \left( V_{it}, W_i \right)\)
\[
b_1 \left( V_{it}, W_i \right) = \left( \mathbb{E} \left[ X_{it}X_{it}' | V_{it}, W_i \right] \right)^{-1} \mathbb{E} \left[ X_{it}Y_{it} | V_{it}, W_i \right]. \]
When Assumption 4′ is used, I take partial derivative of both sides of (A.4) with respect to \(X_{it}\) and obtain
\[
b_1 (V_{it}, W_i) = \frac{\partial \mathbb{E} [Y_{it} | X_{it}, V_{it}, W_i]}{\partial X_{it}}.
\]

Given \(b_1 (V_{it}, W_i)\), I use the LIE to identify \(b = \mathbb{E} \beta_{it}\) and \(b (X_{it}) := \mathbb{E} [b_1 (V_{it}, W_i) | X_{it}]\), respectively.

When ex-post shock \(\epsilon_{it}\) is included additively in (2.1) and \(\mathbb{E} \epsilon_{it} = 0\), equations (A.3) and (A.4) still hold, so the identification result (A.5) holds without any changes.

When ex-post shock \(\upsilon_{it}\) is also included in \(\beta_{it} := \beta (A_i, \eta_{it}, v_{it})\), I need to change \(b_2 (A_i, V_{it}, W_i)\) in (A.2) to be \(b_2t (A_i, V_{it}, W_i)\) because although \(\upsilon_{it}\) is independent of all other variables, the PDF of \(\upsilon_{it}\) itself may depend on \(t\), i.e.,
\[
\mathbb{E} [\beta_{it} | X_{it}, A_i, V_{it}, W_i]
= \mathbb{E} [\beta (A_i, \epsilon_{it}, \upsilon_{it}) | g (Z_{it}, A_i, \eta_{it}), A_i, V_{it}, W_i]
= \mathbb{E} [\beta (A_i, \epsilon_{it}, \upsilon_{it}) | A_i, V_{it}, W_i]
=: b_{it} (A_i, V_{it}, W_i).
\]

As a result, I need to change \(b (X_{it})\) to be \(b_t (X_{it})\) in (A.5). The rest of the proof goes through as before.

When exogenous regressors are included in (2.1), I can define \(X_{it} = (U_{it}', Z_{1,it}')'\) and \(Z_{it} = (Z_{1,it}', Z_{2,it}')'\), and rewrite model (2.1)–(2.2) to be
\[
Y_{it} = X_{it}' \beta (A_i, \epsilon_{it}),
U_{it} = g (Z_{it}, A_i, \eta_{it}).
\]

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.4) to identify $b_1 (V_{it}, W_i)$. As discussed in footnote 16, the residual variation requirement of Assumption 4 or 4′ is easier to be satisfied when exogenous $Z_{1,it}$ is included in $X_{it}$ as the support of $Z_{1,it}$ 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 $\hat{b} (x)$, since the result for $\hat{P}$ follows immediately by setting $r^{M_3} (\cdot) \equiv 1$. Let $p^{m_2}_{i} := p^{m_2} (V_i, W_i)$, $r_i := r^{M_3} (X_i)$, and $\bar{p}_i = \bar{p}^{M_2} (S_i)$. Following Imbens and Newey (2009), I can normalize $\mathbb{E} p^{m_2}_{i} p^{m_2'}_{i} = I_{m_2}$ by Assumption 5 and $R := \mathbb{E} r_i r'_i = I_{M_3}$ by Assumption 6, which imply $\hat{P} := \mathbb{E} \hat{p}_i \bar{p}_i = \mathbb{E} \left[ I_{d_X} \otimes (\hat{p}^{m_2}_{i} \hat{p}^{m_2'}_{i}) \right] = I_{M_2}$. Furthermore, I have $\lambda_{\min} (\hat{R}) \geq C > 0$ with probability approaching one by Newey (1997). Let $\hat{B} := (b_1 (\hat{v}_1, w_1), ..., b_1 (\hat{v}_n, w_n))'$.

By (4.4),

$$\left\| n^{1/2} \left( \hat{\rho}^{M_3} - \rho^{M_3} \right) \right\|^2 / 4 \leq \left( \hat{B} - \bar{B} \right)' \hat{R}^{-1} r' \left( \hat{B} - \bar{B} \right) + \left( \hat{B} - B \right)' \hat{R}^{-1} r' \left( \hat{B} - B \right) + \left( B - B^X \right)' \hat{R}^{-1} r' \left( B - B^X \right) \left( B^X - r \rho^{M_3} \right)' \hat{R}^{-1} r' \left( B^X - r \rho^{M_3} \right), \quad (A.6)$$

where $B := (b_1 (v_1, w_1), ..., b_1 (v_n, w_n))'$ and $B^X := (b (x_1), ..., b (x_n))'$. Hence, $\xi = B - B^X$. I analyze the RHS of (A.6) term by term.

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

$$n^{-2} \left( \hat{B} - \bar{B} \right)' \hat{R}^{-1} r' \left( \hat{B} - \bar{B} \right) \leq C n^{-1} \left( \hat{B} - \bar{B} \right)' \left( \hat{B} - \bar{B} \right) = C n^{-1} \sum_i \left( \hat{p}_i (\hat{\alpha}^{M_2} - \alpha^{M_2}) + \left( \hat{p}_i' \alpha^{M_2} - b_1 (\hat{v}_i, w_i) \right) \right)^2 \leq C \left\| \hat{\alpha}^{M_2} - \alpha^{M_2} \right\|^2 + C \sup_{s \in S} \left\| \bar{p}^{M_2} (s)' \alpha^{M_2} - b_1 (v, w) \right\|^2 = O_p \left( \Delta^2_{2n} \right), \quad (A.7)$$

where the first inequality holds because $n^{-1} r \hat{R}^{-1} r'$ is idempotent, the last inequality holds by the Cauchy-Schwarz inequality (CS) and $\left\| n^{-1} \sum_i \hat{p}_i \bar{p}_i - I \right\| = o_p (1)$, and the last equality uses Lemma 1.
Next,
\[
n^{-2} (\tilde{B} - B)' r \tilde{R}^{-1} r' (\tilde{B} - B) \leq C n^{-1} \sum_{i} (b_i (\tilde{v}_i, w_i) - b_i (v_i, w_i))^2 \leq C n^{-1} \sum_{i} (\tilde{v}_i - v_i)^2 = O_p \left( \Delta_{1n}^2 \right),
\]
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.6),
\[
n^{-2} \mathbb{E} \left[ \left( B - B^\Delta \right)' r \tilde{R}^{-1} r' \left( B - B^\Delta \right) \right] = n^{-2} \text{tr} \left\{ \mathbb{E} \left[ \xi' r \tilde{R}^{-1} r' \xi \right] \right\} = n^{-2} \text{tr} \left\{ \mathbb{E} \left[ \xi' \right] r \tilde{R}^{-1} r' \right\} \leq n^{-2} \text{tr} \left\{ C I r \tilde{R}^{-1} r' \right\} = C n^{-1} \text{tr} \left\{ \tilde{R}^{-1} \right\} = C n^{-1} M_3,
\]
and
\[
n^{-2} (B^\Delta - R \rho_3)' r \tilde{R}^{-1} r' (B^\Delta - R \rho_3) \leq n^{-1} \| B^\Delta - R \rho_3 \|^2 = O_p \left( M_3^{-2d_3/j_3} \right).
\]

By \( \lambda_{\text{min}} (\hat{R}) \geq C > 0 \) and the result that \( \mathbb{E} [|Y_n| | Z_n] = O_p (r_n) \) implies \( Y_n = O_p (r_n) \) (Conditional Markov’s inequality, henceforth denoted as “CM”),
\[
\| \hat{\rho}_3 - \rho_3 \|^2 = O_p \left( \Delta_{2n}^2 + n^{-1} M_3 + M_3^{-2d_3/j_3} \right) =: O_p \left( \Delta_{3n}^2 \right),
\]
which yields
\[
\int \| \hat{b} (x) - b (x) \|^2 dF (x) \leq \int \left( r^{M_3} (x)' (\hat{\rho}_3 - \rho_3) + \left( r^{M_3} (x)' \rho_3 - b (x) \right) \right)^2 dF (x) \leq 2 \| \hat{\rho}_3 - \rho_3 \|^2 + 2 \sup_{x \in \mathcal{X}} b (x) - r^{M_3} (x)' \rho_3 \|^2 = O_p \left( \Delta_{3n}^2 \right),
\]
and
\[
\sup_{x \in \mathcal{X}} \| \hat{b} (x) - b (x) \| \leq \sup_{x \in \mathcal{X}} \| r^{M_3} (x)' \| \| \hat{\rho}_3 - \rho_3 \| + \sup_{x \in \mathcal{X}} b (x) - r^{M_3} (x)' \rho_3 \| = O_p \left( \zeta (M_3) \Delta_{3n} \right).
\]
\[\square\]
**Proof of Lemma 2.** Define

\[ \Omega_1 := \mathbf{P}^{M_2} (v, w) \left( \Sigma + \hat{\Sigma} \right) \left( \mathbf{P}^{M_2} (v, w) \right)^{-1}, \quad \Sigma := \mathbb{E} p_i p_i' u_i^2, \]  
(A.8)

\[ \Sigma_1 := \mathbb{E} p_i' \hat{\Sigma}_i p_i, \quad p_i := \mathbf{P}^{M_2} (S_i), \quad q_i := q^M_i (X_i, Z_i, W_i), \]

\[ \hat{\Sigma}_i := \mathbb{E} \left[ G_V (S_j) \hat{\tau}' (V_j) p_j q_j' Q^{-1} q_j v_{ji} | \mathcal{I}_i \right], \quad u_i := Y_i - G (S_i), \]

\[ G_V (S_j) := \partial G (s) / \partial v | s = S_j, \quad \text{and } v_{ji} := 1 \{ x_i \leq x_j \} - F (x_j | z_i, w_i). \]

and

\[ \hat{\Omega}_1 := \mathbf{P}^{M_2} (v, w) \left( \hat{\Sigma} + \hat{\Sigma}_1 \right) \left( \mathbf{P}^{M_2} (v, w) \right)^{-1}, \]  
(A.9)

\[ \hat{\Sigma} := n^{-1} \sum_{i=1}^n \hat{p}_i \hat{p}_i' (y_i - \mathcal{G} (\hat{s}_i))^2, \quad \hat{\Sigma}_1 := n^{-1} \sum_{i=1}^n \hat{p}_i' \hat{p}_i'', \]

\[ \hat{\mu}_i := n^{-1} \sum_{j=1}^n \mathcal{G}_V (\hat{s}_j) \hat{q}_j q_j' Q^{-1} q_j v_{ji}, \quad \text{and } v_{ji} := \left( \mathbb{I} \{ x_i \leq x_j \} - \bar{F} (x_j | z_i, w_i) \right). \]

IN02 have proved asymptotic normality for known and scalar-valued functionals of \( G (s) \). I apply their results to \( c' \hat{b}_1 (v, w) \) for any constant vector \( c' c = 1 \) and obtain

\[ c' n^{1/2} \Omega^{-1/2}_1 \left( \hat{b}_1 (v, w) - b_1 (v, w) \right) \xrightarrow{d} N (0, 1) \quad \text{and} \]

\[ \left( c' \Omega_1 c \right)^{-1} \left[ c' \left( \hat{\Omega}_1 - \Omega_1 \right) c \right] \xrightarrow{p} 0. \]  
(A.10)

By (A.10) and Assumption 7(f),

\[ c' \left( c_1 \hat{\Omega}_1 - c_1 \Omega_1 \right) c \xrightarrow{p} 0, \]

which implies

\[ c_1 \hat{\Omega}_1 \xrightarrow{p} \Omega_1. \]  
(A.11)

Then,

\[ n^{1/2} \hat{\Omega}_1^{-1/2} \left( \hat{b}_1 (v, w) - b_1 (v, w) \right) \]

\[ = \left( c_1 \hat{\Omega}_1 \right)^{-1/2} \left( c_1 \Omega_1 \right)^{1/2} n^{1/2} \Omega_1^{-1/2} \left( \hat{b}_1 (v, w) - b_1 (v, w) \right) \]

\[ \xrightarrow{d} \Omega_1^{1/2} \Omega_1^{1/2} N (0, I) = \Omega_1 N (0, I), \]

where the convergence holds by (A.10), the Cramér–Wold device, (A.11), Assumption
Proof of Theorem 3. Following the proof of Lemma 2, under regularity conditions one may use the Cramér–Wold device to extend the results from scalar-valued functionals to vector-valued functionals of $G(s)$. Therefore, I maintain that $b(x)$ is a scalar in this proof. First, I derive the influence functions for $\hat{b}(x)$ 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 $\hat{b}(x)$, which can be used to construct asymptotically valid confidence intervals. I write $r_{M3}(x)$ as $r(x)$ when there is no confusion. The proof relies on the results from IN02, and I point out differences.

Define

$$
\Omega_2 := \Omega_{21} + \Omega_{22}, \quad \Omega_{21} := \mathbb{E} \left( A_1 P^{-1} p_i u_i \right) \left( A_1 P^{-1} p_i u_i \right)' ,
$$

(A.12)

$$
\Omega_{22} := \mathbb{E} \left( A_1 P^{-1} \bar{p}_i' - r_{M3}(x)' \left( \bar{p}_i' + r_i \xi_i \right) \right) \left( A_1 P^{-1} \bar{p}_i' - r_{M3}(x)' \left( \bar{p}_i' + r_i \xi_i \right) \right)',
$$

$$
A_1 := r_{M3}(x)' \mathbb{E} r_i \bar{p}_i , \quad \bar{p}_i' := \mathbb{E} \left[ r_j b_{1,v} (V_j, W_j) q_j Q^{-1} q_i v_{ji} \bigg| I_i \right] ,
$$

and $b_{1,v} (V_j, W_j) := \partial b_1 (v, w) / \partial v |_{v=V_j, w=W_j}$, and $F := \Omega_2^{-1/2}$, which is well-defined because $\Omega_2 = \Omega_{21} + \Omega_{22}$ and

$$
\Omega_{21} = A_1 P^{-1} \left( \mathbb{E} p_i p_i' u_i^2 \right) P^{-1} A_1' \\
= A_1 P^{-1} \left( \mathbb{E} \left[ p_i p_i' \mathbb{E} \left( u_i^2 \big| X_i, Z_i, W_i \right) \right] \right) P^{-1} A_1' \\
\geq CA_1 P^{-1} A_1' = C r(x)' \left( \mathbb{E} r_i \bar{p}_i \right) P^{-1} \left( \mathbb{E} r_i \bar{p}_i \right)' r(x) > 0,
$$

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

Define functionals:

$$
\hat{a} \left( \hat{b}_1, \hat{V} \right) := \hat{\mathbb{E}} \left[ \hat{b}_1 (\hat{V}, W) \bigg| X = x \right] = \hat{b}(x),
$$

$$
\hat{a} \left( b_1, \bar{V} \right) := \hat{\mathbb{E}} \left[ b_1 (\bar{V}, W) \bigg| X = x \right],
$$

$$
\hat{a} \left( b_1, V \right) := \hat{\mathbb{E}} [b_1 (V, W) | X = x] ,
$$

and

$$
a \left( b_1, V \right) = \mathbb{E} [b_1 (V, W) | X = x] = b(x).$$
I expand
\[ n^{1/2} F \left( \hat{a} (b_1, \bar{V}) - a (b_1, V) \right) \]
\[ = n^{1/2} F \left( \hat{a} (b_1, \bar{V}) - \hat{a} (b_1, \bar{V}) + \hat{a} (b_1, \bar{V}) - \hat{a} (b_1, V) + \hat{a} (b_1, V) - a (b_1, V) \right) \]
\[ = n^{-1/2} \sum_{i} (\psi_{1i} + \psi_{2i} + \psi_{3i}) + o_p (1), \]
and show that
\[ \psi_{1i} = H_1 \left( p_i u_i - \overline{\mu}_i \right), \psi_{2i} = H_2 \overline{\mu}_i^I, \text{ and } \psi_{3i} = H_2 r_i \xi_i, \]
where \( H_1 := FA_1 P^{-1}, \ A_1 = r (x)' R^{-1} \mathbb{E} r_i \overline{\mu}_i = r (x)' \mathbb{E} r_i \overline{\mu}_i, \ r_i := r (x_i), \ \overline{\mu}_i := \mathbb{E} [G_V (S_j) \tau' (V_j) p_j q_j Q^{-1} q_i v_j] L_i, \)
\( G_V (S_j) := \partial G (s) / \partial v \big|_{v=s_j}, \ v_j := 1 \{ x_i \leq x_j \} - F (x_j | z_i, w_i), \)
\( H_2 := FA_2 R^{-1} = FA_2, \ A_2 := r (x), \ \overline{\mu}_i^I := \mathbb{E} [r_j b_{1,V} (V_j, W_j) q_j' Q^{-1} q_i v_j] L_i, \)
\( b_{1,V} (V_j, W_j) = \partial b_1 (v, w) / \partial v \big|_{v=v_j}, \ \text{and } \xi_i = b_1 (v_i, w_i) - b (x_i). \)

Let \( \hat{H}_1 := FA_1 \hat{P}^{-1}, \ \hat{A}_1 := r (x)' \hat{R}^{-1} n^{-1} \sum_{i=1}^{n} r_i \hat{\overline{\mu}}_i, \ \hat{R} := n^{-1} \sum_{i=1}^{n} r_i r_i' \hat{\overline{\mu}}_i := \overline{\mu} (\hat{s}_i), \)
\( \hat{H}_2 := FA_2 \hat{R}^{-1}, \ G := (G (s_1), ..., G (s_n))', \ \text{and } \hat{G} := (G (\hat{s}_1), ..., G (\hat{s}_n))'. \)

First, for \( \psi_{1i}, \)
\[ n^{1/2} F \left( \hat{a} (b_1, \bar{V}) - \hat{a} (b_1, \bar{V}) \right) \]
\[ = n^{-1/2} F r (x)' \hat{R}^{-1} r' \left( \hat{B} - \hat{B} \right) \]
\[ = n^{-1/2} F r (x)' \hat{R}^{-1} r' \left( n^{-1} \overline{\mu} \hat{P}^{-1} \hat{\overline{\mu}} Y - \hat{B} \right) \]
\[ = n^{-1/2} F r (x)' \hat{R}^{-1} r' \left[ n^{-1} \overline{\mu} \hat{P}^{-1} \hat{\overline{\mu}} \left( Y - G + G - \hat{G} + \hat{G} - \hat{\mu} x M_2 \right) + \left( \hat{\mu} x M_2 - \hat{B} \right) \right] \]
\[ = n^{-1/2} \sum_{i} \hat{H}_1 \hat{p}_i [u_i - (G (\hat{s}_i) - G (s_i))] + n^{-1/2} \hat{H}_1 \hat{p} r' \left( \hat{G} - \hat{\mu} x M_2 \right) \]
\[ + n^{-1/2} \hat{H}_2 r' \left( \hat{\mu} x M_2 - \hat{B} \right) \]
\[ =: D_{11} + D_{12} + D_{13}. \]

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
\[ D_{11} = n^{-1/2} \sum_{i} \psi_{1i} + o_p (1) \quad (A.13) \]
is analogous to that of Lemma B7 and B8 of IN02, except that I need to establish
\[ \| \hat{H}_1 - H_1 \| = o_p(1) \text{ and } \| \hat{H}_2 - H_2 \| = o_p(1). \]

To prove these two claims, notice that

\[ \| H_1 \| = O(1) \text{ and } \| H_2 \| = O(1), \]

because \( \| H_1 \|^2 \leq C A_1 A_1' / \Omega_2 \leq C \) and \( \| H_2 \|^2 = A_2 A_2' / \Omega_2 \leq C A_1 A_1' / \Omega_2 \leq C. \) In addition, \( \| \hat{P} - P \| = o_p(1), \| \hat{R} - I \| = o_p(1), \) and \( \| n^{-1} \sum_i r_i \hat{p}_i - \mathbb{E} r_i \hat{p}_i \| = o_p(1) \) by Lemma A3 of [202] and Newey (1997). I also know that \( \lambda_{\min}(P) \geq C > 0 \) and \( \| P \| = O(1) \) by Assumption 5. By Slutsky’s theorem, \( \| \hat{R} - I \| = o_p(1) \) and \( \| \hat{P} - P^{-1} \| = o_p(1). \) By the CS inequality and Lemma A3 of [202],

\[ \left\| n^{-1} \sum_i r_i (p_i - \hat{p}_i) \right\|^2 \leq n^{-1} \sum_i \| r_i \|^2 \times n^{-1} \sum_i \| p_i - \hat{p}_i \|^2 \]

\[ = O_p \left( M_3 \delta (m_2)^2 \Delta^2 \right) = o_p(1). \]

Therefore, by the triangle inequality with probability approaching one,

\[ \| \hat{H}_1 - H_1 \|^2 \]

\[ = \| FA_1 \hat{P}^{-1} - FA_1 P^{-1} \|^2 \]

\[ \leq 2 \| F (A_1 - A_1) \hat{P}^{-1} \|^2 + 2 \| FA_1 (\hat{P}^{-1} - P^{-1}) \|^2 \]

\[ = 2 \| F (r(x)' (I + o_p(1)) (\mathbb{E} r_i \hat{p}_i + o_p(1)) - r(x)' \mathbb{E} r_i \hat{p}_i) \hat{P}^{-1} \|^2 \]

\[ + 2 \| FA_1 P^{-1} (P - \hat{P}) \hat{P}^{-1} \|^2 \]

\[ \leq \| H_2 \|^2 o_p(1) + \| H_1 \|^2 o_p(1) = o_p(1), \]

and similarly \( \| \hat{H}_2 - H_2 \| = o_p(1), \) which establish (A.13).

Next, recall that by Assumption 7

\[ n^{-1} \left( \hat{G} - \hat{p} \hat{r} M_z \right)' \left( \hat{G} - \hat{p} \hat{r} M_z \right) = O_p \left( M_2^{-2d_i/j} \right), \]

and

\[ n^{-1} \left( \hat{p} \hat{r} M_z - \bar{B} \right)' \left( \hat{p} \hat{r} M_z - \bar{B} \right) = O_p \left( M_2^{-2d_i/j} \right). \]

Therefore, for \( D_{12}, \)

\[ \left| n^{-1/2} \hat{H}_1 \hat{P}' \left( \hat{G} - \hat{p} \hat{r} M_z \right) \right|^2 \leq n \left[ \hat{H}_1 \hat{P} \hat{H}_1' \right] \left[ n^{-1} \left( \hat{G} - \hat{p} \hat{r} M_z \right)' (\hat{G} - \hat{p} \hat{r} M_z) \right] \]

\[ \leq \| \hat{H}_1 \|^2 O_p \left( n M_2^{-2d_i/j} \right) = o_p(1). \] (A.14)
For $D_{13}$, similarly to (A.14),

$$
\left| n^{-1/2} \hat{H}_2 r' \left( \hat{\mu}_{M_2}^T - \hat{B} \right) \right|^2 \leq n \left[ \hat{H}_2 \hat{R} \hat{H}_2 \right] \left[ n^{-1} \left( \hat{\mu}_{M_2} - \hat{B} \right) ' \left( \hat{\mu}_{M_2} - \hat{B} \right) \right] \leq \left\| \hat{H}_2 \right\|^2 O_p \left( nM_2^{-2d_4/j_4} \right) = o_p \left( 1 \right).
$$

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

$$
\psi_{1i} = H_1 \left( p_i u_i - \overline{p}_i \right).
$$

To prove $\psi_{2i} = H_2 \psi_{II}^I$, first notice that by Taylor expansion,

$$
n^{1/2} F \left( \hat{a} \left( b_1, \hat{V} \right) - \hat{a} \left( b_1, V \right) \right)
= n^{-1/2} Fr \left( x \right) ' \hat{R}^{-1} r' \left( \hat{B} - B \right)
= \hat{H}_2 n^{-1/2} \sum_i r_i \left( b_1 (\hat{v}_i, w_i) - b_1 (v_i, w_i) \right)
= \hat{H}_2 n^{-1/2} \sum_i r_i b_{1,V} (v_i, w_i) (\hat{v}_i - v_i) + \hat{H}_2 n^{-1/2} \sum_i r_i b_{1,V} (\hat{v}_i, w_i) (\hat{v}_i - v_i)^2/2
=: D_{21} + D_{22},
$$

where $b_{1,V} (v_i, w_i) \equiv \partial b_1 (v, w) / \partial v^2 \big|_{v=v_i, w=w_i}$ and $\hat{v}_i$ lies between $v_i$ and $\hat{v}_i$. I prove $D_{21} = n^{-1/2} \sum_i H_2 \psi_{II}^I + o_p \left( 1 \right)$ and $D_{22} = o_p \left( 1 \right)$.

For $D_{21}$,

$$
D_{21} = \hat{H}_2 n^{-1/2} \sum_i r_i b_{1,V} (v_i, w_i) (\hat{v}_i - v_i)
= H_2 n^{-1/2} \sum_i r_i b_{1,V} (v_i, w_i) \Delta_i + \left( \hat{H}_2 - H_2 \right) n^{-1/2} \sum_i r_i b_{1,V} (v_i, w_i) (\hat{v}_i - v_i)
+ H_2 n^{-1/2} \sum_i r_i b_{1,V} (v_i, w_i) \left( \Delta^{II}_i + \Delta^{III}_i \right)
=: D_{211} + D_{212} + D_{213},
$$

where

$$
\delta_{ij} = F_{x_i|z_j,w_j} (x_i|z_j, w_j) - q_j ' \gamma_{M_1} (x_i), \quad \Delta_i^{I} = q_i ' \hat{Q}^{-1} n^{-1} \sum_j q_j v_{ij},
\Delta_i^{II} = q_i ' \hat{Q}^{-1} n^{-1} \sum_j q_j \delta_{ij}, \text{ and } \Delta_i^{III} = -\delta_{ii}.
$$
Following an argument identical to the proof of Lemma B7 of IN02,

\[ D_{211} = n^{-1/2} \sum_i H_2 \bar{\mu}_i^I + o_p(1). \]

For \( D_{212} \),

\[ |D_{212}|^2 \leq Cn \left[ (\hat{H}_2 - H_2) \hat{R} \left( \hat{H}_2 - H_2 \right)^T \right] \left[ n^{-1} \sum_i (\hat{v}_i - v_i)^2 \right] = o_p \left\{ n \left( n^{-1} \zeta (M_3^2 M_3) \Delta^2_{1n} \right) \right\} = o_p(1). \]

For \( D_{213} \),

\[ |D_{213}|^2 \leq Cn \left[ H_2 \hat{R} H_2^T \right] \left[ n^{-1} \sum_i \left( (\Delta_i^I)^2 + (\Delta_i^{II})^2 \right) \right] = o_p \left( n M_1^{1-2d_1/j_1} \right) = o_p(1), \]

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

Next, for \( D_{22} \),

\[ |D_{22}| \leq Cn^{1/2} \| \hat{H}_2 \| \sup_{x \in \mathcal{X}} \| r(x) \| \left| n^{-1} \sum_i (\hat{v}_i - v_i)^2 \right| = o_p \left( n^{1/2} \zeta (M_3) \Delta^2_{1n} \right) = o_p(1). \]

Combining the results for \( D_{21} \) and \( D_{22} \),

\[ n^{1/2} F \left( \hat{a} (b_1, \hat{V}) - a (b_1, V) \right) = n^{-1/2} \sum_i H_2 \bar{\mu}_i^I + o_p(1). \]

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

\[ n^{1/2} F \left( \hat{a} (b_1, V) - a (b_1, V) \right) = n^{-1/2} \sum_i \hat{H}_2 r_i b_1 (v_i, w_i) - n^{1/2} Fb(x) \]
\[ = n^{-1/2} \sum_i H_2 r_i (b_1 (v_i, w_i) - b(x_i)) + n^{-1/2} \sum_i (\hat{H}_2 - H_2) r_i (b_1 (v_i, w_i) - b(x_i)) \]
\[ + n^{-1/2} \sum_i \hat{H}_2 r_i \left( b(x_i) - r_i \rho^M \right) - n^{1/2} F \left( b(x) - r(x) \rho^M \right) \]
\[ =: D_{31} + D_{32} + D_{33} + D_{34}, \]
where I use the definition of \( \hat{H}_2 \) and \( \hat{R} \) 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 \( D_{32} \),

\[
\mathbb{E} \left[ |D_{32}|^2 \right] = n^{-1} \left( \hat{H}_2 - H_2 \right)^T r' \mathbb{E} \left[ \xi' \right] r \left( \hat{H}_2 - H_2 \right) \\
\leq C \left( \hat{H}_2 - H_2 \right)^T \hat{R} \left( \hat{H}_2 - H_2 \right) \\
\leq C \| \hat{H}_2 - H_2 \|^2 \left( 1 + \| \hat{R} - I \| \right) \\
= O_p \left( \| \hat{H}_2 - H_2 \|^2 \right) \\
= O_p \left( n^{-1} \zeta (M_3)^2 M_3 \right) = o_p(1),
\]

where the first inequality holds by Assumption 6(c) and the fact that \( \hat{H}_2 \) and \( r \) are functions of \( X \) only, the second equality holds by \( \| \hat{R} - I \| = o_p(1) \), 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,

\[
|D_{33}|^2 \leq n \left( \hat{H}_2 \hat{R} \hat{H}_2' \right) n^{-1} \sum_i \left( b(x_i) - r_i' \rho_{M_3} \right)^2 \\
= O_p \left( n M_3^{-2d_3/j_3} \right) = o_p(1),
\]

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

Finally, for \( D_{34} \),

\[
|D_{34}|^2 = nF^2 \left( b(x) - r(x)' \rho_{M_3} \right)^2 = O_p \left( n M_3^{-2d_3/j_3} \right) = o_p(1).
\]

Combining the results for \( D_{31}, D_{32}, D_{33}, \) and \( D_{34} \), I have

\[
n^{1/2} F \left( \hat{a} (b_1, V) - a (b_1, V) \right) = n^{-1/2} \sum_i H_2 r_i \xi_i + o_p(1).
\]

In sum, I have proved

\[
n^{1/2} F \left( \hat{a} (b_1, V) - a (b_1, V) \right) = n^{-1/2} \sum_i (\psi_{1i} + \psi_{2i} + \psi_{3i}) + o_p(1),
\]
where
\[ \psi_{1i} = H_1 \left( p_i u_i - \mu_i^t \right), \quad \psi_{2i} = H_2 \mu_i^{II}, \text{ and } \psi_{3i} = H_2 r_i \xi_i. \]

Furthermore, notice that
\[ H_1 p_i u_i \perp \left( H_1 \mu_i^t, H_2 \mu_i^{II}, H_2 r_i \xi_i \right) \]
because \( \mathbb{E}[u_i | X_i, V_i, W_i] = 0. \)

Let \( \Psi_{in} = n^{-1/2} (\psi_{1i} + \psi_{2i} + \psi_{3i}) \). It is clear that \( \mathbb{E} \Psi_{in} = 0 \) and \( \mathbb{V} (\Psi_{in}) = n^{-1}. \)

For any \( \epsilon > 0 \), under Assumptions 7 and 8,
\[
n \mathbb{E} \left[ \mathbb{I} \{ |\Psi_{in}| > \epsilon \} \right] \Psi_{in}^2 \leq n \mathbb{E} \left[ \mathbb{I} \{ |\Psi_{in}| > \epsilon \} \right] (\Psi_{in}/\epsilon)^4 \leq n \mathbb{E} \Psi_{in}^4 \leq C n^{-1} \mathbb{E} \left[ (H_1 p_i u_i)^4 + (H_1 \mu_i^t)^4 + (H_2 \mu_i^{II})^4 + (H_2 r_i \xi_i)^4 \right] \leq C n^{-1} \left( \zeta (M_2)^2 M_2 + \zeta (M_2)^4 \zeta (M_1)^4 M_1 + \zeta (M_3)^4 \zeta (M_1)^4 M_1 + \zeta (M_3)^2 M_3 \right) \to 0,
\]
where the last inequality holds by Lemma B5 of IN02. Then, by the Lindeberg–Feller CLT,
\[ n^{1/2} \Omega_2^{-1/2} \left( \tilde{a} \left( \hat{b}_1, \hat{V} \right) - a (b_1, V) \right) \xrightarrow{d} N (0, I), \]
where \( \tilde{a} (\hat{b}_1, \hat{V}) = \tilde{b} (x) \) and \( a (b_1, V) = b (x) \).

To construct a feasible confidence interval, one needs a consistent estimator of the covariance matrix \( \Omega_2 \). Define
\[
\hat{\Omega}_2 := \hat{\Omega}_{21} + \hat{\Omega}_{22}, \quad \hat{\Omega}_{21} := \hat{A}_1 \hat{P}^{-1} \left( n^{-1} \sum_i \hat{p}_i \hat{p}_i^t \right) \hat{P}^{-1} \hat{A}_1, \quad (A.15)
\]
\[
\hat{\Omega}_{22} := n^{-1} \sum_i \left( \hat{A}_1 \hat{P}^{-1} \hat{p}_i - r M_3 (x) \right) \left( \hat{p}_i^t + r_i \hat{\xi}_i \right) \left( \hat{A}_1 \hat{P}^{-1} \hat{p}_i - r M_3 (x) \right) \left( \hat{p}_i^t + r_i \hat{\xi}_i \right),
\]
\[ \hat{A}_1 := r M_3 (x) \hat{R}^{-1} \left( n^{-1} \sum_i r_i \hat{p}_i^t \right), \quad \hat{\xi}_i := \hat{b}_1 (v_i, w_i) - \hat{b} (x_i). \]

I show \( \hat{\Omega}_2 / \Omega_2 - 1 \xrightarrow{p} 0. \) Recall that
\[ \Omega_2 = \mathbb{E} \left( A_1 \hat{P}^{-1} p_i u_i \right)^2 + \mathbb{E} \left( A_1 \hat{P}^{-1} \mu_i^t - A_2 \left( \mu_i^{II} + r_i \xi_i \right) \right)^2 = \Omega_{21} + \Omega_{22} \]
and

\[ \Omega_2 = n^{-1} \sum_{i} \left( \hat{A}_1 \hat{P}^{-1} \hat{p}_i \hat{u}_i \right)^2 + n^{-1} \sum_{i} \left( \hat{A}_1 \hat{P}^{-1} \hat{p}_i' - \hat{A}_2 \hat{R}^{-1} \left( \hat{\mu}_i' + r_i \hat{\xi}_i \right) \right)^2 =: \Omega_{21} + \Omega_{22}. \]

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

For \( \Omega_{22} \), I need to show

\[ n^{-1} \sum_{i} \left( \hat{H}_1 \hat{\mu}_i' - H_1 \mu_i' \right)^2 = o_p(1), \]

\[ n^{-1} \sum_{i} \left( \hat{H}_2 \hat{\mu}_i'' - H_2 \mu_i'' \right)^2 = o_p(1), \] and

\[ n^{-1} \sum_{i} \left( \hat{H}_2 r_i \hat{\xi}_i - H_2 r_i \xi_i \right)^2 = o_p(1). \] (A.16)

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

\[ \hat{H}_2 r_i \hat{\xi}_i - H_2 r_i \xi_i \]

\[ = \hat{H}_2 r_i \left( \hat{\xi}_i - \xi_i \right) + \left( \hat{H}_2 - H_2 \right) r_i \xi_i \]

\[ = \hat{H}_2 r_i \left( \hat{b}_1 (\hat{v}_i, w_i) - \hat{b} (x_i) - b_1 (v_i, w_i) + b(x_i) \right) + \left( \hat{H}_2 - H_2 \right) r_i \xi_i \]

\[ = \hat{H}_2 r_i \left( \hat{b}_1 (\hat{v}_i, w_i) - b_1 (\hat{v}_i, w_i) \right) + \hat{H}_2 r_i \left( b_1 (\hat{v}_i, w_i) - b_1 (v_i, w_i) \right) \]

\[ + \hat{H}_2 r_i \left( b(x_i) - \hat{b} (x_i) \right) + \left( \hat{H}_2 - H_2 \right) r_i \xi_i \]

\[ =: D_{412} + D_{422} + D_{432} + D_{442}. \]

For \( D_{41} \),

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

\[ \leq C \zeta (M_3)^2 n^{-1} \sum_{i} \left[ \left( \hat{\mu}_i' \left( \hat{\alpha}_M^2 - \alpha_M^2 \right) \right) ^2 + \left( \hat{\mu}_i' \alpha_M^2 - b_1 (\hat{v}_i, w_i) \right) ^2 \right] \]

\[ = O_p \left( \zeta (M_3)^2 \Delta_{2n}^2 \right) = o_p(1), \] (A.17)
where the second inequality holds by \( \| \hat{H}_2 \| = O_p(1) \) and Assumption 8(a) and the first equality holds by (A.7).

For \( D_{42} \),

\[
n^{-1} \sum_i D_{42i}^2 \leq \left\| \hat{H}_2 \right\|^2 \sup_{x \in X} \| r(x) \|^2 n^{-1} \sum_i (b_1(\hat{v}_i, w_i) - b_1(v_i, w_i))^2 \leq C \zeta (M_3)^2 n^{-1} \sum_i (\hat{v}_i - v_i)^2 = O_p(\zeta (M_3)^2 \Delta_{1n}^2) = o_p(1),
\]

where the first equality holds by Lemma 1.

The proof of \( n^{-1} \sum D_{43i}^2 = o_p(1) \) is completely analogous to (A.17) and is omitted.

For \( D_{44} \),

\[
E \left[ n^{-1} \sum_i D_{44i}^2 \middle| X \right] = (\hat{H}_2 - H_2) n^{-1} \sum_i r_i r_i' E \left( \xi_i^2 \middle| X_i \right) (\hat{H}_2 - H_2)'
\leq C (\hat{H}_2 - H_2) \hat{R} (\hat{H}_2 - H_2)'
\leq C \left\| \hat{H}_2 - H_2 \right\|^2 = o_p(1),
\]

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

\[
n^{-1} \sum_i D_{44i}^2 = o_p(1).
\]

Combining the results for \( D_{41}, D_{42}, D_{43}, \) and \( D_{44} \), I obtain

\[
n^{-1} \sum_i \left( \hat{H}_2 r_i \hat{\xi}_i - H_2 r_i \xi_i \right)^2 = o_p(1).
\]

Therefore, by (A.16),

\[
n^{-1} \sum_i \left( (\hat{H}_1 \hat{\mu}_i - \hat{H}_2 \hat{\mu}_i^{II} - \hat{H}_2 r_i \hat{\xi}_i) - (H_1 \mu_i - H_2 \mu_i^{II} - H_2 r_i \xi_i) \right)^2
\leq 3n^{-1} \sum_i \left( \hat{H}_1 \hat{\mu}_i - H_1 \mu_i \right)^2 + 3n^{-1} \sum_i \left( \hat{H}_2 \hat{\mu}_i^{II} - H_2 \mu_i^{II} \right)^2
+ 3n^{-1} \sum_i \left( \hat{H}_2 r_i \hat{\xi}_i - H_2 r_i \xi_i \right)^2 = o_p(1).
\]
Since \( E \left( H_1 \mu_i^I - H_2 \mu_i^{II} - H_2 r_i \xi_i \right)^2 = \Omega_{22}/\Omega_2 \leq 1 \), by Markov’s inequality and Lemma B6 of IN02,
\[
\left| \hat{\Omega}_{22}/\Omega_2 - n^{-1} \sum_i \left( H_1 \mu_i^I - H_2 \mu_i^{II} - H_2 r_i \xi_i \right)^2 \right| = o_p(1).
\]
By the law of large numbers,
\[
\left| n^{-1} \sum_i \left( H_1 \mu_i^I - H_2 \mu_i^{II} - H_2 r_i \xi_i \right)^2 - \Omega_{22}/\Omega_2 \right| = o_p(1).
\]
Therefore, by the triangle inequality,
\[
\hat{\Omega}_{22}/\Omega_2 - \Omega_{22}/\Omega_2 = o_p(1).
\]
Combining the results for \( \hat{\Omega}_{21} \) and \( \hat{\Omega}_{22} \), I obtain
\[
\hat{\Omega}_2/\Omega_2 - 1 \overset{p}{\to} 0.
\]
For the last result of Theorem 3, first define
\[
\hat{\Omega}_3 := \hat{\Omega}_{31} + \hat{\Omega}_{32}, \quad \hat{\Omega}_3 := \hat{A}_3 \bar{\mu}_i \left( n^{-1} \sum_i \hat{p}_i \hat{v}_i^2 \right) \bar{\mu}_i^{-1} \hat{A}_3;
\]
\[
\hat{\Omega}_{32} := n^{-1} \sum_i \left( \hat{A}_3 \bar{\mu}_i^{-1} \bar{\mu}_i^{II} - \left( \hat{\mu}_i^I + \hat{\xi}_i \right) \right) \left( \hat{A}_3 \bar{\mu}_i^{-1} \bar{\mu}_i^{I} - \left( \hat{\mu}_i^{II} + \hat{\xi}_i \right) \right),
\]
\[
\hat{A}_3 := n^{-1} \sum_i \hat{p}_i, \text{ and } \hat{\mu}_i^{II} := n^{-1} \sum_j \hat{b}_{1,V} (\hat{v}_j, w_j) q_j^I \hat{Q}^{-1} \hat{q}_i \hat{v}_{ji}.
\]
The results for \( n^{1/2} \hat{\Omega}_3^{-1/2} \left( \hat{\theta} - \bar{\theta} \right) \overset{d}{\to} N(0, I) \) follows immediately by setting \( r(x) \equiv 1 \) in the proof above.

**B Sufficient Conditions for Assumption 2**

Proposition 1 (Sufficient Conditions for Assumption 2 based on Altonji and Matzkin (2005)). Suppose the following conditions hold:

(a) \( f_{\eta_i | A_i} = f_{\bar{\eta}_i | A_i} \) for any permutation \( \bar{\eta}_i \) 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) $f_{A_i|X_i,Z_i}$ is continuous in $(X_i,Z_i)$.

Then, Assumption 2 is satisfied.

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

$$f_{A_i|X_i,Z_i} = f_{A_i|\tilde{X}_i,\tilde{Z}_i},$$

where $(\tilde{X}_i,\tilde{Z}_i)$ is any permutation of $(X_i,Z_i)$ in time. It is worth emphasizing that (B.1), which will be proved in the following, is different from Assumption 2.3 of Altonji and Matzkin (2005).\(^{27}\) Without loss of generality, we prove (B.1) for $T = 2$ since any ordering of $\{1,...,T\}$ for finite $T$ can be achieved via a finite number of pairwise permutations. Then, we prove that one can construct $W_i$ such that Assumption 2 holds. For simplicity of notation, we 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,

$$f_{A_i,\eta_{it}}(a,h_1,h_2) = f_{A_i,\eta_{it}}(a,h_2,h_1).$$ \hspace{1cm} (B.2)

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

$$D_1 \equiv \begin{vmatrix} \frac{\partial A}{\partial X_1} & \frac{\partial A}{\partial X_2} & \frac{\partial A}{\partial X_3} \\ \frac{\partial g^{-1}(X_1,Z_1,A)}{\partial X_1} & \frac{\partial g^{-1}(X_1,Z_1,A)}{\partial X_2} & \frac{\partial g^{-1}(X_1,Z_1,A)}{\partial X_3} \\ \frac{\partial g^{-1}(X_2,Z_2,A)}{\partial X_1} & \frac{\partial g^{-1}(X_2,Z_2,A)}{\partial X_2} & \frac{\partial g^{-1}(X_2,Z_2,A)}{\partial X_3} \end{vmatrix} (X_1,X_2,Z_1,Z_2,A) = (x_1,x_2,z_1,z_2,a)$$

\(^{27}\)To be specific, Altonji and Matzkin (2005) assume an exchangeability condition involving both $A$ and $\eta_t$, i.e., $f_{A_i,\eta_{it}|X_i} = f_{A_i,\eta_{it}|\tilde{X}_i}$, which effectively rules out time-varying endogeneity through the random coefficients since it requires that the density of $\eta_{it}$ given $X_{it} = x_{it}$ is the same as that given $X_{it} = x_{is}$ for any $s \neq t$.  

60
\[
M = \begin{vmatrix}
\frac{\partial g^{-1}(X_1, Z_1, A)}{\partial X_1} & 0 & \frac{1}{\partial A} \\
0 & \frac{\partial g^{-1}(X_2, Z_2, A)}{\partial X_2} & 0 \\
0 & 0 & \frac{\partial g^{-1}(X_1, Z_1, A)}{\partial A}
\end{vmatrix}
\]
\[
= \left. \frac{\partial g^{-1}(X, Z, A)}{\partial X} \right|_{(X, Z, A) = (x_1, z_1, a)} \times \left. \frac{\partial g^{-1}(X, Z, A)}{\partial X} \right|_{(X, Z, A) = (x_2, z_2, a)}
\]

and similarly
\[
D_2 := \begin{vmatrix}
\frac{\partial g(Z_1, A, \eta_1)}{\partial A} & \frac{\partial g(Z_1, A, \eta_1)}{\partial \eta_1} & \frac{\partial g(Z_1, A, \eta_1)}{\partial \eta_2} \\
\frac{\partial g(Z_2, A, \eta_2)}{\partial A} & \frac{\partial g(Z_2, A, \eta_2)}{\partial \eta_1} & \frac{\partial g(Z_2, A, \eta_2)}{\partial \eta_2}
\end{vmatrix}
\]
\[
= \left. \frac{\partial g(Z, A, \eta)}{\partial \eta} \right|_{(Z, A, \eta) = (z_2, a, h_2, h_1)} \times \left. \frac{\partial g(Z, A, \eta)}{\partial \eta} \right|_{(Z, A, \eta) = (z_1, a, h_1)}
\]

Then,
\[
\begin{align*}
& f_{X_1, X_2, A|Z_1, Z_2}(x_1, x_2, a | z_1, z_2) \\
& = f_{A, \eta_1, \eta_2|Z_1, Z_2}(a, g^{-1}(x_1, z_1, a), g^{-1}(x_2, z_2, a) | z_1, z_2) | D_1 | \\
& = f_{A, \eta_1, \eta_2|Z_1, Z_2}(a, g^{-1}(x_2, z_2, a), g^{-1}(x_1, z_1, a) | z_2, z_1) | D_1 | \\
& = f_{X_1, X_2, A|Z_1, Z_2}(x_2, x_1, a | z_2, z_1) | D_2 D_1 | \\
& = f_{X_1, X_2, A|Z_1, Z_2}(x_2, x_1, a | z_2, z_1) , \tag{B.3}
\end{align*}
\]

where the first equality holds by the change of variables for \( \eta_1 \) and \( \eta_2 \), the second equality uses (B.2) 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 (B.3), I integrate \( A \) out in (B.3) to obtain
\[
\begin{align*}
f_{X_1, X_2|Z_1, Z_2}(x_1, x_2 | z_1, z_2) &= f_{X_1, X_2|Z_1, Z_2}(x_2, x_1 | z_2, z_1) , \tag{B.4}
\end{align*}
\]
which implies

\[
\begin{align*}
&f_{A|X_1, X_2, Z_1, Z_2}(a | x_1, x_2, z_1, z_2) \\
&\quad = f_{X_1, X_2, A|Z_1, Z_2}(x_1, x_2, a | z_1, z_2) / f_{X_1, X_2|Z_1, Z_2}(x_1, x_2 | z_1, z_2) \\
&\quad = f_{X_1, X_2, A|Z_1, Z_2}(x_2, x_1, a | z_2, z_1) / f_{X_1, X_2|Z_1, Z_2}(x_2, x_1 | z_2, z_1) \\
&\quad = f_{A|X_1, X_2, Z_1, Z_2}(a | x_2, x_1, z_2, z_1),
\end{align*}
\]

(B.5)

where the second equality uses (B.3) and (B.4).

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 \(f_{A|X_1, X_2, Z_1, Z_2}(a | x_1, x_2, z_1, z_2)\) can be approximated arbitrarily closely by a function of the form \(f_{A|W}(a | w)\), where \(W\) is a vector-valued function symmetric in the paired elements of \((X, Z)\). See footnote 10 for the detailed description of \(W\). 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,

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

(B.6)

Let

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

denote the simple average of \(f_{A|X_1, X_2, Z_1, Z_2}(a | x_1, x_2, z_1, z_2)\) over all \(T!\) (here \(T = 2\)) unique permutations of \((x_t, z_t)\), and similarly for \(\overline{f}^w_{A|X_1, X_2, Z_1, Z_2}(a | x_1, x_2, z_1, z_2)\). By (B.5),

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

By (B.5), (B.6), and the triangle inequality,

\[
\begin{align*}
&\left| f_{A|X_1, X_2, Z_1, Z_2}(a | x_1, x_2, z_1, z_2) - \overline{f}^w_{A|X_1, X_2, Z_1, Z_2}(a | x_1, x_2, z_1, z_2) \right| \\
&\quad = \left| \overline{f}_{A|X_1, X_2, Z_1, Z_2}(a | x_1, x_2, z_1, z_2) - \overline{f}^w_{A|X_1, X_2, Z_1, Z_2}(a | x_1, x_2, z_1, z_2) \right|
\end{align*}
\]

62
\[ \leq T! \times (\delta/T!) = \delta. \] (B.7)

Since \( f^w \) can be chosen to make \( \delta \) arbitrarily small, equation (B.7) implies that \( f_{A[X_1,X_2,Z_1,Z_2]} (a|x_1, x_2, z_1, z_2) \) can be approximated arbitrarily closely by a polynomial \( \mathcal{F}^w \) 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, \( \mathcal{F}^w \) can be written as a polynomial function of the elementary symmetric functions \( W(\cdot) \) of \((x_1, z_1), (x_2, z_2)\). We denote this function by \( f_{A|W} (a|w) \) and obtain that \( f_{A[X_1,X_2,Z_1,Z_2]} (a|x_1, x_2, z_1, z_2) \) can be approximated arbitrarily closely by \( f_{A|W} (a|W) \). Let \( \delta \to 0 \) in (B.7).\(^{28}\) Then, for any \( t \in \{1, ..., T\} \) and on-support \((x_t, z_t, a, w)\)

\[ f_{A[X_t,Z_t,W]} (a|x_t, z_t, w) = f_{A|W} (a|w). \]

\[ \square \]

\section{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 function applications is first provided. Then, I present the baseline results of estimating the APE \( \bar{b} := \mathbb{E} \beta_{it} \). Next, I show how the method performs for estimating the LAR \( b_t (x) := \mathbb{E} [\beta_{it}|X_{it} = x] \). Finally, as robustness checks, I (i) vary the number of firms and periods, (ii) change the order of basis functions used for series estimation, and (iii) include ex-post shocks \( \epsilon_{it} \) and \( \nu_{it} \) to the DGP.

\subsection{C.1 DGP}

The baseline revenue-based DGP is

\[ Y_{it} = k_{it}\beta_{it}^K + l_{it}\beta_{it}^L + \omega_{it} + \epsilon_{it}, \]

\(^{28}\)Note that to let \( \delta \to 0 \) we need the order of polynomials in the approximating function \( f^w \) to increase. However, it does not affect the order of the elements of \( W \) since by the fundamental theorem of symmetric functions, the elements of \( W \) has a fixed order of \( T \). The key is we have two different orders of polynomials; one is the order of the approximating functions \( f^w \), the other is the order of the arguments of \( W \). For example, one may let the order \( P_1 \) 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 order of its symmetric arguments (in this example, it is \( x+y \) and \( xy \)) fixed. See footnote 9 and 10 of Altonji and Matzkin (2005) for a more detailed discussion.
where \( \omega_{it}, \beta^K_{it}, \) and \( \beta^L_{it} \) are all functions of \((A_i, \epsilon_{it}, \upsilon_{it})\). \( k_{it} \) 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. Denote \( \mathcal{U}[a, b] \) to be the uniform distribution over the interval of \([a, b]\). I draw \( A_i \sim \text{i.i.d.} \mathcal{U}[1, 2] \) and \( \epsilon_{it} \sim \text{i.i.d.} \mathcal{U}[1, 2] \). I let true \( \omega_{it} = \ln (A_i + \epsilon_{it}/2 + 1) \), \( \beta^K_{it} = (A_i + \epsilon_{it})/10 \), and \( \beta^L_{it} = (A_i + \epsilon_{it})/10 \), and write \( \beta_{it} \equiv (\omega_{it}, \beta^K_{it}, \beta^L_{it}) \). I compute the true \( \bar{w} := \mathbb{E}\omega_{it} = 1.1736 \), \( \bar{B}^K := \mathbb{E}\beta^K_{it} = .3 \), and \( \bar{B}^L := \mathbb{E}\beta^L_{it} = .3 \). 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 \( \epsilon_{it} = v_{it} = 0 \) 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 \( \epsilon_{it} \) with noise \( \lambda^K_{it} \) and \( \lambda^L_{it} \):

\[
\eta^K_{it} = \epsilon_{it} + \lambda^K_{it} \quad \text{and} \quad \eta^L_{it} = \epsilon_{it} + \lambda^L_{it},
\]

where \( \lambda^K_{it} \) and \( \lambda^L_{it} \sim \text{i.i.d.} \mathcal{U}[-.05, .05] \). 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, and solve the firm’s profit maximization problem to obtain

\[
k_{it} = \frac{r_{it} - (A_i + \eta^K_{it})(r_{it} - w_{it})/10 - \ln \left( \frac{(A_i + \eta^K_{it})}{10} \right) - \ln \left( \frac{A_i + \eta^K_{it}}{2 + 1} \right)}{\ln \left( \frac{A_i + \eta^K_{it}}{5 - 1} \right)},
\]

\[
l_{it} = \frac{w_{it} - (A_i + \eta^L_{it})(w_{it} - r_{it})/10 - \ln \left( \frac{(A_i + \eta^L_{it})}{10} \right) - \ln \left( \frac{A_i + \eta^L_{it}}{2 + 1} \right)}{\ln \left( \frac{A_i + \eta^L_{it}}{5 - 1} \right)}.
\]

Let \( X_{it} = (k_{it}, l_{it})' \). It is clear that \( X_{it} \) is correlated with \( \beta_{it} \) in each period via \((A_i, \eta^K_{it}, \eta^L_{it})\) in a nonseparable way. I use \( N, T, \) and \( M \) to denote the total number of firms, periods, and simulations, respectively. The observed data is \( \{X_{it,m}, Y_{it,m}, Z_{it,m}\} \) for \( i = 1, ..., n, \ t = 1, ..., T, \) and \( m = 1, ..., M, \) which are used to construct \( \tilde{b}_m \) and \( \tilde{b}_t(x) \) to estimate \( \bar{b} \) and \( b_t(x) \), respectively, via the estimation procedure outlined in Section 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.
Table 3: Performance of $\hat{b}$

| Formula      | $\hat{\omega}$ | $\hat{b}_K$ | $\hat{b}_L$ |
|--------------|-----------------|--------------|--------------|
| Bias         | $M^{-1} \sum_m \left( \frac{\hat{b}_m^{(d)} - \bar{b}^{(d)}}{\bar{b}^{(d)}} \right)$ | 2.81% | 3.22% | 3.24% |
| rMSE         | $\sqrt{M^{-1} \sum_m \left( \frac{\hat{b}_m^{(d)} - \bar{b}^{(d)}}{\bar{b}^{(d)}} \right)^2}$ | 2.89% | 3.65% | 3.62% |
| MND          | $M^{-1} \sum_m \left( \frac{\hat{b}_m^{(d)} - \bar{b}^{(d)}}{\bar{b}^{(d)}} \right)$ | 2.81% | 3.26% | 3.26% |

C.2 Results

For the baseline results, I set $N = 1,000$ and $T = 2$, and use a second-order 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 $V_{it} := (F_{it|Z_{it},W_i}(k_{it}|Z_{it},W_i), F_{it|Z_{it},W_i}(l_{it}|Z_{it},W_i))'$. The performance of $\hat{\omega}, \hat{b}_K,$ and $\hat{b}_L$ is summarized in Table 3. 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 3 reports the normalized bias of each coordinate of $\hat{b}$. 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 $b$. The second row shows the normalized rMSE of each coordinate of $\hat{b}$. My method achieves low normalized rMSEs between 2.89% and 3.65% of the size of the corresponding coordinate $b^{(d)}$ of $b$. The last row presents the normalized MNDs of each coordinate of $\hat{b}$. Again, the method performs well with an MND between 2.81% and 3.26% of the size of the corresponding coordinate $b^{(d)}$ of $b$.

Next, I investigate the performance of $\hat{b}_{t,m}(x)$, which is obtained by regressing the estimated $\hat{b}_{1t,m}(\hat{V}_{it,m}, W_{i,m})$ on the second-degree polynomial spline basis of $X_{it,m}$ with its knot at the median for each $(t, m)$ combination. 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 $\beta_{it,m}$ on the same spline basis functions of $X_{it,m}$.

Table 4 presents the results. Note that by definition the bias of $\hat{b}_t(x)$ is the same
Table 4: Performance of \( \tilde{b}_t(x) \)

| Formula | \( \tilde{\omega}_t(x) \) | \( \tilde{b}_t^K(x) \) | \( \tilde{b}_t^L(x) \) |
|---------|-----------------|-----------------|-----------------|
| rMSE    | \( \sqrt{(NTM)^{-1} \sum_{i,t,m} \left( \hat{b}_{t,m}^{(d)}(x_{it,m}) - b_t^{(d)}(x_{it,m}) \right)^2 / \| \tilde{b}^{(d)} \| } \) | 3.36% | 5.57% | 5.59% |
| MND     | \( (NTM)^{-1} \sum_{i,t,m} \left[ \left| \hat{b}_{t,m}^{(d)}(x_{it,m}) - b_t^{(d)}(x_{it,m}) \right| / \| \tilde{b}^{(d)} \| \right) \) | 2.87% | 4.36% | 4.37% |

as \( \tilde{b} \) so I omit it here. The normalized rMSE of \( \tilde{b}_t(x) \) is bigger than that of \( \tilde{b} \), with a magnitude between 3.36% and 5.59% of the size of the corresponding \( \tilde{b}^{(d)} \). The normalized MND follows a similar pattern. The performance of \( \tilde{b}_t(x) \) for estimating \( b_t(x) \) is not as good as that of \( \hat{b}_t^{(d)} \), 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 \( \tilde{b} \) and \( b_t(x) \), I conduct another set of exercises. I evaluate the performance of my estimator using rMSE defined as \( \sqrt{M^{-1} \sum_m \| \hat{b}_m - \tilde{b} \|^2 / \| \tilde{b} \| } \) for the whole vector of \( \tilde{b} \), and \( \sqrt{(NTM)^{-1} \sum_{i,t,m} \| \hat{b}_{t,m}(x_{it,m}) - b_t(x_{it,m}) \|^2 / \| \tilde{b} \| } \) for the whole vector of \( b_t(x) \). First, I vary \( N \) and \( T \), and present the results in Table 5. As expected, a larger \( N \) is good for overall performance. However, the magnitude in the improvement of performance is mild, possibly caused by the fact that 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 of my method, 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 5: Performance under Varying \( N \) and \( T \)

| \( N \) (\( T = 2 \)) | rMSE: \( \tilde{b} \) | rMSE: \( b_t(x) \) | \( T \) (\( N = 1,000 \)) | rMSE: \( \tilde{b} \) | 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 basis functions used to construct the series
estimators for all steps, and present the results in Table 6. The knot is placed at the median for all cases. 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. I attribute it to the fact that there are many more regressors in each step of regression when $d = 3$, leading to a possible over-fitting problem. Based on this result, I use the second-degree splines in the empirical application.\footnote{One may also use the AIC criterion to select the degree of basis functions.}

\begin{table}[h]
\centering
\begin{tabular}{lcc}
\hline
Degree of Basis Functions & rMSE: $\theta$ & rMSE: $b_t(x)$ \\
\hline
1 & 5.27\% & 6.41\% \\
2 & 2.99\% & 3.69\% \\
3 & 4.82\% & 6.40\% \\
\hline
\end{tabular}
\caption{Performance under Varying Degree of Basis Functions}
\end{table}

Lastly, I examine how including the ex-post shocks $\epsilon_{it}$ and $\upsilon_{it}$ into the model affects the performance of my estimators. I draw $\epsilon_{it} \sim U[-.25,.25]$, the ex-post shock to the main equation (2.1), independently of all the other variables. One may also interpret $\epsilon_{it}$ as an ex-post shock to $\omega_{it}$. For $\upsilon_{it}$, since ex-post shock to $\omega_{it}$ has been considered by $\epsilon_{it}$, I draw $\upsilon_{it}^s \sim U[-.1,.1]$ 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 7. Adding $\epsilon_{it}$ or $\upsilon_{it}$ negatively affects the performance of the proposed estimator. When $\epsilon_{it}$ is included, the rMSE of $\hat{\theta}$ for estimating $\theta$ increases from 2.99\% to 3.64\% and the rMSE of $\hat{b}_t(x)$ for estimating $b_t(x)$ rises from 3.64\% to 5.43\%. The effect of including $\upsilon_{it}$ on the performance is similar. Overall, these results show that my method is relatively robust against the inclusion of ex-post shocks.

\begin{table}[h]
\centering
\begin{tabular}{lcrr}
\hline
Add $\epsilon_{it}$ to $Y_{it}$? & rMSE: $\theta$ & rMSE: $b_t(x)$ \\
\hline
No & 2.99\% & 3.69\% \\
Yes & 3.64\% & 5.43\% \\
\hline
Add $\upsilon_{it}$ to $\left(\beta_{K_{it}}^K, \beta_{L_{it}}^L\right)$? & rMSE: $\theta$ & rMSE: $b_t(x)$ \\
\hline
No & 2.99\% & 3.69\% \\
Yes & 3.57\% & 5.19\% \\
\hline
\end{tabular}
\caption{Performance with and without Ex Post Shocks}
\end{table}