Nonparametric Instrumental Variables Identification and Estimation of Nonseparable Panel Models

Bradley Setzler*

December 8, 2016

Abstract

This paper considers identification and estimation of ceteris paribus effects of continuous regressors in nonseparable panel models with instrumental variables. From the insight that the passing of time creates a triangular structure in which shocks realized in the future are excluded at present, a novel recursive control function is shown to control for persistence in the unobservables. The recursive control function achieves nonparametric identification for any number of time periods and endogenous regressors without restrictions on the joint dependence of unobservables over time or time homogeneity, but requires that the instruments satisfy a full support condition. Then, it is shown that a semiparametric version of the recursive control function identifies the model under semiparametric shape restrictions even if the instruments have small support. Semiparametric estimators based on quantile regression are introduced that do not suffer a curse of dimensionality and perform well in small samples. In a novel empirical application to the population of matched firms and workers in Norway, the ceteris paribus elasticities of firm production with respect to capital and labor are estimated without separability of productivity shocks from capital or labor.

*PhD Candidate, Department of Economics, University of Chicago, bradley.setzler@gmail.com. I acknowledge the support of the NSF Graduate Research Fellowship and the NIH National Institute on Aging Predoctoral Fellowship. I am grateful to Stéphane Bonhomme, Magne Mogstad, Azeem Shaikh, and Alex Torgovitsky for valuable input and guidance.

JEL classification: C14, C33, C36, C51

Keywords: Control variables, Instrumental variables, Nonseparable model, Nonparametric identification, Panel data, Production function

Contents

1 Introduction **3**

2 Illustration of the Identification Approach **5**

2.1 Motivation from Heterogeneous Firm Productivity 5

2.2 Illustration of the Identification Approach 8

2.3 Discussion and Preview of General Results 10

3 Nonparametric Identification **11**

3.1 Triangular Panel Model 11

3.2 Identifying Assumptions 13

3.3 Identification Results 15

4 Semiparametric Identification and Estimation **16**

4.1 Random Coefficient and Markov Chain Assumptions 16

4.2 Identification with Small-support Instruments 19

4.3 Practical Estimators and Monte Carlo Study 22

5 Data and Empirical Results **25**

5.1 Norwegian Tax Records on Firms and Workers 25

5.2 Empirical Results 25

6 Conclusions **26**

References **27**

A Appendix **30**

A.1 Proof of Theorem 1 30

A.2 Proof of Theorem 2 31

A.3 Proof of Theorem 3 33

A.4 Proof of Theorem 4 35

1 Introduction

This paper is concerned with the identification of the *ceteris paribus* effect of an input X_t on an output Y_t at time t. Such an effect is challenging to identify when unobservable determinants of Y_t depend nonseparably on unobservable determinants of X_t. An existing literature achieves identification of this effect by assuming that the distribution of nonseparable unobservables and their relationship to X_t and Y_t are constant over time, which Chernozhukov et al. (2013, 2015) refer to as "time being randomly assigned" or *time-homogeneity*. This assumption is satisfied when all dependence over time (persistence) in the nonseparable unobservables is due to time-invariant unobservables (fixed effects), which I will refer to as the *fixed effects assumption*. The fixed effects assumption is central to identification results in nonseparable panel models by Altonji and Matzkin (2005), Bester and Hansen (2012), Graham and Powell (2012), and Chernozhukov et al. (2013, 2015). However, the fixed effects assumption rules out many common persistence structures, including first-order Markov chains (e.g., the random walk) and, more generally, k^{th}-order Markov chains (e.g., the k^{th}-order autoregressive process). This literature also rules out impulse responses, that is, *ceteris paribus* effects of X_t or the unobservables at t on $Y_{t'}, t' > t$.

By contrast, this paper considers the case in which the relationship between X_t, Y_t, and the unobservables changes over time: "new" unobservables arrive at t, and both X_t and Y_t depend on new as well as "lagged" unobservables which arrived prior to t. New and lagged unobservables have unrestricted dependence, so persistence may be due to fixed effects, random walks, k^{th}-order autoregressions, or any other joint distribution of unobservables over time. Furthermore, X_t and unobservables at t may affect $Y_{t'}, t' > t$, allowing for impulse responses. Any of these relationships may vary with time without restrictions. The model may grow in complexity over time, as more observables and unobservables accumulate.

While the weak restrictions considered in this paper violate time-homogeneity in general and the fixed effects assumption in particular, this paper achieves identification using the *triangularity* generated by time: unobservables that arrive after t are excluded from the relationship between X_t and Y_t. Results from the literature on the identification of triangular models in cross-sectional contexts using instrumental variables, especially results by Imbens and Newey (2009), are extended to identify the triangular panel model using a recursive control function introduced here that controls for the persistence in unobservables over time. Identification is achieved for any dimension of inputs X_t and any number of time periods, as the approach simplifies to existing cross-sectional identification results when there is only one time period. The outcome models identified in this paper include those identified in the existing literature cited above, as shown explicitly through examples in the text. However, like the related cross-sectional literature, my approach requires assumptions on the input equations, such as monotonic responses and the availability of instrumental variables, which are not required in the literature that imposes the fixed effects assumption. My identification approach also requires observing the beginning time period of the process, so that there exists an observed time period at which there are no lagged unobservables to consider.

Identification from the nonparametric recursive control function introduced here places strong demands on the data, as it requires that the initial time period be observed and instrumental variables that satisfy a full support condition be included in the data. Any corresponding estimator that controls for all lagged unobservables will necessarily suffer a curse of dimensionality, as the number of lagged unobservables grows with time. As a practical solution, I show that a semiparametric version of the control function identifies the nonseparable panel model under shape restrictions that maintain the qualitative features of interest even when the instruments have small support and without a curse of dimensionality. I do so by extending to a panel context the cross-sectional insight of Ma and Koenker

(2006), Jun (2009), and, especially, Masten and Torgovitsky (2016), and by imposing that unobservables satisfy the k^{th}-order Markov assumption, where k is finite. In small-sample Monte Carlo studies with binary instruments, it is shown that the estimator suffers little bias, while other potential estimators are shown to be severely biased and the biases become worse over time. Note that, in both the nonparametric and semiparametric approaches, identification is proven separately for the case in which lagged unobservables in the outcome equation also enter the input equation, and for the case in which measurement error in the output equation is of infinite dimension.

In an empirical application, I study the productivity of capital and labor in the population of firms in Norway using a 12-year panel of administrative balance sheet data matched to tax records for the population of workers in Norway. The existing literature on firm productivity under endogeneity assumes that the elasticities of firm production with respect to capital and labor are constant across firms – a property inherent to the traditional Cobb-Douglas specification of the firm production function (Marschak and Andrews, 1944). I relax this assumption by allowing for these elasticities to depend nonseparably on unobserved firm productivity, so that firm productivity shocks may be labor-augmenting, capital-augmenting, or both. Like Olley and Pakes (1996) and the related literature, I control flexibly for unobserved productivity using monotonicity of the input equations. Unlike Olley and Pakes (1996) and the related literature, I permit nonseparability of productivity shocks, allow for productivity shocks to have greater than first-order Markov dependence, and allow that the firm imperfectly observes its own productivity shock.

2 Illustration of the Identification Approach

2.1 Motivation from Heterogeneous Firm Productivity

To introduce notation, motivate the relaxation of time-homogeneity, and introduce the empirical context, consider the model of Olley and Pakes (1996), but adding heterogeneity to the production function as well as the information set. For simplicity, ignore the depreciation of capital, the exit option of firms, and exogenous covariates. Denote Y_t as revenues, X_t as the capital stock, and U_t as unobserved productivity, which is assumed to be continuously-distributed and scalar. The firm's production function is,

$$Y_t = g_t\left(X_t, U_t, \epsilon_t\right). \tag{2.1}$$

where ϵ_t denotes *iid* measurement error, which may be vector-valued. Combining equations (1) and (3) from Olley and Pakes (1996), the profit-maximizing firm solves,

$$\mathcal{V}_t\left(X_t, U_t\right) = \max_{X_{t+1}} \mathbb{E}\left[Y_t - C_t\left(X_{t+1} - X_t, Z_t\right) + \beta \mathcal{V}_{t+1}\left(X_{t+1}, U_{t+1}\right) \middle| \mathcal{J}_t\right], \tag{2.2}$$

where C_t is the cost function, Z_t is the price of capital, β is the discount rate, and $\mathcal{J}_t$ is the information set. The solution to the firm's problem is,

$$X_{t+1} = h_t\left(X_t, Z_t, \mathcal{J}_t\right). \tag{2.3}$$

which is equation (5) from Olley and Pakes (1996) but with a general information set.

Consider three cases to motivate the identification challenge addressed by this paper that cannot be addressed by existing approaches. For the first case, suppose U_t follows a first order Markov chain and $U_t \in \mathcal{J}_t$. Then, (2.1) and (2.3) have the form,

$$\begin{aligned} Y_t &= g_t\left(X_t, U_t, \epsilon_t\right), \\ X_t &= h_{t-1}\left(X_{t-1}, Z_{t-1}, U_{t-1}\right), \end{aligned} \tag{2.4}$$

which is equivalent to the model of Olley and Pakes (1996) when g_t has the Cobb-Douglas specification and Z_t is constant across firms. The identification approach of Olley and Pakes (1996), Levinsohn and Petrin (2003), and a related literature uses the insight that the only unobservable in the output equation that may depend on X_t, U_t, also appears in the input equation for X_{t+1}. This motivates the isolation of U_t from an input equation so that it can be controlled as if it were an observable when estimating the output equation. In practice, the equations for optimal investment or material expenditure at t are often used instead of X_{t+1} to isolate U_t; see the survey by Ackerberg et al. (2015).

For the second case, suppose $U_t \notin \mathcal{J}_t$ but $V_t \in \mathcal{J}_{t-1}$, where V_t is a signal about U_t. For example, $V_t = U_t + \zeta_t$, where ζ_t is an *iid* measurement error. Then, (2.1) and (2.3) become,

$$\begin{aligned} Y_t &= g_t\left(X_t, U_t, \epsilon_t\right), \\ X_t &= h_{t-1}\left(X_{t-1}, Z_{t-1}, V_t\right). \end{aligned} \tag{2.5}$$

This differs from the model of Olley and Pakes (1996) in that the unobservable in the output equation that may depend on X_t, U_t, is not identical to the unobservable that can be isolated from the input equation for X_{t+1}, V_{t+1}, so the identification approach from this literature does not apply. Instead, (2.5) is the triangular model developed in a cross-sectional context by Chesher (2003) and Imbens and Newey (2009), where the price of capital, Z_t, is now interpreted as an excluded instrument. Briefly, if h_{t-1} is monotonic in V_t and $Z_{t-1} \perp\!\!\!\perp (V_t, U_t)$, then (2.5) satisfies the key identifying assumptions of Imbens and Newey (2009) when Z_{t-1} also satisfies a full-support condition, and those of Torgovitsky (2015) when ϵ_t is excluded from g_t and Z_t varies at each value in the support of X_t.

For the final case, consider again the second case, but allowing the process U_t to be a Markov chain of unrestricted order ($k = \infty$). The firm's solution is,

$$\begin{aligned} Y_t &= g_t\left(X_t, U_t, \epsilon_t\right), \\ X_t &= h_{t-1}\left(X_{t-1}, Z_{t-1}, V_1, V_2, \ldots, V_t\right), \end{aligned} \tag{2.6}$$

where the solution includes all lags of V_t because, e.g., V_1 is a predictor of U_1 and U_1 is a predictor of U_t when $k = \infty$. (2.6) differs from the model of Chesher (2003) and Imbens and Newey (2009) in that there are multiple unobservables in the input equation for X_t. If U_t were a fixed effect, i.e., $U_t = \overline{U}, \forall t$, then the identification approach of Chernozhukov et al. (2013, 2015) would apply. Otherwise, this model has not previously been identified without separability, and is identified in this paper. This paper also identifies more general cases that are difficult to motivate in the firm production model but may be relevant in other contexts, e.g., allowing V_t, U_t, and X_t to appear in the equation for $Y_{t'}$, $t' > t$.

2.2 Illustration of the Identification Approach

Here, I consider a correlated random coefficients (Heckman and Vytlacil, 1998) specification of the firm productivity model in (2.6). I show identification in two time periods with a binary instrument to illustrate the key insight of this paper in a simplified context. Omitting the measurement error ϵ_t for simplicity, a random coefficients specification of the firm production function in (2.1) is,

$$Y_t = A_t(U_t) + B_t(U_t)X_t, \tag{2.7}$$

where $A_t = U_t + \alpha_t$, $B_t = U_t + \beta_t$, U_t is an unobserved random variable normalized to have mean zero, and $\alpha_t > 0, \beta_t > 0$ are constants. The parameter of interest is the average marginal effect of X_t on Y_t, defined as $\mathbb{E}[B_t]$, which equals β_t since U_t has mean zero.

If $X_t \perp\!\!\!\perp U_t$, then linear regression of Y_t on a constant and X_t identifies β_t as the slope on X_t (Amemiya, 1985). Formally, if $X_t \perp\!\!\!\perp U_t$, the OLS estimator of β_t has probability limit,

$$\frac{\text{Cov}\,(Y_t, X_t)}{\text{Var}\,(X_t)} = \frac{\text{Cov}\,(U_t + \alpha_t + (U_t + \beta_t)X_t, X_t)}{\text{Var}\,(X_t)} = \beta_t, \tag{2.8}$$

where the second equality uses that $\text{Cov}\,(U_t, X_t) = 0$ and the third equality uses that,

$$\text{Cov}\,((U_t + \beta_t)X_t, X_t) = \mathbb{E}[(U_t + \beta_t)X_t^2] - \mathbb{E}[(U_t + \beta_t)X_t]\mathbb{E}[X_t] = \beta_t \text{Var}(X_t), \tag{2.9}$$

since $\mathbb{E}[U_t X_t] = 0$ and $\mathbb{E}[U_t X_t^2] = 0$. The identification challenge is that X_t depends on U_t. However, suppose there exists a control function R_t satisfying the property that $X_t \perp\!\!\!\perp U_t | R_t = r_t$. Then, if $\text{Var}(X_t | R_t = r_t) > 0$, OLS local to R_t identifies β_t, as,

$$\frac{\text{Cov}\,(Y_t, X_t | R_t = r_t)}{\text{Var}\,(X_t | R_t = r_t)} = \frac{\text{Cov}\,(U_t + \alpha_t + (U_t + \beta_t)X_t, X_t | R_t = r_t)}{\text{Var}\,(X_t | R_t = r_t)} = \beta_t, \tag{2.10}$$

since $\text{Cov}\,(U_t, X_t | R_t = r_t) = 0$, $\mathbb{E}[U_t X_t | R_t = r_t] = 0$, and $\mathbb{E}[U_t X_t^2 | R_t = r_t] = 0$.

In order to recover a control function R_t satisfying $X_t \perp\!\!\!\perp U_t | R_t = r_t$, consider the equations that determine X_t. Consider $Z_t \in \{0, 1\}$, and let V_1 and V_2 have any continuous joint distribution. A random coefficients specification of (2.6) at $t = 1$ is,

$$X_1 = \tau_1(V_1) + \theta_1(V_1)Z_0 \tag{2.11}$$

where $\tau_1(V_1) = V_1 + \overline{\tau}_1$, $\gamma_1(V_1) = V_1 + \overline{\gamma}_1$, and $\theta_1(V_1) = V_1 + \overline{\theta}_1$, and a random coefficients specification of (2.6) at $t = 2$ is,

$$X_2 = \tau_2(V_1, V_2) + \gamma_2(V_1, V_2)X_1 + \theta_2(V_1, V_2)Z_1. \tag{2.12}$$

where $\tau_2(V_2) = V_2 + \overline{\tau}_2 + \psi_2(V_1)$, $\gamma_2(V_2) = V_1 + V_2 + \overline{\gamma}_2$, and $\theta_2(V_2) = V_1 + V_2 + \overline{\theta}_2$.

To identify β_1, notice that the *rank* of X_1 given Z_0 is equal to the *rank* of V_1 given Z_0 because X_1 is a strictly increasing function of V_1 given Z_0. Formally, denoting $F_{A|B}(a|b) \equiv \Pr(A \leq a|B = b)$ for any random variables A and B, the rank of X_1 given Z_0 is defined by $R_1 \equiv F_{X_1|Z_0}(X_1|Z_0)$. The monotonicity of the input equation in V_1 implies $R_1 = F_{V_1|Z_0}(V_1|Z_0)$. Suppose also that $Z_0 \perp\!\!\!\perp (V_1, U_1)|X_0$. Then, the conditioning on Z_0 can be relaxed, so that $R_1 = F_{V_1}(V_1)$. Thus, the data (X_1, Z_0) is sufficient information to recover the unconditional rank of V_1. Notice that V_1 being continuously-distributed implies that $F_{V_1}^{-1}$ exists. Fixing $R_1 = r_1$, there exists v_1 such that $v_1 = F_{V_1}^{-1}(r_1)$. Finally, $X_1 \perp\!\!\!\perp U_1|V_1 = v_1$ because all dependence between X_1 and U_1 is due to V_1, and conditioning on $R_1 = r_1$ is equivalent to conditioning on $V_1 = v_1$, so $X_1 \perp\!\!\!\perp U_1|R_1 = r_1$, and β_1 is identified by (2.10), subject to the support condition that $\text{Var}(X_1|R_1 = r_1) > 0$, $\forall r_1 \in [0, 1]$. See Masten and Torgovitsky (2016) for additional details about and extensions to this result.

The previous result is known, while the following result is, to my knowledge, novel. In order to identify β_2, notice that X_2 is a monotonic function of V_2 given not only the observables X_1 and Z_1, but also V_1, so that $F_{X_2|V_1,X_1,Z_2}(X_2|V_1, X_1, Z_2) = F_{V_2|V_1,X_1,Z_2}(V_2|V_1, X_1, Z_2)$. While V_1 is unobservable, it was shown above that conditioning on $V_1 = v_1$ is equivalent to conditioning on $R_1 = r_1$. Defining $R_2 \equiv F_{X_2|R_1,X_1,Z_2}(X_2|R_1, X_1, Z_2)$, it follows that $R_2 = F_{V_2|V_1,X_1,Z_2}(V_2|V_1, X_1, Z_2)$. Since X_1 depends only on V_1 and Z_1, and V_1 is already conditioned on, X_1 can be replaced with Z_1 as $R_2 = F_{V_2|V_1,Z_1,Z_2}(V_2|V_1, Z_1, Z_2)$. Suppose that $(Z_1, Z_2) \perp\!\!\!\perp (U_2, V_2)|V_1$. Conditioning on (Z_1, Z_2) can be relaxed, so $R_2 = F_{V_2|V_1}(V_2|V_1)$. Thus, (X_1, X_2, Z_1, Z_2) is sufficient data to recover the conditional rank of V_2 given V_1.

Assuming V_2 is continuously-distributed at each value in the support of V_1, then $F_{V_2|V_1}^{-1}$ exists. Fixing $(R_1, R_2) = (r_1, r_2)$, there exists (v_1, v_2) such that $v_1 = F_{V_1}^{-1}(r_1)$ and $v_2 = F_{V_2|V_1}^{-1}(r_2|v_1)$. Thus, conditioning jointly on $(V_1, V_2) = (v_1, v_2)$ is equivalent to conditioning jointly on $(R_1, R_2) = (r_1, r_2)$, so $(X_1, X_2) \perp\!\!\!\perp U_2|(V_1, V_2)$ implies $(X_1, X_2) \perp\!\!\!\perp U_2|(R_1, R_2)$, so that (R_1, R_2) provides a (vector-valued) control function such that β_2 is identified by (2.10), subject to the support condition that $\text{Var}(X_2|R_1 = r_1, R_2 = r_2) > 0$, $\forall(r_1, r_2) \in [0, 1]^2$. This is the insight of this paper: the recursively-constructed vector of control functions (R_1, R_2) permits identification of the effect of X_t on Y_t with unrestricted persistence of the nonseparable unobservables over time, under monotonicity restrictions on the unobservables and independence and support conditions on the instruments.

2.3 Discussion and Preview of General Results

Before proceeding to general results, it is useful to briefly discuss three properties of the identification results above that will be discussed in greater detail in a general context below. First, the identification results above did not impose any parametric restrictions on the joint distribution of the unobservables over time. For example, it was not necessary to assume that the dependence in V_1 and V_2 is due to a fixed effect, as required by Chernozhukov et al. (2013, 2015). The fixed effect assumption is clearly a special case, e.g., if $V_t = \overline{V} + \eta_t$, where η_t is *iid*, the above identification arguments for β_1 and β_2 remain valid.

Second, this identification approach allows for the number of nonseparable unobservables to be greater than the number of instruments in the input equation, and does not impose independence of all unobservables from all instruments over time. In (2.12), there are two nonseparable unobservables, V_1 and V_2, but only one instrument, Z_1, and Z_1 may be arbitrarily dependent on (V_1, U_1). It is shown below that, with T time periods, identification may be achieved even when the dimension of the unobservables is many times greater than that of the instruments, and the instruments and unobservables may evolve jointly over time.

Third, identification at $t = 2$ requires that one of the unobservables, V_2, enters the input equation monotonically and satisfies a conditional independence condition with respect to the instrument. However, other unobservables in the input equation do not need to satisfy monotonicity and may be correlated with the instrument at $t = 2$. For example, in (2.12), $\frac{\partial X_2}{\partial V_2} = 1 + X_1 + Z_1$, which is always positive, while $\frac{\partial X_2}{\partial V_1} = \psi'(V_1) + X_1 + Z_1$, which may be positive for some values of V_1 and negative for others. This may be less objectionable than the assumption that there is only one unobservable and it enters monotonically; see the discussion by Hoderlein and Mammen (2007) and Kasy (2011).

3 Nonparametric Identification

3.1 Triangular Panel Model

This subsection presents the notation and equations required for the general statement of the triangular panel model, which is the most general model identified in this paper. Time is indexed by $t = 1, 2, \ldots, T$, where T is the final observed time period. We use upper-case letters to denote random variables, lower-case letters to denote constants, and bold font to denote vectors. For any random variable A_t, denote the *history* of A_t by $A^{(t)} \equiv (A_1, A_2, \ldots, A_t)$. For convenience, denote $A^{(0)} \equiv \emptyset$. The CDF of A_t is denoted F_{A_t} and the conditional CDF of A_t given B_t is denoted $F_{A_t|B_t}$.

The observable random variables are defined as follows: Let $X_{t,k}$ denote the k^{th} scalar input, $\mathbf{X}_t \equiv (X_{t,1}, X_{t,2}, \ldots, X_{t,K})$ denote the vector of inputs at time t, Y_t denote the scalar output, $\mathbf{Z}_t$ denote the vector of instruments at time t, and $\mathbf{W}$ denote the vector of exogenous observable heterogeneity. The observed data is $(Y^{(T)}, \mathbf{W}, \mathbf{X}^{(T)}, \mathbf{Z}^{(T)})$.[1] The unobservable random variables are defined as follows: Let $V_{t,k}$ denote the scalar unobservable heterogeneity associated with input $X_{t,k}$, and $\mathbf{V}_t \equiv (V_{t,1}, V_{t,2}, \ldots, V_{t,K})$ denote the vector of input heterogeneity at time t. Let U_t denote the scalar unobservable heterogeneity associated with output Y_t. Let ϵ_t denote measurement error associated with Y_t, with possibly infinite dimension.

The *triangular panel model* is defined by,

$$\begin{aligned} X_{t,k} &= h_{t,k}\left(\mathbf{W}, \mathbf{X}^{(t-1)}, \mathbf{Z}^{(t)}, U^{(t-1)}, \mathbf{V}^{(t-1)}, V_{t,k}\right),\ k = 1, 2, \ldots, K \\ Y_t &= g_t\left(\mathbf{W}, \mathbf{X}^{(t)}, U^{(t)}, \mathbf{V}^{(t)}, \epsilon_t\right), \end{aligned} \tag{3.1}$$

[1] All identification results will hold if there are multiple outputs or if the exogenous covariates vary over time in a known way (e.g., age increases by one unit in each time period), but I focus on a single output Y_t and time-invariant exogenous observables $\mathbf{W}$ for simplicity.

where the conditional CDF of the history of the unobservables $\mathcal{F}_{\mathbf{V}^{(T)},U^{(T)}|\mathbf{X}^{(t-1)},\mathbf{Z}^{(t-1)},V^{(t-1)},U^{(t-1)}}$ is unrestricted.[2] The set of equations for $\mathbf{X}_t$ are referred to as the first stage or input equations at t, and the equation for Y_t as the second stage or output equations at t.

Remark 2.1: Notice that the model in the initial period $t = 1$ reduces to,

$$\begin{aligned} X_{1,k} &= h_{1,k}\left(\mathbf{W}, \mathbf{Z}_1, V_{1,k}\right),\ k = 1, 2, \ldots, K \\ Y_1 &= g_1\left(\mathbf{W}, \mathbf{X}_1, U_1, \mathbf{V}_1, \epsilon_1\right), \end{aligned} \tag{3.2}$$

If there is only one endogenous variable ($K = 1$), this is the triangular model identified by Imbens and Newey (2009), so the triangular model of Imbens and Newey (2009) can be thought of as defining the initial conditions for the triangular panel model considered here.

Remark 2.2: There are more first stage unobservables than endogenous inputs if $t \geq 2$. In particular, $1+(K+1)(t-1)$ unobservables appear in $h_{t,k}$, which is greater than K if $t \geq 2$ and increases by $K + 1$ unobservables in each time period. The second stage similarly adds $K + 1$ endogenous unobservables in each period, as well as the exogenous unobservables ϵ_t, which may be infinite in dimension.

Remark 2.3: Suppose that (i) the output equation is *memoryless* in the sense that no lagged observables or unobservables are included, (ii) $\mathbf{V}_t = V_1, \forall t$, so that $\mathbf{V}_t$ is a *fixed effect*, and (iii) the only time-varying unobservables included in the output equation are *purely transitory* (i.e., $U_t = 0, \forall t$). Then, the output equation of (3.1) reduces to,

$$Y_t = g_t\left(\mathbf{W}, \mathbf{X}_t, \mathbf{V}_1, \epsilon_t\right). \tag{3.3}$$

which is the nonseparable fixed effects model studied by Altonji and Matzkin (2005), Chernozhukov et al. (2013, 2015). Thus, the output model in (3.1) is strictly more general than

[2]In Skorokhod's representation, unobservables evolve according to,

$$\begin{aligned} V_{t,k} &= \zeta_t^k\left(\mathbf{X}^{(t-1)}, \mathbf{Z}^{(t-1)}, V^{(t-1)}, U^{(t-1)}, \nu_t^k\right),\ k = 1, 2, \ldots, K, \\ U_t &= \zeta_t^U\left(\mathbf{X}^{(t-1)}, \mathbf{Z}^{(t-1)}, V^{(t-1)}, U^{(t-1)}, \nu_t^U\right), \end{aligned}$$

where $\nu_t^k\big|\left(V^{(t-1)}, U^{(t-1)}\right) \sim \text{Uniform}(0,1), k = 1, 2, \ldots, K,\ \nu_t^U\big|\left(V^{(t-1)}, \epsilon^{(t-1)}\right) \sim \text{Uniform}(0,1)$, the functions $(\zeta_t^1, \zeta_t^2, \ldots, \zeta_t^K, \zeta_t^U)$ and the copula of $\left(\nu_t^1, \nu_t^2, \ldots, \nu_t^K, \nu_t^U\right)$ is unrestricted. This representation makes explicit that unobservables at time t may depend generally on inputs and instruments from previous time periods, and may follow any persistent process, such as a random walk or moving average process.

the nonseparable fixed effects output model.

3.2 Identifying Assumptions

For notational convenience, define the historical conditioning set at t by,

$$\mathbf{M}_t \equiv \left(\mathbf{X}^{(t-1)}, \mathbf{Z}^{(t-1)}, U^{(t-1)}, \mathbf{V}^{(t-1)}\right). \tag{3.4}$$

Identification in this section relies on the conditional rank device (or control function),

$$R_{t,k} \equiv F_{X_{t,k}|\mathbf{W},\mathbf{Z}_t,\mathbf{M}_t}(X_{t,k}|\mathbf{W}, \mathbf{Z}_t, \mathbf{M}_t), \tag{3.5}$$

which extends the conditional rank device of Imbens and Newey (2009) to a panel environment. The identifying assumptions imposed on the triangular panel model are as follows:

A.1 $(U_t, \mathbf{V}_t) \perp\!\!\!\perp (\mathbf{Z}_t, \mathbf{W})|\mathbf{M}_t$;

A.2 Conditional on $\mathbf{M}_t$, $V_{t,k}$ is a continuously-distributed scalar with conditional distribution, $\mathcal{F}_{V_{t,k}|\mathbf{M}_t}$, that is strictly increasing across its support;

A.3 $h_{t,k}$ is strictly monotonic in $V_{t,k}$ with probability one;

A.4 Support $\left(\mathbf{R}^{(t)}|\mathbf{W} = \mathbf{w}, \mathbf{X}^{(t)} = \mathbf{x}^{(t)}\right) = [0,1]^{tK}, \forall(\mathbf{w}, \mathbf{x}^{(t)}) \in$ Support $\left(\mathbf{W}, \mathbf{X}^{(t)}\right)$;

A.5 $\epsilon_t \perp\!\!\!\perp \epsilon_{t'}$, $\forall t' \neq t$, and $\epsilon_t \perp\!\!\!\perp (\mathbf{W}, \mathbf{M}_{T+1})$; and,

A.6 $\mathbb{E}[|Y_t|] < \infty$.

A.1 extends assumption (i) of Theorem 1 by Imbens and Newey (2009) to a panel environment. Because it only requires contemporaneous independence at time t, it permits $\mathbf{V}_t$ to depend generally on any lagged values of the instruments. For example, in the motivating model of firms and workers discussed above, this allows that labor productivity responds to changes in capital costs, e.g., through on-the-job training.

A.2 extends assumption (ii) of Theorem 1 by Imbens and Newey (2009) to a panel environment. This assumptions means that $\mathbf{V}_t$ has a smooth conditional distribution, conditional on lagged values of observables and unobservables. This restricts us from considering models in which the error follows a discrete Markov chain with finite support, which is often assumed in empirical macroeconomic models.

A.3 extends model equation 2.2 by Imbens and Newey (2009) to a panel environment. The strength of first stage monotonicity assumptions in cross-sectional contexts has been discussed by Hoderlein and Mammen (2007) and Torgovitsky (2015), and a test has been developed by Hoderlein et al. (2016). However, this restriction is much less severe here as it allows for many unobservables in the first stage that do not satisfy monotonicity, $U^{(t-1)}$ and $\mathbf{V}^{(t-1)}$, while only one unobservable must satisfy monotonicity, $V_{t,k}$.

A.4 extends assumption 2 by Imbens and Newey (2009) to a panel environment. It requires that the instrument has sufficient variation to trace out the full support of the conditional distribution of $V_{t,k}$ at each time period. **A.4** is testable and depends crucially on the instruments available in the data. Full support assumptions of this form are often found not to hold empirically, so the next section of this paper presents semiparametric model restrictions under which **A.4** is not necessary, but **A.4** is maintained in this section.

A.5 defines explicitly what it means for ϵ_t to be measurement error in Y_t: it is strictly independent of all observable and unobservable model components other than Y_t, while **A.6** ensures that the outcome has a finite first moment.

Lastly, we introduce two assumptions, where only one assumption will be imposed at a time, and substantially different identification results are achieved depending on which of the two assumptions is imposed for each t:

B.1 $U^{(t)}$ is excluded from $h_{t,k}, \forall k$; and,

B.2 g_t excludes ϵ_t and is strictly monotonic in U_t with probability one.

B.1 allows for second stage unobservables to have infinite dimension, but imposes that second stage unobservables are excluded from the first stage. In practice, this will allow us to integrate out the second stage unobservables in order to identify averages of g_t, without needing to account for second stage unobservables in the first stage.

B.2 is analogous to the restriction imposed on the model of Imbens and Newey (2009) by Torgovitsky (2015) in order to relax the full-support condition on the ranks. However, it is less objectionable, as it allows for many unobservables in the output equation, $\left(U^{(t)}, \mathbf{V}^{(t)}\right)$, with g_t only required to be monotonic in one of these unobservables, U_t.

3.3 Identification Results

The nonparametric identification results of this paper are now presented in two theorems:

Theorem 1 (Nonparametric identification of the triangular panel model with infinite-dimensional second stage unobservables)**.** *Suppose assumptions **A.1** -**A.6** and **B.1** hold. Then, for each t,*

(a) $R_{t,k} = F_{V_{t,k}|\mathbf{M}_t}(V_{t,k}|\mathbf{M}_t), \forall k$*;*

(b) $(U_t, \epsilon_t) \perp\!\!\!\perp \mathbf{X}_t | \mathbf{R}^{(t)}, \mathbf{M}_t$*; and,*

(c) $\mathbb{E}[\mathbb{E}[Y_t|\mathbf{W} = \mathbf{w}, \mathbf{X}^{(t)} = \mathbf{x}^{(t)}, \mathbf{R}^{(t)}]] = \mathbb{E}[g_t(\mathbf{w}, \mathbf{x}^{(t)}, \mathbf{V}^{(t)}, U^{(t)}, \epsilon_t)]$.

Proof. See Appendix A.1. □

The proof is carried out by induction. First, it is established at time $t = 1$ following Imbens and Newey (2009), recalling that the model in the initial time period $t = 1$ differs only from the model identified by Imbens and Newey (2009) in that there are multiple endogenous inputs (see Remark 2.1). Second, it is established that the theorem holds for arbitrary $t \geq 2$ if the theorem holds for $t - 1$ using the conditional rank device. The second step relies crucially on the exclusion of $U^{(t)}$ from the first stage, as this allows for the isolation of $V_{t,k}$ through the one-to-one mapping between $V_{t,k}$ and $X_{t,k}$ for each t, k, and the isolation of each $V_{t,k}$ permits controlling for $\mathbf{V}^{(t-1)}$ in the second stage to solve the endogeneity problem. By contrast, the next theorem identifies similar objects when including $U^{(t)}$ in the first stage but excluding the infinite-dimensional exogenous unobservable ϵ_t from the second stage:

Theorem 2 (Nonparametric identification of the triangular panel model with output unobservables included recursively in the input equations)**.** *Suppose assumptions* ***A.1*** *-* ***A.6*** *and* ***B.2*** *hold. Furthermore, define,*

$$Q_t \equiv F_{Y_t|\mathbf{W},\mathbf{V}_t,\mathbf{M}_t}(Y_t|\mathbf{W},\mathbf{V}_t,\mathbf{M}_t), \tag{3.6}$$

Then,

(a) $R_{t,k} = F_{V_{t,k}|\mathbf{M}_t}(V_{t,k}|\mathbf{M}_t), \forall k,$ *and* $Q_t = F_{U_t|\mathbf{V}_t,\mathbf{M}_t}(U_t|\mathbf{V}_t,\mathbf{M}_t)$*;*

(b) $(U_t \perp\!\!\!\perp \mathbf{X}_t)|(\mathbf{R}_t,\mathbf{M}_t)$*; and,*

(c) $\mathbb{E}[\mathbb{E}[Y|\mathbf{W}=\mathbf{w}, X^{(t)}=x^{(t)}, \mathbf{R}^{(t)}, Q^{(t)}]] = \mathbb{E}[g_t(\mathbf{w}, x^{(t)}, \mathbf{V}^{(t)}, U^{(t)})]$.

Proof. See Appendix A.2. □

The proof is very similar to that of Theorem 1, with the additional challenge of proving the second statement in Theorem 2(a) by induction.

4 Semiparametric Identification and Estimation

4.1 Random Coefficient and Markov Chain Assumptions

There are three reasons why directly implementing the sample counterpart to the nonparametric estimator in Theorem 1(a) or Theorem 2(a) for the input equations and Theorem 1(c) or Theorem 2(c) for the output equations is impractical. First, the estimator suffers from a cross-sectional curse of dimensionality: even at time $t = 1$, there may be many controls to include in the nonparametric regression, as discussed by Imbens and Newey (2009) and Torgovitsky (2015). Second, the estimator suffers from a panel curse of dimensionality: because new conditional rank devices, instruments, and inputs are added to the model over time as part of the histories of these variables, the number of controls to include in the nonparametric regression increases rapidly over time. Third, the estimator suffers from unavailability of full support instruments: instruments often have small support, so an identification approach that relies on large support instruments is not always useful.

In order to address the cross-sectional curse of dimensionality as well as the unavailability of full support instruments, this section follows the approach of Ma and Koenker (2006), Jun (2009), and Masten and Torgovitsky (2016), who impose a semiparametric functional form in both the input and output equations. I extend this to approach to a panel environment: $V_{t,k}$ enters the input equation for $X_{t,k}$ through correlated random coefficients. When **B.1** is imposed (so that U_t is excluded), ϵ_t enters the output equation for Y_t through correlated random coefficients; when **B.2** is imposed (so that ϵ_t is excluded), U_t enters the output equation for Y_t through correlated random coefficients.

The specification may include transformations of the observable variables, such as squared terms or interaction terms, which Masten and Torgovitsky (2016) refer to as *derived* endogenous variables, in contrast to (untransformed) *basic* endogenous variables. The panel context has the added complexity that lagged *unobservables* are included in the basic endogenous variables, and may also be included in the derived endogenous variables. Note that, although this semiparametric restriction maintains nonseparability between contemporaneous unobservables and any basic or derived endogenous variables, it requires us to explicitly specify any interaction between lagged unobservables and any other endogenous variables, or interactions among any observables. The fully nonparametric approach can be thought of as including all possible (infinitely-many) derived endogenous variables, so the semiparametric approach avoids the curse of dimensionality by choosing a finite subset these.

To address the panel curse of dimensionality, this section proposes imposing an s^{th}-order Markov-like assumption on the panel model, which I will refer to as the *memory limit of order s*. The memory limit is modeled in two parts: an exclusion part and a Markov chain part. For the exclusion part, any observables and unobservables at time $t - s$ or earlier are excluded from the input and output equations at t. For the Markov chain part, the unobservables are jointly assumed to be a Markov chain of order s. For notational purposes, while the history of any random variable A_t is defined as $A^{(t)} \equiv (A_1, A_2, \ldots, A_t)$ above, this section defines the memory of any random variable A_t as $A^{(t,s)} \equiv (A_{t-s+1}, A_{t-s+2}, \ldots, A_t)$, which has s components. For convenience, denote $A^{(t,0)} \equiv \emptyset$. Memory limits are common in panel applications: a first-order Markov chain (such as a random walk) satisfies $s = 1$, an autoregressive process of order 2 satisfies $s = 2$, and the nonseparable fixed effects model of Chernozhukov et al. (2015) satisfies $s = 0$ (memoryless).

Combining the assumptions above with the memory limit s, the triangular panel model in (3.1) simplifies to,

$$\begin{aligned} X_{t,k} &= \mathbf{M}_t' \mathbf{\Pi}_{t,k}(V_{t,k}), k = 1, 2, \ldots, K \\ Y_t &= \mathbf{P}_t' \mathbf{\Psi}_t(U_t, \epsilon_t), \end{aligned} \tag{4.1}$$

where $\mathbf{\Pi}_{t,k}$ and $\mathbf{\Psi}_t$ are vector-valued random coefficients and the matrices $\mathbf{M}_t$ and $\mathbf{P}_t$ include lagged *observable and unobservable* components, including derived variables, as,

$$\begin{aligned} \mathbf{M}_t &= m_t(\mathbf{W}, \mathbf{Z}^{(t,s)}, U^{(t-1,s)}, \mathbf{V}^{(t-1,s)}) \\ \mathbf{P}_t &= p_t(\mathbf{W}, \mathbf{X}^{(t,s)}, U^{(t-1,s)}, \mathbf{V}^{(t-1,s)}) \end{aligned} \tag{4.2}$$

where m_t and p_t select the basic and derived endogenous variables to include. That $\mathbf{M}_t$ and $\mathbf{P}_t$ include unobserved components is part of the identification challenge, and is not involved in the correlated random coefficients literature cited above.

For example, suppose $K = 1$ (so that the k subscript can be omitted), $\mathbf{Z}_t$ is scalar, and $\mathbf{W} = (1)$ (the intercept only). If $s = 0$ and only basic endogenous variables are selected by m_t and p_t, the model is,

$$\begin{aligned} X_t &= (1, Z_t, Z_{t-1})' \mathbf{\Pi}_{t,k}(V_t), \\ Y_t &= (1, X_t, V_{t-1}, X_t V_{t-1})' \mathbf{\Psi}_t(U_t, \epsilon_t), \end{aligned} \tag{4.3}$$

which is the model considered as an illustrative example in Subsection 2.2. If $s = 1$ and all interactions between lagged unobservables and current observables are selected, it is,

$$\begin{aligned} X_t &= (1, Z_t, Z_{t-1}, U_{t-1}, V_{t-1}, Z_t V_{t-1}, Z_t U_{t-1})' \mathbf{\Pi}_{t,k}(V_t), \\ Y_t &= (1, X_t, X_{t-1}, U_{t-1}, V_{t-1}, X_t V_{t-1}, X_t U_{t-1})' \mathbf{\Psi}_t(U_t, \epsilon_t), \end{aligned} \tag{4.4}$$

which is already a very flexible specification, e.g., $(0, X_t, 0, 0, 0, V_{t-1}, U_{t-1})' \mathbf{\Psi}_t(U_t, \epsilon_t)$ is the marginal effect of X_t on Y_t, which includes many dimensions of heterogeneity from unobservables, observables, and their interaction.

Remark 3.1: In (4.1), the model at the initial time period $t = 1$ reduces to,

$$\begin{aligned} X_{1,k} &= (\mathbf{W}, \mathbf{Z}_1)'\Pi_{1,k}(V_{1,k}), k = 1, 2, \ldots, K \\ Y_1 &= (\mathbf{W}, \mathbf{X}_1)'\Psi_1(\epsilon_1), \end{aligned} \tag{4.5}$$

where U_1 is ignored to simplify notation, which is exactly the model identified by Masten and Torgovitsky (2016). Furthermore, if ϵ_1 is scalar and $K = 1$, this is exactly the model identified by Ma and Koenker (2006) under additional parametric restrictions and by Jun (2009) without additional parametric restrictions. As a result, the model of Jun (2009) can be thought of as defining the initial conditions if ϵ_1 is scalar and $K = 1$, and the model of Masten and Torgovitsky (2016) can be thought of as defining the initial conditions otherwise.

Remark 3.2: In (4.1), there are more unobservables in the input equation than endogenous inputs in the output equation at t if $t \geq 2$ and $s \geq 1$. In particular, if only the basic endogenous variables are included, $K+2$ unobservables are nonseparable while the remaining $s(K+1) - K - 1$ unobservables are separable from the observables. If desired, interactions between lagged unobservables and endogenous observables can be specified as in (4.4).

Remark 3.3: In (4.1), suppose $s = 0$ (memoryless), ϵ_t is excluded from the output equation, and $U_t = \nu + \eta_t$, where ν is a fixed effect and η_t is a strictly exogenous measurement error. Then, the output equation at t becomes,

$$Y_t = (\mathbf{W}, \mathbf{X}_t)'\Psi_t(U_t), \tag{4.6}$$

which is the random coefficients panel model considered by Graham and Powell (2012) and is similar to the random location-scale panel model considered by Chernozhukov et al. (2015).

4.2 Identification with Small-support Instruments

This subsection first presents identification results under high-dimensional unobservables in the output equation under **B.1**, then presents identification results permitting lagged unobservables from the output equation to be included in the input equation. Throughout this section, **A.1**, **A.2**, **A.3**, **A.5**, and **A.6** are assumed to hold. The conditional rank devices and instrumental support assumption **A.4** are modified to take advantage of the functional form restrictions.

For identification with high-dimensional unobservables in the output equation, consider the conditional rank device,

$$R^*_{t,k} = \int_0^1 \mathbb{1}[\mathbf{M}_t' \Pi_{t,k}(r) \leq X_{t,k}]dr, \tag{4.7}$$

which makes use of the pre-arrangement operator of Chernozhukov et al. (2010). Furthermore, replace **A.4** by,

A.4* $\mathbb{E}[\mathbf{M}_t\mathbf{M}_t'|\mathbf{R}_t^* = \mathbf{r}]$ and $\mathbb{E}[\mathbf{P}_t\mathbf{P}_t'|\mathbf{R}_t^* = \mathbf{r}]$ are invertible, for almost every $\mathbf{r} \in [0,1]^K$.

Because $\mathbf{Z}_t$ is a component of $\mathbf{M}_t$, the first part of **A.4*** requires that there is variation in the instruments at each value of the conditional rank device, a condition that may very well be satisfied by instruments with small support. It is the same as the familiar instrumental relevance condition in linear instrumental variables contexts, but local to $\mathbf{R}_t^*$.

The following theorem holds:

Theorem 3 (Semiparametric identification of the triangular panel model with infinite-dimensional output equation unobservables)**.** *Suppose that assumptions* ***A.1****,* ***A.2****,* ***A.3****,* ***A.4*****,* ***A.5****,* ***A.6****, and* ***B.1*** *hold. Then, for each t,*

(a) $R^*_{t,k} = F_{V_{t,k}|\mathbf{V}^{(t-1)}}(V_{t,k}|\mathbf{V}^{(t-1)}), \forall k;$

(b) $(U_t, \mathbf{V}_t, \epsilon_t \perp\!\!\!\perp \mathbf{X}_t)|(\mathbf{W}, \mathbf{X}^{(t-1)}, \mathbf{R}^{*(t-1)})$*; and,*

(c) $\mathbb{E}\left[\mathbb{E}[\mathbf{P}_t\mathbf{P}_t'|\mathbf{R}_t^*]^{-1}\mathbb{E}[\mathbf{P}_tY_t|\mathbf{R}_t^*]\right] = \mathbb{E}[\Psi_t(U_t, \mathbf{V}_t, \epsilon_t)].$

Proof. See Appendix C. □

The proof follows by induction. At $t = 1$, the theorem can be shown following Masten and Torgovitsky (2016), as the model at $t = 1$ is identical to the model identified by Masten and Torgovitsky (2016). Then, it is shown that the theorem holds at t if it holds at $t - 1$ using similar reasoning to the proof of Theorem 1. In particular, it is shown that $\mathbf{M}_t$ and $\mathbf{P}_t$ are recovered from the conditional rank device at t, given $\mathbf{M}_t$ and $\mathbf{P}_{t-1}$.

Next, identification results are presented under **B.2**. To this end, define the conditional rank devices,

$$R^{\dagger}_{t,k} = \int_0^1 \mathbb{1}[\mathbf{M}'_t \Pi_{t,k}(Q^{\dagger}_{t-1}, r) \leq X_{t,k}]dr, \ \forall k, \tag{4.8}$$

and,

$$Q^{\dagger}_t = \int_0^1 \mathbb{1}[\mathbf{P}'_t \Psi_t(q, \mathbf{R}^{\dagger}_t) \leq Y_t]dq. \tag{4.9}$$

In place of **A.4**, consider the following assumption:

A.4† $\mathbb{E}[\mathbf{M}_t\mathbf{M}'_t|\mathbf{R}^{\dagger}_t = \mathbf{r}]$ and $\mathbb{E}[\mathbf{P}_t\mathbf{P}'_t|\mathbf{R}^{\dagger}_t = \mathbf{r}]$ are invertible, for almost every $\mathbf{r} \in [0,1]^K$, and Ψ_t is strictly monotonic in U_t with probability one.

The invertibility assumptions in **A.4**† are analogous to those of **A.4***, but incorporating the conditional rank devices $R^{\dagger}_{t,k}$ and $Q^{\dagger}_t$.

The following theorem holds:

Theorem 4 (Semiparametric identification of the triangular panel model with lagged unobservables from the output equation included in the input equation)**.** *Suppose that assumptions* ***A.1****,* ***A.2****,* ***A.3****,* ***A.4***†*,* ***A.5****,* ***A.6****, and* ***B.2*** *hold. Then, for each* t*,*

(a) $R^{\dagger}_{t,k} = F_{V_{t,k}|\mathbf{M}_t}(V_{t,k}|\mathbf{M}_t), \forall k,$ *and* $Q^{\dagger}_t = F_{U_t|\mathbf{V}_t,\mathbf{M}_t}(U_t|\mathbf{V}_t, \mathbf{M}_t)$*;*

(b) $(U_t, \mathbf{V}_t, \epsilon_t \perp\!\!\!\perp \mathbf{X}_t)|(\mathbf{W}, \mathbf{X}^{(t-1)}, \mathbf{R}^{\dagger(t-1)})$*; and,*

(c) $\mathbb{E}\left[\mathbb{E}[\mathbf{P}_t\mathbf{P}'_t|\mathbf{R}^{\dagger}_t]^{-1}\mathbb{E}[\mathbf{P}_tY_t|\mathbf{R}^{\dagger}_t]\right] = \mathbb{E}[\Psi_t(\mathbf{V}_t, U_t)]$*.*

Proof. See Appendix D. □

As with Theorem 3, the theorem can be shown at $t = 1$ following Masten and Torgovitsky (2016), as the model at $t = 1$ is identical to the model identified by Masten and Torgovitsky (2016). Then, it is shown that the theorem holds at t if it holds at $t-1$ using similar reasoning to the proof of Theorem 2. In particular, it is shown that $\mathbf{M}_t$ and $\mathbf{P}_t$ are recovered from both the input and out equation conditional rank devices at t, given $\mathbf{M}_t$ and $\mathbf{P}_{t-1}$.

4.3 Practical Estimators and Monte Carlo Study

Estimators that directly implement the identification approaches of Theorems 3 and 4 are now proposed. These estimators are practical in that they do not suffer a curse of dimensionality and do not rely on the instruments having large support. Then, data is simulated from models satisfying the model assumptions here but violating the assumptions of existing identification approaches. The new estimators are shown to outperform existing approaches when applied to the simulated data even in small samples.

Consider the identification results of Theorems 3. At $t = 1$, Masten and Torgovitsky (2016) show that $\mathbf{R}_1^*$ can be recovered by applying the linear quantile regression of $X_{1,k}$ on $\mathbf{W}, \mathbf{Z}_1$ to estimate $\widehat{\Pi}_{1,k}(r)$ at various ranks r using the method of Koenker and Bassett (1978), and then applying the pre-arrangement operator of Chernozhukov et al. (2010). To extend the implementation of Masten and Torgovitsky (2016) to a panel environment, iteratively replace $\mathbf{M}_t \equiv (\mathbf{W}, \mathbf{Z}^{(t,s)}, \mathbf{X}^{(t-1,s)}, \mathbf{R}^{(t-1,s)})$ by $\widehat{\mathbf{M}}_t \equiv (\mathbf{W}, \mathbf{Z}^{(t,s)}, \mathbf{X}^{(t-1,s)}, \widehat{\mathbf{R}}^{(t-1,s)})$, and iteratively replace $\mathbf{P}_t \equiv (\mathbf{W}, \mathbf{X}^{(t,s)}, \mathbf{R}^{(t-1,s)})$ by $\widehat{\mathbf{P}}_t \equiv (\mathbf{W}, \mathbf{X}^{(t,s)}, \widehat{\mathbf{R}}^{(t-1,s)})$.

In particular, the estimator for the conditional rank device at t is,

$$\widehat{R}_{t,k}^* = \int_0^1 \mathbb{1}[\widehat{\mathbf{M}}_t' \widehat{\Pi}_{t,k}(r) \leq X_{t,k}] dr. \tag{4.10}$$

while the estimator for $\Psi_t(\mathbf{r})$ is,

$$\widehat{\Psi}_t(\mathbf{r}) = \left(\sum_{i=1}^N \kappa\left(\frac{\widehat{\mathbf{R}}_{t,i}^* - \mathbf{r}}{h} \right) \widehat{\mathbf{P}}_{t,i} \widehat{\mathbf{P}}_{t,i}' \right)^{-1} \left(\sum_{i=1}^N \kappa\left(\frac{\widehat{\mathbf{R}}_{t,i}^* - \mathbf{r}}{h} \right) \widehat{\mathbf{P}}_{t,i} Y_t \right). \tag{4.11}$$

Because R_t^* has the uniform distribution by construction, the average effect of $\mathbf{X}_t$ on Y_t can be estimated by $\int_0^1 \widehat{\Psi}_t(r) dr$, where the integral is replaced by a numerical approximation in practice. Together (4.10) and (4.11) define the estimator that directly implements the identification strategy of Theorem 3.

Next, to implement the estimator corresponding to Theorem 4, iteratively replace $\mathbf{M}_t \equiv (\mathbf{W}, \mathbf{Z}^{(t,s)}, \mathbf{X}^{(t-1,s)}, Q^{(t-2,s)}, \mathbf{R}^{(t-1,s)})$ by $\widehat{\mathbf{M}}_t \equiv (\mathbf{W}, \mathbf{Z}^{(t,s)}, \mathbf{X}^{(t-1,s)}, \widehat{Q}^{\dagger(t-2,s)}, \widehat{\mathbf{R}}^{\dagger(t-1,s)})$, and iteratively replace $\mathbf{P}_t \equiv (\mathbf{W}, \mathbf{X}^{(t,s)}, Q^{(t-1,s)}, \mathbf{R}^{(t-1,s)})$ by $\widehat{\mathbf{P}}_t \equiv (\mathbf{W}, \mathbf{X}^{(t,s)}, Q^{\dagger(t-1,s)}, \widehat{\mathbf{R}}^{\dagger(t-1,s)})$.

Estimators for the conditional rank devices at t are,

$$\widehat{R}_{t,k}^{\dagger} = \int_0^1 \mathbb{1}[\widehat{\mathbf{M}}_t' \widehat{\Pi}_{t,k}(\widehat{Q}_t^{\dagger}, r) \leq X_{t,k}] dr. \tag{4.12}$$

and,

$$\widehat{Q}_t^{\dagger} = \int_0^1 \mathbb{1}[\widehat{\mathbf{P}}_t' \widehat{\Psi}_{t,k}(q, \widehat{\mathbf{R}}_t^{\dagger}) \leq Y_t] dq. \tag{4.13}$$

and then apply the estimator for the output equation in (4.11). The conditional rank devices can be implemented using the kernel-weighted approximations,

$$\widehat{R}_{t,k}^{\dagger} \approx \sum_{\mathbf{r} \in \mathcal{S}^K} \sum_{q \in \mathcal{S}} \mathbb{1}[\widehat{\mathbf{M}}_t' \widehat{\Pi}_{t,k}(q, r) \leq X_{t,k}] \kappa \left((\widehat{Q}_t^{\dagger} - q)/h \right), \tag{4.14}$$

and,

$$\widehat{Q}_t^{\dagger} \approx \sum_{q \in \mathcal{S}} \sum_{\mathbf{r} \in \mathcal{S}^K} \mathbb{1}[\widehat{\mathbf{P}}_t' \widehat{\Psi}_{t,k}(q, r) \leq Y_t] \kappa \left((\widehat{\mathbf{R}}_t^{\dagger} - \mathbf{r})/h \right), \tag{4.15}$$

where $\mathcal{S}$ is a discrete approximation of the unit-interval.

Now, consider a model satisfying the assumptions of Theorem 3 with $K = 1$, $s = 3$, and bivariate measurement error in the outcome equation. At time $t = 1$, specify,

$$X_1 = V_1 + Z_1 V_1,$$
$$Y_1 = A_1 + B_1 X_1.$$

Next, at time $t = 2$, specify,

$$X_2 = V_2 + (Z_2 + V_1) V_2,$$
$$Y_2 = A_2 + X_2 B_2 + V_1.$$

Lastly, for $t = 3, 4, \ldots, T$, specify,

$$X_t = V_t + (Z_t + V_{t-1}) V_2,$$
$$Y_t = A_t + X_t B_t + V_{t-2} + V_{t-1}.$$

Specify $T = 10$, $A_t = V_t + \epsilon_t^A$, $B_t = 2(V_t + \epsilon_t^B)$, $\epsilon_t^j \sim_{iid} \mathcal{N}(0, 1/10), j = A, B$, $Z_t \sim_{iid}$ *Bernoulli*$(1/2)$, and $(V_1, V_2, \ldots, V_T)$ is drawn from the Gaussian copula with covariance

Table 1: Monte Carlo Study of Estimation Bias

Panel A. Output Equation Estimation Bias					
Bias in Estimator of $\mathbf{E}[B_t] = 1$	$t = 1$	$t = 2$	$t = 3$	$t = 5$	$t = 10$
Ordinary Least Squares	1.308	1.374	1.495	1.492	1.496
	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)
Two Stage Least Squares	0.335	0.333	0.338	0.337	0.324
	(0.036)	(0.050)	(0.046)	(0.048)	(0.048)
Cross-sectional Masten & Torgovitsky	0.006	0.100	0.492	0.503	0.466
	(0.063)	(0.116)	(0.192)	(0.188)	(0.192)
New Estimator	0.006	0.009	-0.056	0.010	0.001
	(0.063)	(0.064)	(0.059)	(0.075)	(0.078)
Panel B. Recovery of Ranks from Input Equation					
Correlation between V_t and $\widehat{V}_t$	$t = 1$	$t = 2$	$t = 3$	$t = 5$	$t = 10$
Cross-sectional Masten & Torgovitsky	1.000	0.965	0.885	0.885	0.886
New Estimator	1.000	0.994	0.975	0.971	0.967

Notes: Estimates of bias and correlation are averaged across 100 simulations of size $N = 5,000$. The kernel κ is biweight with bandwidth $h = 0.025$. Standard errors are based on $B = 40$ bootstrap draws.

matrix where each diagonal element is 1 and each off-diagonal element if 1/10. It follows that $\mathbb{E}[B_t] = 1$, which is the parameter of interest.

Table 1, Panel A, demonstrates the success of the new estimator relative to the other potential estimators. OLS and TSLS exhibit large positive biases. The method of Masten and Torgovitsky (2016) applied to the cross-sectional data recovers the true parameter at time $t = 1$, but has a small bias when $t = 2$ due to the omission of V_{t-1} and a larger bias when $t > 2$ due to the omission of both V_{t-1} and V_{t-2}. The new estimator is very close to the true parameter at all time periods, as it explicitly accounts for the history of V_t.

Table 1, Panel B provides additional information on why the new estimator performs better than the cross-sectional approach. The cross-sectional estimator recovers V_t at $t = 1$ but not at $t > 1$, while the new estimator also recovers V_t at $t > 1$ with little bias. Of particular interest, V_t is still recovered with little bias at $t = 10$.

5 Data and Empirical Results

5.1 Norwegian Tax Records on Firms and Workers

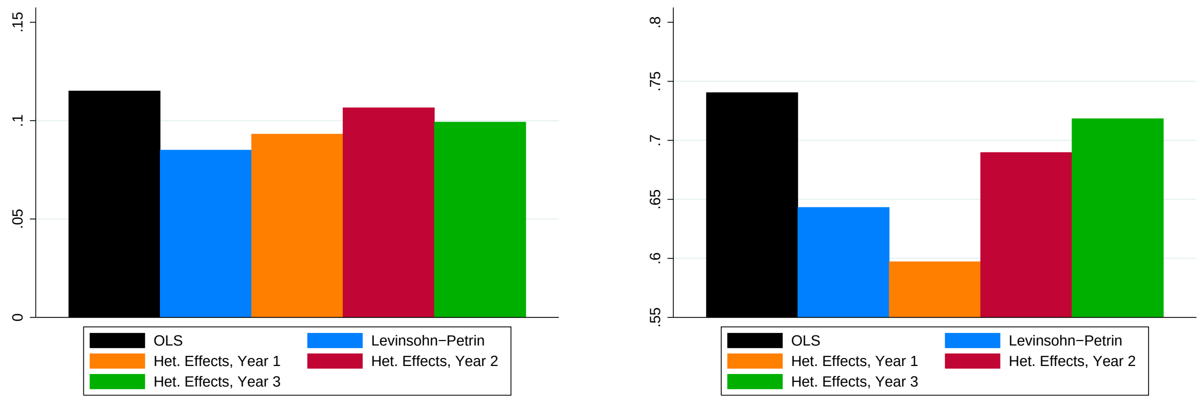
The empirical context is the population of firms in Norway, observed in a 12 year panel of tax records of annual frequency from 2003 until 2014. Each observation-year includes gross revenues, labor costs, intermediate input costs (other than labor costs), accumulated capital (fixed assets), industry code, municipal code, and the year the firm was founded. I form value-added as the difference between gross revenues and intermediate input costs. Furthermore, to permit the identification of a panel process with persistence of at least three periods, I require that all variables are observed for at least three adjacent periods. Because the identification approach requires that the initial period of production be observed, I only consider firms that are founded within the data window. The resulting sample size is $19,613$.

To construct the local price of labor, I match the region and industry of the firm to the earnings tax records of all workers in the same region and industry. Assuming that each firm is a price-taker, prices within the industry and region are independent of firm-specific productivity shocks. Furthermore, as the price of labor rises, firms are expected to respond by changing capital investment and labor expenditures. Together, these imply that local prices are both relevant and independent of firm-specific unobservables, which are the conditions required for the local price of labor to serve as a valid instrumental variable.

5.2 Empirical Results

There are three empirical questions of interest. First, what is the average *ceteris paribus* elasticity of production with respect to capital and labor? This is the question addressed by the existing literature, including Olley and Pakes (1996) and Levinsohn and Petrin (2003). Second, do these averages vary over time? Time-variation is ruled out *ex ante* by pooled models that assume elasticities do not vary with time, including those employed by Levinsohn and Petrin (2003). Third, do the elasticities of production with respect to capital and labor vary with the firm productivity shock? This question is ruled out *ex ante* by the Cobb-

Figure 1: Time Variation in Average Elasticities of Capital and Labor



(a) Capital (b) Labor

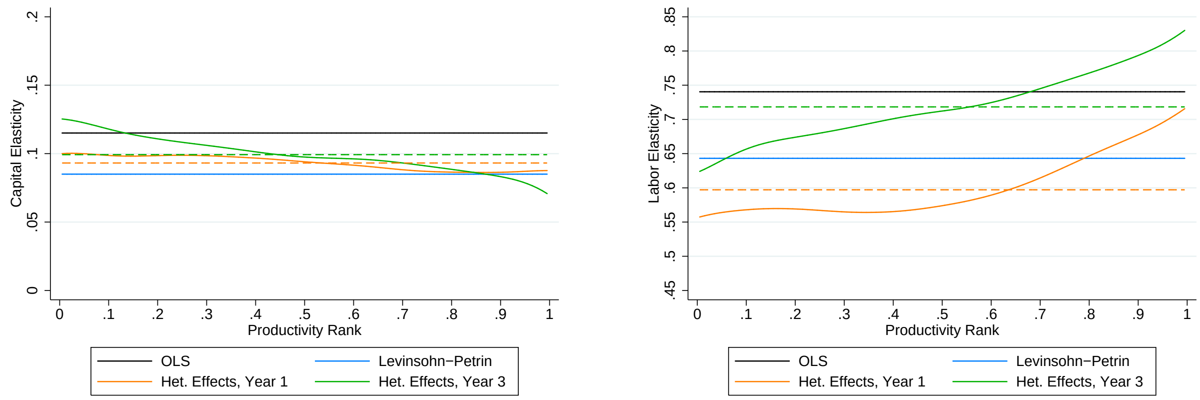
Douglas specification standard in the literature (Ackerberg et al., 2015).

Figure 1 demonstrates that the average elasticity of production with respect to capital is rising over the first three years of firm production. The pooled estimator of Levinsohn and Petrin (2003) successfully recovers average productivity of labor across all years, but misses the year-to-year increase. Figure 2 demonstrates that firm productivity shocks augment the productivity of labor, as higher firm productivity raises the elasticity with respect to labor. By contrast, higher firm productivity does not appear to augment the elasticity with respect to capital.

6 Conclusions

This paper has developed an approach to identify *ceteris paribus* effects of continuous regressors in nonseparable panel models without time homogeneity. From the insight that the passing of time creates a triangular structure in which shocks realized in the future are excluded at present, a novel recursive control function was proven to control for persistence in the unobservables. To overcome the curse of dimensionality and deal with the empirical reality that many instruments have limited support, it was shown that the recursive control function identifies the model under semiparametric restrictions even if the instruments have

Figure 2: Unobserved Heterogeneity in Elasticities of Capital and Labor



(a) Capital

(b) Labor

small support, and the associated estimator does not suffer from a curse of dimensionality.

In a novel empirical application, the *ceteris paribus* elasticities of firm production with respect to capital and labor are estimated without separability of firm productivity shocks from capital or labor. It was found that existing approaches that require the Cobb-Douglas functional form of firm production identify the average elasticities of capital and labor, but mask heterogeneity that is found to be especially important when labor is divisible into skill types. In particular, it is found that high-productivity firms have higher elasticity of high-skill labor productivity and lower elasticity of low-skill labor productivity.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Altonji, J. G. and R. L. Matzkin (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica 73*(4), 1053–1102.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard university press.

Bester, C. A. and C. Hansen (2012). Identification of marginal effects in a nonparametric correlated random effects model. *Journal of Business & Economic Statistics*.

Blundell, R. and S. Bond (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric reviews 19*(3), 321–340.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica 78*(3), 1093–1125.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013). Average and quantile effects in nonseparable panel models. *Econometrica 81*(2), 535–580.

Chernozhukov, V., I. Fernandez-Val, S. Hoderlein, H. Holzmann, and W. Newey (2015). Nonparametric identification in panels using quantiles. *Journal of Econometrics*.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica 71*(5), 1405–1441.

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*(5), 2105–2152.

Griliches, Z. and J. Mairesse (1995). Production functions: the search for identification. Technical report, National Bureau of Economic Research.

Heckman, J. and E. Vytlacil (1998). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources*, 974–987.

Hoderlein, S. and E. Mammen (2007). Identification of marginal effects in nonseparable models without monotonicity. *Econometrica 75*(5), 1513–1518.

Hoderlein, S., L. Su, H. White, and T. T. Yang (2016). Testing for monotonicity in unobservables under unconfoundedness. *Journal of Econometrics 193*(1), 183–202.

Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica 77*(5), 1481–1512.

Jun, S. J. (2009). Local structural quantile effects in a model with a nonseparable control variable. *Journal of Econometrics 151*(1), 82–97.

Kasy, M. (2011). Identification in triangular systems using control functions. *Econometric Theory 27*, 663–671.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica*, 33–50.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*(2), 317–341.

Ma, L. and R. Koenker (2006). Quantile regression methods for recursive structural equation models. *Journal of Econometrics 134*(2), 471–506.

Marschak, J. and W. H. Andrews (1944). Random simultaneous equations and the theory of production. *Econometrica*, 143–205.

Masten, M. A. and A. Torgovitsky (2016). Identification of instrumental variable correlated random coefficients models. *Review of Economics and Statistics* (0).

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Syverson, C. (2004). Market structure and productivity: A concrete example. *Journal of Political Economy 112*(6), 1181–1222.

Torgovitsky, A. (2015). Identification of nonseparable models using instruments with small support. *Econometrica 83*(3), 1185–1197.

A Appendix

A.1 Proof of Theorem 1

This section proves that Theorem 1 holds at time $t = 1$, and Theorem 1 holds at time t if it holds at time $t-1$. By induction, Theorem 1 holds for all t.

Step 1. *Proof of Theorem 1 at time $t = 1$.*

Fix k. Because there is only one random unobservable $V_{1,k}$ in $h_{1,k}$, and $X_{1,k}$ is monotonic in $V_{1,k}$ by **A.3**, $h_{1,k}$ can be partially inverted with respect to $V_{1,k}$ such that,

$$R_{1,k} \equiv F_{X_{1,k}|\mathbf{W},\mathbf{Z}_1}\left(X_{1,k}|\mathbf{W},\mathbf{Z}_1\right) = F_{V_{1,k}|\mathbf{W},\mathbf{Z}_1}\left(h_{1,k}^{-1}\left(\mathbf{W}, X_{1,k}, \mathbf{Z}_1\right)|\mathbf{W},\mathbf{Z}_1\right). \tag{A.1}$$

By **A.1**, $V_{1,k}$ does not depend on $\mathbf{W}, \mathbf{Z}_1$, so,

$$R_{1,k} = F_{V_{1,k}}\left(h_{1,k}^{-1}\left(\mathbf{W}, X_{1,k}, \mathbf{Z}_1\right)\right) = F_{V_{1,k}}\left(V_{1,k}\right), \tag{A.2}$$

where $F_{V_{1,k}}$ is unique by **A.2**, so $\mathbf{R}_1 = F_{\mathbf{V}_1}(\mathbf{V}_1)$, establishing Theorem 1(a) at time $t = 1$.

Furthermore, $\mathbf{X}_1$ is determined by $\mathbf{W}$, $\mathbf{Z}_1$, $\mathbf{V}_1$, and $(\mathbf{W}, \mathbf{Z}_1) \perp\!\!\!\perp U_1$, so any dependence between $\mathbf{X}_1$ and U_1 must operate through $\mathbf{V}_1$. Since Theorem 1(a) implies conditioning on $\mathbf{R}_1$ is equivalent to conditioning on $\mathbf{V}_1$, then $X_{1,k} \perp\!\!\!\perp U_1|, \mathbf{R}_1$, so Theorem 1(b) holds $t = 1$. A direct implication of Theorem 1(b) and **A.5** is,

$$\mathbb{E}[Y_1|\mathbf{W} = \mathbf{w}, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{R}_1 = \mathbf{r}_1] = \mathbb{E}[g_1(\mathbf{w}, \mathbf{x}_1, U_1, \mathbf{V}_1, \epsilon_1)|\mathbf{R}_1 = \mathbf{r}_1]. \tag{A.3}$$

Finally, using that the support of $\mathbf{R}_1$ is full and does not vary with $\mathbf{W}, \mathbf{X}_1$ by A.4, the equivalence between conditioning on $\mathbf{R}_1$ and $\mathbf{V}_1$ by Theorem 1(a), and that $\mathbb{E}[|Y|] < \infty$, the law of iterated expectations implies that,

$$\mathbb{E}[\mathbb{E}[Y_1|\mathbf{W} = \mathbf{w}, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{R}_1 = \mathbf{r}]] = \mathbb{E}[g_1(\mathbf{w}, \mathbf{x}_1, U_1, \mathbf{V}_1, \epsilon_1)], \tag{A.4}$$

establishing Theorem 1(c) at $t = 1$.

Step 2. *Proof that Theorem 1 holds at time t if it holds at time $t-1$.*

Suppose Theorem 1 holds at time $t-1$. Fix k. There is only one random unobservable $V_{t,k}$ in $h_{t,k}$ given $\mathbf{M}_t$, where we can condition on the unobservable components of $\mathbf{M}_t$, $\mathbf{V}^{(t-1)}$, because Theorem 1(a) holds at $t-1$. Since $X_{t,k}$ is monotonic in $V_{t,k}$ by **A.3**, $h_{t,k}$ can be partially inverted with respect to $V_{t,k}$ such that,

$$\begin{aligned} R_{t,k} &\equiv F_{X_{t,k}|\mathbf{W},\mathbf{Z}_t,\mathbf{M}_t}\left(X_{t,k}|\mathbf{W},\mathbf{Z}_t,\mathbf{M}_t\right) \\ &= F_{V_{t,k}|\mathbf{W},\mathbf{Z}_t,\mathbf{M}_t}\left(h_{t,k}^{-1}\left(X_{t,k},\mathbf{W},\mathbf{Z}_t,\mathbf{M}_t\right)|\mathbf{W},\mathbf{Z}_t,\mathbf{M}_t\right). \end{aligned} \tag{A.5}$$

By **A.1**, $V_{t,k}$ does not depend on $\mathbf{W},\mathbf{Z}_t$ given $\mathbf{M}_t$, so,

$$R_{t,k} = F_{V_{t,k}|\mathbf{M}_t}\left(V_{t,k}|\mathbf{M}_t\right), \tag{A.6}$$

where $F_{V_{t,k}|\mathbf{M}_t}$ is unique by **A.2**, so $\mathbf{R}_t = F_{\mathbf{V}_t|\mathbf{M}_t}(\mathbf{V}_t|\mathbf{M}_t)$, establishing Theorem 1(a) at time t.

Given $\mathbf{M}_t$, $\mathbf{X}_t$ is determined by $\mathbf{W}$, $\mathbf{Z}_t$, $\mathbf{V}_t$, and $(\mathbf{W},\mathbf{Z}_t) \perp\!\!\!\perp U_t|\mathbf{M}_t$, so any dependence between $\mathbf{X}_t$ and U_t must operate through $\mathbf{V}_t$ given $\mathbf{M}_t$. Since Theorem 1(a) implies conditioning on $\mathbf{R}_t$ is equivalent to conditioning on $\mathbf{V}_t$, then $X_{t,k} \perp\!\!\!\perp U_t|,\mathbf{R}_t,\mathbf{M}_t$, so Theorem 1(b) holds at t. A direct implication of Theorem 1(b) and **A.5** is,

$$\mathbb{E}[Y_t|\mathbf{W}=\mathbf{w},\mathbf{X}^{(t)}=\mathbf{x}^{(t)},\mathbf{R}^{(t)}=\mathbf{r}^{(t)}] = \mathbb{E}[g_t(\mathbf{w},\mathbf{x}^{(t)},\mathbf{V}^{(t)},U^{(t)},\epsilon_t)|\mathbf{V}^{(t)}=\mathbf{v}^{(t)}] \tag{A.7}$$

Finally, using that the support of $\mathbf{R}^{(t)}$ is full and does not vary with $\mathbf{W},\mathbf{X}^{(t)}$ by A.4, the equivalence between conditioning on $\mathbf{R}^{(t)}$ and $\mathbf{V}^{(t)}$ by Theorem 1(a), and that $\mathbb{E}[|Y_t|] < \infty$, the law of iterated expectations implies that,

$$\mathbb{E}[\mathbb{E}[Y_t|\mathbf{W}=\mathbf{w},\mathbf{X}^{(t)}=\mathbf{x}^{(t)},\mathbf{R}^{(t)}=\mathbf{r}]] = \mathbb{E}[g_t(\mathbf{w},\mathbf{x}^{(t)},U^{(t)},\mathbf{V}^{(t)},\epsilon_1)], \tag{A.8}$$

establishing Theorem 1(c) at t.

A.2 Proof of Theorem 2

This section proves that Theorem 2 holds at time $t=1$, and Theorem 2 holds at time t if it holds at time $t-1$. By induction, Theorem 2 holds for all t.

Step 1. *Proof of Theorem 2 at time $t = 1$.*

Notice that the triangular panel model at time $t = 1$ under the assumptions of Theorem 2 is nested by the triangular panel model at time $t = 1$ under the assumptions of Theorem 1, so the same proof holds from Step 1 in Appendix A.1. The only additional result at time $t = 1$ for Theorem 2 is $Q_t = F_{U_1|\mathbf{V}_1}(U_1|\mathbf{V}_1)$, which is now established:

The rank device Q_1 can be written,

$$Q_1 \equiv F_{Y_1|\mathbf{W},\mathbf{X}_1,\mathbf{R}_1}(Y_1|\mathbf{W},\mathbf{X}_1,\mathbf{R}_1), \tag{A.9}$$

where it is used that conditioning on $\mathbf{R}_1$ is equivalent to conditioning on $\mathbf{V}_1$. Because there is only one random unobservable U_1 in g_1 given $\mathbf{R}_1$, and Y_1 is monotonic in U_1 by **B.2**, g_1 can be partially inverted with respect to U_1 such that,

$$Q_1 = F_{U_1|\mathbf{W},\mathbf{X}_1,\mathbf{R}_1}(g_1^{-1}(\mathbf{W},\mathbf{X}_1,\mathbf{R}_1)\,|\mathbf{W},\mathbf{X}_1,\mathbf{R}_1) = F_{U_1|\mathbf{R}_1}(U_1|\mathbf{R}_1). \tag{A.10}$$

where the second equality uses $U_1 \perp\!\!\!\perp (\mathbf{W},\mathbf{X}_1)|\mathbf{R}_1$ by **A.1** and $F_{U_1|\mathbf{R}_1}$ is unique by **B.2**.

Step 2. *Proof that Theorem 2 holds at time t if it holds at time $t-1$.*

Suppose Theorem 2 holds at time $t-1$. Fix k. There is only one random unobservable $V_{t,k}$ in $h_{t,k}$ given $\mathbf{M}_t$, since conditioning on $\mathbf{M}_t$ implies conditioning on $U^{(t-1)}, \mathbf{V}^{(t-1)}$ by Theorem 2(a) at $t-1$. Since $X_{t,k}$ is monotonic in $V_{t,k}$ by **A.3**, $h_{t,k}$ can be partially inverted with respect to $V_{t,k}$ such that,

$$\begin{aligned} R_{t,k} &\equiv F_{X_{t,k}|\mathbf{W},\mathbf{Z}_t,\mathbf{M}_t}\left(X_{t,k}|\mathbf{W},\mathbf{Z}_t,\mathbf{M}_t\right) \\ &= F_{V_{t,k}|\mathbf{W},\mathbf{Z}_t,\mathbf{M}_t}\left(h_{t,k}^{-1}\left(X_{t,k},\mathbf{W},\mathbf{Z}_t,\mathbf{M}_t\right)|\mathbf{W},\mathbf{Z}_t,\mathbf{M}_t\right). \end{aligned} \tag{A.11}$$

By **A.1**, $V_{t,k}$ does not depend on $\mathbf{W},\mathbf{Z}_t$ given $\mathbf{M}_t$, so,

$$R_{t,k} = F_{V_{t,k}|\mathbf{M}_t}\left(V_{t,k}|\mathbf{M}_t\right), \tag{A.12}$$

where $F_{V_{t,k}|\mathbf{M}_t}$ is unique by **A.2**, so $\mathbf{R}_t = F_{\mathbf{V}_t|\mathbf{M}_t}(\mathbf{V}_t|\mathbf{M}_t)$, establishing the first part of Theorem 1(a) at time t.

Because there is only one random unobservable U_t in g_t given $\mathbf{W}, \mathbf{X}_t, \mathbf{R}_t, \mathbf{M}_t$ and Y_t is monotonic in U_t by **B.2**, g_t can be partially inverted with respect to U_t such that,

$$\begin{aligned} Q_t &= F_{U_t|\mathbf{W},\mathbf{X}_t,\mathbf{R}_t,\mathbf{M}_t}(g_t^{-1}\left(Y_t, \mathbf{W}, \mathbf{X}^{(t)}, U^{(t)}, \mathbf{V}^{(t)}, \epsilon_t\right)|\mathbf{W}, \mathbf{X}_t, \mathbf{R}_t, \mathbf{M}_t) \\ &= F_{U_t|\mathbf{R}_t,\mathbf{M}_t}(U_t|\mathbf{R}_t, \mathbf{M}_t), \end{aligned} \tag{A.13}$$

where the second equality uses $U_t \perp\!\!\!\perp (\mathbf{W}, \mathbf{X}_t)|\mathbf{R}_t, \mathbf{M}_t$ by **A.1** and $F_{U_t|\mathbf{R}_t,\mathbf{M}_t}$ is unique by **B.2**, establishing the second part of Theorem 1(a) at time t.

Given $\mathbf{M}_t$, $\mathbf{X}_t$ is determined by $\mathbf{W}$, $\mathbf{Z}_t$, $\mathbf{V}_t$, and $(\mathbf{W}, \mathbf{Z}_t) \perp\!\!\!\perp U_t|\mathbf{M}_t$, so any dependence between $\mathbf{X}_t$ and U_t must operate through $\mathbf{V}_t$ given $\mathbf{M}_t$. Since Theorem 1(a) implies conditioning on $\mathbf{R}_t$ is equivalent to conditioning on $\mathbf{V}_t$, then $X_{t,k} \perp\!\!\!\perp U_t|, \mathbf{R}_t, \mathbf{M}_t$, so Theorem 1(b) holds at t. A direct implication of Theorem 1(b) and **A.5** is,

$$\begin{aligned} &\mathbb{E}[Y_t|\mathbf{W} = \mathbf{w}, \mathbf{X}^{(t)} = \mathbf{x}^{(t)}, Q^{(t-1)} = q^{(t-1)}, \mathbf{R}^{(t)} = \mathbf{r}^{(t)}] \\ &= \mathbb{E}[g_t(\mathbf{w}, \mathbf{x}^{(t)}, \mathbf{V}^{(t)}, U^{(t)}, \epsilon_t)|Q^{(t-1)} = q^{(t-1)}, \mathbf{R}^{(t)} = \mathbf{r}^{(t)}]. \end{aligned} \tag{A.14}$$

Finally, using that the support of $Q^{(t-1)}, \mathbf{R}^{(t)}$ is full and does not vary with $\mathbf{W}, \mathbf{X}^{(t)}$ by A.4, the equivalence between conditioning on $Q^{(t-1)}, \mathbf{R}^{(t)}$ and $U^{(t-1)}, \mathbf{V}^{(t)}$ by Theorem 1(a), and that $\mathbb{E}[|Y_t|] < \infty$, the law of iterated expectations implies that,

$$\mathbb{E}[\mathbb{E}[Y_t|\mathbf{W} = \mathbf{w}, \mathbf{X}^{(t)} = \mathbf{x}^{(t)}, Q^{(t-1)}, \mathbf{R}^{(t)}]] = \mathbb{E}[g_t(\mathbf{w}, \mathbf{x}^{(t)}, U^{(t)}, \mathbf{V}^{(t)}, \epsilon_t)], \tag{A.15}$$

establishing Theorem 2(c) at $t = 1$.

A.3 Proof of Theorem 3

This section proves that Theorem 3 holds at time $t = 1$, and Theorem 3 holds at time t if it holds at time $t - 1$. By induction, Theorem 3 holds for all t.

Step 1. *Proof of Theorem 3 at time $t = 1$.*

Consider $t = 1$. Fix k. Note that the input equation can be written $X_{1,k} = h_{1,k}(\mathbf{M}_1, V_{1,k})$ and the inverse, if it exists, can be written $V_{1,k} = h_{1,k}^{-1}(X_{1,k}, \mathbf{M}_1)$. From the functional form in (4.1),

$$h_{1,k}^{-1}(X_{1,k}, \mathbf{M}_1) = \Pi_1^{-1}\left(\mathbb{E}[\mathbf{M}_1\mathbf{M}_1'|V_{1,k}]^{-1}\mathbb{E}[\mathbf{M}_1 X_{1,k}|V_{1,k}]\right). \tag{A.16}$$

Since the inverse of $\mathbb{E}[\mathbf{M}_1\mathbf{M}_1'|V_{1,k}]$ exists by **A.4***, and Π_1^{-1} exists uniquely by **A.3**, it follows that $h_{1,k}^{-1}$ exists and is unique. Thus, for any x in the support of $X_{1,k}$ and $\mathbf{m}$ in the support of $\mathbf{M}_1$, there exists unique v such that $v = h_{1,k}^{-1}(x, \mathbf{m})$. It follows that,

$$\int_0^1 \mathbb{1}[\mathbf{M}'\Pi(V_{1,k}) \leq x]|\mathbf{M}_1 = \mathbf{m}] = \int_0^1 \mathbb{1}[V_{1,k} \leq h_{1,k}^{-1}(x, \mathbf{M}_1)]|\mathbf{M}_1 = \mathbf{m}]. \tag{A.17}$$

Because $V_{1,k} \perp\!\!\!\perp \mathbf{M}_1$ by **A.1**,

$$\int_0^1 \mathbb{1}[V_{1,k} \leq h_{1,k}^{-1}(x, \mathbf{m})]|\mathbf{M}_1] = \int_0^1 \mathbb{1}[V_{1,k} \leq h_{1,k}^{-1}(x, \mathbf{m})] \equiv v, \tag{A.18}$$

so $R_{1,k}^*$ uniquely maps $(X_{1,k}, \mathbf{M}_1)$ to $V_{1,k}$, which establishes Theorem 3(a) at $t = 1$.

Furthermore, since $U_1 \perp\!\!\!\perp \mathbf{M}_1$, and $\mathbf{X}_1$ depends only on $\mathbf{M}_1$ and $\mathbf{V}_1$, any dependence between $\mathbf{X}_1$ and U_1 is due to $\mathbf{V}_1$, so $\mathbf{X}_1 \perp\!\!\!\perp U_1|\mathbf{R}_1^*$, where conditioning on $\mathbf{R}_1^*$ is equivalent to conditioning on $\mathbf{V}_1$ by Theorem 3(a) at $t = 1$. Finally, rearranging the output equation in (4.1) and again applying Theorem 3(a) at $t = 1$,

$$\mathbb{E}[\mathbb{E}[\mathbf{P}_1\mathbf{P}_1'|\mathbf{R}_1^* = \mathbf{r}_1]^{-1}\mathbb{E}[\mathbf{P}_1 Y_1|\mathbf{R}_1^* = \mathbf{r}]] = \mathbb{E}[\Psi_1(U_1, \mathbf{V}_1, \epsilon_1)|\mathbf{V}_1 = \mathbf{v}] \tag{A.19}$$

where $\mathbb{E}[\mathbf{P}_t\mathbf{P}_t'|\mathbf{R}_t^*]$ exists by **A.4***, and Theorem 3(c) follows immediately.

Step 2. *Proof that Theorem 3 holds at time t if it holds at time* $t-1$.

Fix t. Suppose Theorem 3 holds at time $t-1$, which, importantly, implies that $\mathbf{M}_t$ is known at t, where $\mathbf{M}_t$ includes $\mathbf{R}^{*(t-1,s)}$ as a component. Given $\mathbf{M}_t$, the proof of Theorem 3 at t is essentially identical to the proof of Theorem 3 at $t = 1$:

Fix k. Note that the input equation can be written $X_{t,k} = h_{t,k}(\mathbf{M}_t, V_{t,k})$ and the inverse, if it exists, can be written $V_{t,k} = h_{t,k}^{-1}(X_{t,k}, \mathbf{M}_t)$. From the functional form in (4.1),

$$h_{t,k}^{-1}(X_{t,k}, \mathbf{M}_t) = \Pi_t^{-1}\left(\mathbb{E}[\mathbf{M}_t\mathbf{M}_t'|V_{t,k}]^{-1}\mathbb{E}[\mathbf{M}_t X_{t,k}|V_{t,k}]\right). \tag{A.20}$$

Since the inverse of $\mathbb{E}[\mathbf{M}_t\mathbf{M}_t'|V_{t,k}]$ exists by **A.4***, and Π_t^{-1} exists uniquely by **A.3**, it follows that $h_{t,k}^{-1}$ exists and is unique. Thus, for any x in the support of $X_{t,k}$ and $\mathbf{m}$ in the support of $\mathbf{M}_t$, there exists unique v such that $v = F_{V_{t,k}|\mathbf{M}_t}(h_{t,k}^{-1}(x, \mathbf{M}_t)|\mathbf{m})$. It follows that,

$$\int_0^1 \mathbb{1}[\mathbf{M}'\Pi(V_{t,k}) \leq x]|\mathbf{M}_t = \mathbf{m}] = \int_0^1 \mathbb{1}[V_{t,k} \leq h_{t,k}^{-1}(x, \mathbf{M}_t)]|\mathbf{M}_t = \mathbf{m}] \equiv v, \tag{A.21}$$

so $R^*_{t,k}$ uniquely maps $(X_{t,k}, \mathbf{M}_t)$ to $V_{t,k}$, which establishes Theorem 3(a) at t.

Given $\mathbf{M}_t$, any dependence between U_t and $\mathbf{X}_t$ must operate through $\mathbf{V}_t$, so $\mathbf{X}_t \perp\!\!\!\perp U_t | \mathbf{R}^*_t, \mathbf{M}_t$, where conditioning on $\mathbf{R}^*_t$ is equivalent to conditioning on $\mathbf{V}_t$ by Theorem 3(a) at t. Finally, rearranging the output equation in (4.1), noticing that $\mathbb{E}[\mathbf{P}_t\mathbf{P}'_t|\mathbf{R}^*_t, \mathbf{M}_t] = \mathbb{E}[\mathbf{P}_t\mathbf{P}'_t|\mathbf{R}^*_t]$ by **A.2**, and again applying Theorem 3(a) at t,

$$\mathbb{E}[\mathbb{E}[\mathbf{P}_t\mathbf{P}'_t|\mathbf{R}^*_t = \mathbf{r}]^{-1}\mathbb{E}[\mathbf{P}_t Y_t|\mathbf{R}^*_t = \mathbf{r}]] = \mathbb{E}[\Psi_t(U_t, \mathbf{V}_t, \epsilon_t)|\mathbf{V}_t = \mathbf{v}] \tag{A.22}$$

where $\mathbb{E}[\mathbf{P}_t\mathbf{P}'_t|\mathbf{R}^*_t]$ exists by **A.4***, and Theorem 3(c) follows immediately.

A.4 Proof of Theorem 4

This section proves that Theorem 4 holds at time $t = 1$, and Theorem 4 holds at time t if it holds at time $t-1$. By induction, Theorem 4 holds for all t.

Step 1. *Proof of Theorem 4 at time $t = 1$.*

The model and assumptions at time $t = 1$ are identical to those of Theorem 3 at time $t = 1$, except for the second part of Theorem 4(a), which is now proven: Note that the output equation can be written $Y_1 = g_1(\mathbf{P}_1, U_{1,k}, \mathbf{V}_1)$ and the inverse, if it exists, can be written $U_1 = g_1^{-1}(Y_1, \mathbf{P}_1, \mathbf{V}_1)$. From the functional form in (4.1),

$$g_1^{-1}(Y_1, \mathbf{P}_1, \mathbf{V}_1) = \Psi_1^{-1}\left(\mathbb{E}[\mathbf{P}_1\mathbf{P}'_1|\mathbf{V}_1]^{-1}\mathbb{E}[\mathbf{P}_1\mathbf{P}'_1|\mathbf{V}_1\right). \tag{A.23}$$

Since the inverse of $\mathbb{E}[\mathbf{P}_1\mathbf{P}'_1|\mathbf{V}_1]$ exists by **A.4***, and Ψ_1^{-1} exists uniquely by **A.4**†, it follows that g_1^{-1} exists and is unique. Thus, for any y in the support of Y_1, $\mathbf{v}$ in the support of $\mathbf{V}_1$, and $\mathbf{p}$ in the support of $\mathbf{P}_1$, there exists unique u such that $u = F_{U_1|\mathbf{V}_1,\mathbf{P}_1}(g_1^{-1}(y, \mathbf{p}, \mathbf{v})|\mathbf{V}_1 = \mathbf{v}, \mathbf{P}_1 = \mathbf{p})$. It follows that,

$$\int_0^1 \mathbb{1}[\mathbf{P}'_1\Psi(U_1, \mathbf{V}_1) \leq y]|\mathbf{P}_1 = \mathbf{p}, \mathbf{V}_1 = \mathbf{v}] = \int_0^1 \mathbb{1}[U_1 \leq g_1^{-1}(y, \mathbf{p}, \mathbf{v})]|\mathbf{P}_1 = \mathbf{p}, \mathbf{V}_1 = \mathbf{v}] \equiv u, \tag{A.24}$$

so $Q_1^\dagger$ uniquely maps $(Y_1, \mathbf{P}_1, \mathbf{R}_1)$ to U_1, which establishes Theorem 4(a) at $t = 1$.

Step 2. *Proof that Theorem 4 holds at time t if it holds at time $t - 1$.*

Fix t. Suppose Theorem 4 holds at time $t-1$, which, importantly, implies that $\mathbf{M}_t$ and $\mathbf{P}_t$ are known at t. Given $\mathbf{M}_t$ and $\mathbf{P}_t$, the proofs of the first part of Theorem 4(a) as well as Theorems 4(b) and 4(c) are identical to those of Theorem 3 at t, so it only remains to prove the second part of Theorem 4(a). The proof that $Q_t^\dagger$ uniquely maps $(Y_t, \mathbf{P}_t, \mathbf{R}_t)$ to U_t for t given $\mathbf{P}_t, \mathbf{R}_t$ is essentially identical to the proof at $t = 1$.

Note that the output equation can be written $Y_t = g_t(\mathbf{P}_t, U_t, \mathbf{V}_t)$ and the inverse, if it exists, can be written $U_t = g_1^{-1}(Y_t, \mathbf{P}_t, \mathbf{V}_t)$. From the functional form in (4.1),

$$g_t^{-1}(Y_t, \mathbf{P}_t, \mathbf{V}_t) = \Psi_t^{-1}\left(\mathbb{E}[\mathbf{P}_t\mathbf{P}_t'|\mathbf{V}_t]^{-1}\mathbb{E}[\mathbf{P}_t\mathbf{P}_t'|\mathbf{V}_t]\right). \tag{A.25}$$

Since the inverse of $\mathbb{E}[\mathbf{P}_t\mathbf{P}_t'|\mathbf{V}_t]$ exists by **A.4***, and Ψ_t^{-1} exists uniquely by **A.4**†, it follows that g_t^{-1} exists and is unique. Thus, for any y in the support of Y_t, $\mathbf{v}$ in the support of $\mathbf{V}_t$, and $\mathbf{p}$ in the support of $\mathbf{P}_t$, there exists unique u such that $u = F_{U_t|\mathbf{V}_t,\mathbf{P}_t}(g_t^{-1}(y, \mathbf{p}, \mathbf{v})|\mathbf{V}_t = \mathbf{v}, \mathbf{P}_t = \mathbf{p})$. It follows that,

$$\int_0^1 \mathbb{1}[\mathbf{P}_t'\Psi(U_t, \mathbf{V}_t) \leq y]|\mathbf{P}_t = \mathbf{p}, \mathbf{V}_t = \mathbf{v}] = \int_0^1 \mathbb{1}[U_t \leq g_t^{-1}(y, \mathbf{p}, \mathbf{v})]|\mathbf{P}_t = \mathbf{p}, \mathbf{V}_t = \mathbf{v}] \equiv u, \tag{A.26}$$

so $Q_t^\dagger$ uniquely maps $(Y_t, \mathbf{P}_t, \mathbf{R}_t)$ to U_t, which establishes Theorem 4(a) at t.