

Under-Identification of Structural Models Based on Timing and Information Set Assumptions

Daniel Akerberg

Department of Economics
University of Texas at Austin
Austin, 78712, USA

Garth Frazer

Rotman School of Management
University of Toronto
Toronto, M5S 3E6, Canada

Kyoo il Kim

Department of Economics
Michigan State University
East Lansing, 48824, USA

Yao Luo

Department of Economics
University of Toronto
Toronto, M5S 3G7, Canada

Yingjun Su*

Institute for Economic and Social Research
Jinan University
Guangzhou, 510632, China

October 2020

Abstract

We revisit identification based on timing and information set assumptions in structural models, which have been used in the context of production functions, demand equations, and hedonic pricing models (e.g. Olley and Pakes (1996), Blundell and Bond (2000)). First, we demonstrate a general under-identification problem using these assumptions, illustrating this with a simple version of the Blundell-Bond dynamic panel model. In particular, the basic moment conditions can yield multiple discrete solutions: one at the persistence parameter in the main equation and another at the persistence parameter governing the regressor. Second, we propose possible solutions based on sign restrictions and an augmented moment approach. We show the identification of our approach and propose a consistent estimation procedure. Our Monte Carlo simulations illustrate the under-identification issue and finite sample performance of our proposed estimator. Lastly, we show that the problem persists in many alternative models of the regressor but disappears in some models under stronger assumptions.

Keywords: Production Function, Identification, Timing and Information Set Assumptions, Market Persistence Factor, Monte Carlo Simulation

JEL Classifications: C14, C18, D24

*Corresponding author: Yingjun Su, Associate Professor, Institute for Economic and Social Research, Jinan University, Guangzhou, China; Tel: 86-13809776419; Email: yis17@pitt.edu. All errors are our own.

1 Introduction

A number of recent papers address endogeneity in structural models using a panel data identification strategy based on timing and information set assumptions. A preeminent example is estimation of production functions with endogenous input choice - in this context, timing and information set assumptions are used for identification in two distinct lines of literature: estimation methods using a proxy variable approach (e.g. Olley and Pakes (1996), Levinsohn and Petrin (2003), and Akerberg, Caves, and Frazer (ACF, 2015)), and estimation based on what are often called dynamic panel methods (e.g. Blundell and Bond (2000)). The methods proposed in these literatures have been used in thousands of papers, both in the context of production functions and in many other structural models. In this paper, we study an interesting identification problem that can arise in these methods. This under-identification issue has been pointed out more narrowly in prior papers (see footnote 16 of ACF and Kim, Luo, and Su (KLS, 2019)) which concern the proxy variable approach. In this paper, we illustrate how this is a more widespread issue in estimation methods that rely on timing and information set assumptions. In particular, we show that the under-identification problem can also arise in the context of dynamic panel methods. Because dynamic panel methods have a relatively simple representation, we then use this dynamic panel context to further study the under-identification problem and suggest possible solutions.

The identification problem we study is unusual. While typical identification problems involve regions of non-identification, this identification problem does not - in our simple models, the moment conditions are zero at the true parameters and non-zero locally around the true parameters, but also at a second distinct point in the parameter space. So, the model might be described as being locally identified (around the true parameters) but not globally identified. There is a quadratic aspect to the model and moment conditions that leads to the two possible solutions. This form of the non-identification also impacts some of the solutions we propose for the issue - in particular it means that we can leverage sign restrictions to resolve it.

Before proceeding, we discuss a couple of important issues. First, we illustrate the under-identification problem with perhaps the simplest possible dynamic panel model. In particular, our basic model does not include fixed effects, which are relatively common in the empirical literature using these techniques. This is mainly to aid in our initial illustration of the problem and solutions. Later in the paper, we show that the under-identification problem also exists in dynamic panel models with fixed effects.

Second, much of our discussion is couched in the production function context. This is because production functions are perhaps the most frequent empirical context for the application of these identification strategies based on timing and information set assumptions. However,

it is important to note that there are many empirical applications of dynamic panel methods (and timing/information set assumptions more generally) in other scenarios. One interesting example is a set of recent work on demand estimation with potential endogenous product characteristics, e.g. Sweeting (2009), Grennan (2013), Lee (2013), and Sullivan (2017). Another example is some recent work on hedonic price regression models with a time-varying unobserved attribute that is correlated with observed attributes and follows a persistent time-series process, e.g. Bajari, Fruehwirth, Kim, and Timmins (2012) and Bishop and Murphy (2019). The under-identification issues and potential solutions we propose are potentially relevant for all this work as well.

The paper proceeds by first illustrating the under-identification problem in what is perhaps the simplest possible dynamic panel model. We then discuss very explicit potential solutions to the problem, in part based on specifying and estimating a model for the endogenous right-hand-side variables (inputs in the production function context) and making a sign restriction. We illustrate how these solutions can work in practice using Monte Carlo experiments. We then broaden our scope to consider the under-identification problem more generally, showing how it also manifests itself in more general models (e.g. dynamic panel models with fixed effects, different data generating processes for the right-hand-side endogenous variables, and models with multiple right-hand-side endogenous variables). We end by using lessons from our specific solutions for the simple example to propose more general practical advice for researchers using these methods in practice (applicable in both proxy variable and dynamic panel approaches) - in an effort to avoid these under-identification problems.

2 The Under-Identification Problem

Consider a panel regression model

$$y_{it} = \alpha^0 + \beta^0 x_{it} + \omega_{it} + \eta_{it} \tag{1}$$

where x_{it} is correlated with error term ω_{it} but not with η_{it} . Our main context is production function estimation, in which y_{it} is the output, x_{it} is an input, ω_{it} is a serially correlated unobserved productivity shock, and η_{it} is, e.g. measurement error in output that is often assumed to be uncorrelated across time.

Because inputs are endogenously chosen by firms, we want to allow x_{it} to be correlated with productivity shock ω_{it} . To capture this, we assume the following data generating process (DGP) for x_{it} ,

$$x_{it} = \pi^0 + \theta^0 \omega_{it} + \kappa_{it}, \tag{2}$$

which allows firm choice of x_{it} to depend on productivity shock ω_{it} as well as another persistent market factor κ_{it} .

We allow both ω_{it} and κ_{it} to follow first-order autoregressive processes

$$\omega_{it} = \rho_{\omega}^0 \omega_{i,t-1} + \xi_{it} \quad (3)$$

and

$$\kappa_{it} = \rho_x^0 \kappa_{i,t-1} + u_{it}, \quad (4)$$

where ξ_{it} and u_{it} are the innovation components of these processes. If x_{it} is labor input, then κ_{it} can be interpreted as a wage factor that is persistent over time.¹ A simple identification restriction that has been used in the production function literature involves assuming the following timing and information set conditions.

Assumption 1 $E[\xi_{i,t+1}|I_{it}] = E[u_{i,t+1}|I_{it}] = 0$, where I_{it} represents firm i 's information set at t and includes x_{it} up to t and y_{it} up to $t - 1$.

This is a simple form of a “timing and information set” assumption. Loosely speaking, it can be interpreted as following from an economic model where (i) firms do not observe ω_{it} and κ_{it} until t , and (ii) firms choose x_{it} at t - thereby implying (with some additional, more technical assumptions) that future innovations in ω_{it} and κ_{it} , e.g. $\xi_{i,t+1}$ and $u_{i,t+1}$, are uncorrelated with x_{it} . Note how this is consistent with (2), where x_{it} is chosen as a function of ω_{it} and κ_{it} .²

For estimation, one also needs to make an assumption on the measurement error in output, e.g. $E[\eta_{it}|I_{it}] = 0$. Given this, one can consider estimating the model parameters $(\alpha, \beta, \rho_{\omega})$ by ρ_{ω} -differencing the production function

$$\begin{aligned} y_{it} - \rho_{\omega}^0 y_{i,t-1} &= \alpha^0 (1 - \rho_{\omega}^0) + \beta^0 (x_{it} - \rho_{\omega}^0 x_{i,t-1}) + \omega_{it} - \rho_{\omega}^0 \omega_{i,t-1} + \eta_{it} - \rho_{\omega}^0 \eta_{i,t-1} \\ &= \alpha^0 (1 - \rho_{\omega}^0) + \beta^0 (x_{it} - \rho_{\omega}^0 x_{i,t-1}) + \xi_{it} + \eta_{it} - \rho_{\omega}^0 \eta_{i,t-1}. \end{aligned}$$

Since, by assumption, ξ_{it} , η_{it} , and $\eta_{i,t-1}$ are all conditional mean independent of $I_{i,t-1}$, it follows

¹Serial correlation in productivity and/or input price shocks is widely allowed in the literature. Empirical studies that find strong persistence in firm input prices include Alonse-Borrego and Arellano (1999) and Grieco, Li, and Zhang (2016). But it is worth noting that while we allow ω_{it} and κ_{it} to be serially correlated, we do not require it.

²Also note why this is not only a timing assumption, but also an assumption about firms' information sets at various points in time. For example, consider an alternative information set assumption where firms observe the shocks ω_{it} and κ_{it} *one period ahead of time*, i.e. $\omega_{i,t+1}$ and $\kappa_{i,t+1}$ are observed at t . Since, in general, firm choice of inputs x_{it} may have dynamic implications (e.g. capital, labor with any sort of adjustment costs), those choices could optimally depend on $\omega_{i,t+1}$ and $\kappa_{i,t+1}$, implying that $\xi_{i,t+1}$ and $u_{i,t+1}$ would no longer be uncorrelated with x_{it} . On the other hand, suppose $\omega_{i,t+1}$ and $\kappa_{i,t+1}$ continue to be observed at t , except with an alternative timing assumption where firms must commit to x_{it} at $t - 1$. This case brings us back to $\xi_{i,t+1}$ and $u_{i,t+1}$ being uncorrelated with x_{it} once again.

that

$$E[(y_{it} - \rho_\omega y_{i,t-1}) - \alpha(1 - \rho_\omega) - \beta(x_{it} - \rho_\omega x_{i,t-1}) | I_{i,t-1}] = 0 \quad \text{at } (\alpha, \beta, \rho_\omega) = (\alpha^0, \beta^0, \rho_\omega^0). \quad (5)$$

This is a very simple example of a type of dynamic panel moment condition that might be used to estimate a production function (see, e.g., Blundell and Bond (2000) and Akerberg, Caves, and Frazer (2015) for a discussion of how the dynamic panel approach is related to proxy variable approaches). Note that this model is simpler than those used in many applications of dynamic panel methods. In particular, dynamic panel approaches often include a fixed effect α_i in the production function. This typically requires a second differencing to form usable moment conditions.³ Our simplified model (5) excludes fixed effect for expositional purposes. Later, we show that our under-identification result also applies to a model with a fixed effect.

We know moment condition (5) equals zero at true parameters values. Here, we show that this moment condition generally has an additional zero - one where $\rho_\omega = \rho_x^0$, i.e. where the productivity persistence parameter is not equal to its true value, but instead equal to the persistence parameter of the *input* process. This identification issue has been noted in the context of the proxy variable literature on estimating production functions (see footnote 16 of ACF and discussions in Kim, Luo, and Su (2019)). KLS (2019) studies this “pseudo”-solution using Monte Carlo simulations in the context of the proxy variable literature. In this paper, we show that this under-identification is a more general issue that also applies to the dynamic panel literature. Studying the issue in the dynamic panel approach is particularly useful because the framework is simpler in some respects, and this allows us to better understand the nature of the under-identification problem. This allows us to propose more general approaches that we think might help mitigate the problem.

To see the pseudo-solution in the current model, consider moment condition (5), where ρ_ω

³Moreover, many dynamic panel models include lagged y (i.e. models with state dependence) instead of the serially correlated ω in our model. The nature of these alternative models is somewhat different. Including lagged y directly in the structural model means the persistence in y is causal (lagged y causes current y), while in the models we study, the persistence originates from the persistent unobserved state ω . The models we study are used more often in the production function literature (as well as the demand and hedonic literature cited in the introduction), as there is typically more concern about serially correlated unobserved productivity shocks than true state dependent phenomena (although there are exceptions, e.g. a production function with learning by doing has causal state dependence).

is incorrectly set at ρ_x^0 , α is incorrectly set at $\alpha^0 - \frac{\pi^0}{\theta^0}$, and β is incorrectly set at $\beta^0 + \frac{1}{\theta^0}$:

$$\begin{aligned}
& E[(y_{it} - \rho_\omega y_{i,t-1}) - \alpha(1 - \rho_\omega) - \beta(x_{it} - \rho_\omega x_{i,t-1}) | I_{i,t-1}] \\
&= E[(y_{it} - \rho_x^0 y_{i,t-1}) - \left(\alpha^0 - \frac{\pi^0}{\theta^0}\right) (1 - \rho_x^0) - \left(\beta^0 + \frac{1}{\theta^0}\right) (x_{it} - \rho_x^0 x_{i,t-1}) | I_{i,t-1}] \\
&= E[\alpha^0(1 - \rho_x^0) + \beta^0(x_{it} - \rho_x^0 x_{i,t-1}) + \omega_{it} - \rho_x^0 \omega_{i,t-1} + \eta_{it} - \rho_x^0 \eta_{i,t-1} \\
&\quad - \left(\alpha^0 - \frac{\pi^0}{\theta^0}\right) (1 - \rho_x^0) - \left(\beta^0 + \frac{1}{\theta^0}\right) (x_{it} - \rho_x^0 x_{i,t-1}) | I_{i,t-1}] \\
&= E\left[\frac{\pi^0}{\theta^0}(1 - \rho_x^0) - \frac{1}{\theta^0}(x_{it} - \rho_x^0 x_{i,t-1}) + \omega_{it} - \rho_x^0 \omega_{i,t-1} + \eta_{it} - \rho_x^0 \eta_{i,t-1} | I_{i,t-1}\right] \\
&= E\left[\frac{\pi^0}{\theta^0}(1 - \rho_x^0) - \frac{1}{\theta^0}((1 - \rho_x^0) \pi^0 + \theta^0 (\omega_{it} - \rho_x^0 \omega_{i,t-1}) + (\kappa_{it} - \rho_x^0 \kappa_{i,t-1})) \right. \\
&\quad \left. + \omega_{it} - \rho_x^0 \omega_{i,t-1} + \eta_{it} - \rho_x^0 \eta_{i,t-1} | I_{i,t-1}\right] \\
&= E\left[-\frac{1}{\theta^0} (\kappa_{it} - \rho_x^0 \kappa_{i,t-1}) + \eta_{it} - \rho_x^0 \eta_{i,t-1} | I_{i,t-1}\right] \\
&= E\left[-\frac{1}{\theta^0} u_{it} + \eta_{it} - \rho_x^0 \eta_{i,t-1} | I_{i,t-1}\right] \\
&= 0,
\end{aligned}$$

where the second and fourth equalities come from substituting in the true DGP from (1) and (2) respectively. This derivation shows that, at this alternative parameter vector $(\alpha, \beta, \rho) = (\alpha^0 - \frac{\pi^0}{\theta^0}, \beta^0 + \frac{1}{\theta^0}, \rho_x^0)$, our simple dynamic panel moment condition essentially recovers the innovation in the input demand shock κ_{it} (i.e. u_{it}) rather than the innovation in the productivity shock ω_{it} (i.e. ξ_{it}). So given that we assume innovation in κ_{it} has the same properties as innovation in ω_{it} , our moment condition also equals zero at this pseudo-solution. This is perhaps the simplest illustration of the under-identification issue we study, though we develop more intuition for it momentarily.

Because the two zeros of the moment condition arise at different values of ρ , one may think that the identification issue does not arise when $\rho_\omega^0 = \rho_x^0$. While this is not the focus of the paper, we show that in this case the identification problem is even worse in that there is an entire region of non-identification. To see this argument from the input demand equation (2), note that if $\rho_\omega^0 = \rho_x^0 = \rho^0$, then

$$\begin{aligned}
x_{it} - \rho^0 x_{i,t-1} &= \pi^0(1 - \rho^0) + \theta^0(\omega_{it} - \rho^0 \omega_{i,t-1}) + (\kappa_{it} - \rho^0 \kappa_{i,t-1}) \\
&= \pi^0(1 - \rho^0) + \theta^0 \xi_{it} + u_{it}.
\end{aligned}$$

In other words, when $\rho_\omega^0 = \rho_x^0$, the differencing $x_{it} - \rho x_{i,t-1}$ dispenses with all remaining information in x_{it} other than the innovation terms, which are orthogonal to $I_{i,t-1}$. It follows that,

for an arbitrary (α, β) , moment equation (5) at $(\alpha, \beta, \rho_\omega) = (\alpha, \beta, \rho^0)$ satisfies

$$\begin{aligned}
0 &= E[(y_{it} - \rho^0 y_{i,t-1}) - \alpha(1 - \rho^0) - \beta(x_{it} - \rho^0 x_{i,t-1}) | I_{i,t-1}] \\
&= E[(\alpha^0 - \alpha)(1 - \rho^0) + (\beta^0 - \beta)(x_{it} - \rho^0 x_{i,t-1}) + \xi_{it} + \eta_{it} - \rho^0 \eta_{i,t-1} | I_{i,t-1}] \\
&= E[(\alpha^0 - \alpha)(1 - \rho^0) + (\beta^0 - \beta)(\pi^0(1 - \rho^0) + \theta^0 \xi_{it} + u_{it}) + \xi_{it} + \eta_{it} - \rho^0 \eta_{i,t-1} | I_{i,t-1}] \\
&= (\alpha^0 - \alpha)(1 - \rho^0) + (\beta^0 - \beta)[\pi^0(1 - \rho^0)],
\end{aligned}$$

where the second and third equalities come from substituting in the true DGP from (1) and (2), respectively, for which we impose $\rho_\omega^0 = \rho_x^0 = \rho^0$. The last result holds since, given the timing and information set assumptions, the unobservables ξ_{it} , u_{it} , η_{it} , and $\eta_{i,t-1}$ are orthogonal to the information available at time $t - 1$. In other words, the moment equation has arbitrarily many solutions of (α, β) that satisfy

$$(\alpha^0 - \alpha)(1 - \rho^0) + (\beta^0 - \beta)[\pi^0(1 - \rho^0)] = 0.$$

Therefore, moment condition (5) does not contain any information on β when $\rho_\omega^0 = \rho_x^0$. Intuitively this is similar to the non-identification result for time-invariant variables in a fixed effects panel model. In any case, our focus is on the existence of the pseudo-solution that arises from moment condition (5) in a general setting. Therefore, for the remainder of the paper, we avoid this more severe problem by assuming:

Assumption 2 $\rho_\omega^0 \neq \rho_x^0$,

which, again, is analogous to ruling out a time-invariant x_{it} in a fixed effects model.

Returning to our main point, when $\rho_\omega^0 \neq \rho_x^0$, we have verified that the basic moment condition of the dynamic panel approach may fail identification by yielding two solutions: one at the true AR(1) parameter of the productivity shock and the other at the AR(1) parameter of the input demand shock. In the next section, we consider a possible identification approach that utilizes additional moments based on the DGP of x_{it} . Our approach also benefits from the fact that, in the production function context (and likely in many other contexts), one may be willing to make assumptions about the sign of θ , which can be seen as the direction of endogeneity.

3 Identification Using the Input Process

In this section, we investigate an augmented approach that aids identification by adding additional moment conditions based on the input equation (2). This equation (and its unobserved input factor κ_{it}) depends on three additional parameters, π , θ , and ρ_x , so we are now considering estimating an econometric model with six parameters $(\alpha, \beta, \pi, \theta, \rho_\omega, \rho_x)$.

To derive a second moment condition, we can express the input demand equation as

$$\begin{aligned}
x_{it} &= \pi^0 + \theta^0 \omega_{it} + \kappa_{it} \\
&= \pi^0 + \theta^0 (y_{it} - \alpha^0 - \beta^0 x_{it} - \eta_{it}) + \kappa_{it} \\
&= \left(\frac{1}{\beta^0 \theta^0 + 1} \right) (\pi^0 - \theta^0 \alpha^0 + \theta^0 y_{it} - \theta^0 \eta_{it} + \kappa_{it}).
\end{aligned} \tag{6}$$

Since the persistent input factor κ_{it} is correlated with the various observables, a natural strategy is to ρ_x -difference this equation to isolate its innovation component (just like what is done with ω_{it} to estimate the production function). This results in

$$x_{it} - \rho_x^0 x_{i,t-1} = \left(\frac{1}{\beta^0 \theta^0 + 1} \right) ((1 - \rho_x^0) (\pi^0 - \theta^0 \alpha^0) + \theta^0 (y_{it} - \rho_x^0 y_{i,t-1}) - \theta^0 (\eta_{it} - \rho_x^0 \eta_{i,t-1}) + u_{it}).$$

With the model's timing/information set assumptions, the unobserved term in this equation, $\left(\frac{1}{\beta^0 \theta^0 + 1} \right) (-\theta^0 (\eta_{it} - \rho_x^0 \eta_{i,t-1}) + u_{it})$, is orthogonal to $I_{i,t-1}$, i.e. to $x_{i,t-1}, y_{i,t-2}$ and further lags of these variables. Therefore, one can construct a moment condition

$$E[(x_{it} - \rho_x x_{i,t-1}) - \left(\frac{1}{\beta \theta + 1} \right) ((1 - \rho_x) (\pi - \theta \alpha) + \theta (y_{it} - \rho_x y_{i,t-1})) | I_{i,t-1}] = 0, \tag{7}$$

which equals zero at $(\alpha, \beta, \pi, \theta, \rho_\omega, \rho_x) = (\alpha^0, \beta^0, \pi^0, \theta^0, \rho_\omega^0, \rho_x^0)$. Intuitively, (6) (and the moment derived from it, i.e. (7)) might be described as providing information on θ^0 through relating x_{it} to the residual from the production function, which contains the productivity shock. Given this, one might think of combining this second moment with the original moment (5) derived in the previous section, and using both moments together to estimate $\Phi = (\alpha, \beta, \pi, \theta, \rho_\omega, \rho_x)$.

However, the under-identification problem also manifests itself here. To see this, multiply (7) by $-\frac{\beta\theta+1}{\theta} = -(\beta + \frac{1}{\theta})$, move the terms around to form

$$e_2(\Phi) \equiv E[(y_{it} - \rho_x y_{i,t-1}) - (1 - \rho_x) \left(\alpha - \frac{\pi}{\theta} \right) - \left(\beta + \frac{1}{\theta} \right) (x_{it} - \rho_x x_{i,t-1}) | I_{i,t-1}] = 0,$$

and compare this version of the new moment, $e_2(\Phi)$, to the original moment (5):

$$e_1(\Phi) \equiv E[(y_{it} - \rho_\omega y_{i,t-1}) - \alpha(1 - \rho_\omega) - \beta(x_{it} - \rho_\omega x_{i,t-1}) | I_{i,t-1}] = 0.$$

Interestingly, while moments $e_1(\Phi)$ and $e_2(\Phi)$ are distinct in terms of parameters (and thus contain distinct information on the true parameters), their structure is *identical* in terms of observables $y_{it}, y_{i,t-1}, x_{it}$, and $x_{i,t-1}$. What are the implications of this?

First, this provides an alternative proof of our original under-identification result. Because the structure of the two moments is so similar, the fact that the new moment $e_2(\Phi)$ equals zero at the true parameters $\Phi^0 = (\alpha^0, \beta^0, \pi^0, \theta^0, \rho_\omega^0, \rho_x^0)$ clearly implies that the original moment $e_1(\Phi)$ must equal zero at a pseudo-parameter vector that is different from the truth, i.e. $(\alpha, \beta, \rho_\omega) = \left(\alpha^0 - \frac{\pi^0}{\theta^0}, \beta^0 + \frac{1}{\theta^0}, \rho_x^0\right)$.

More importantly at this point, this also shows that simply adding moment $e_2(\Phi)$ does not solve the under-identification problem. In particular, if β is set to $\beta^0 + \frac{1}{\theta^0}$ to “incorrectly” zero moment $e_1(\Phi)$ at the pseudo-solution, then one can also set $\theta = -\theta^0$, which means that the coefficient on $(x_{it} - \rho_x x_{i,t-1})$ in $e_2(\Phi)$ equals

$$\beta + \frac{1}{\theta} = \beta^0 + \frac{1}{\theta^0} + \frac{1}{-\theta^0} = \beta^0,$$

which can help zero $e_2(\Phi)$ as well (because this is the true coefficient on $(x_{it} - \rho_\omega x_{i,t-1})$ in $e_1(\Phi)$). In fact, examination of $e_2(\Phi)$ and $e_1(\Phi)$ together implies there is a *joint* pseudo solution to *both* moments, where

$$\Phi^P = (\alpha^0 - \pi^0/\theta^0, \beta^0 + 1/\theta^0, \pi^0, -\theta^0, \rho_x^0, \rho_\omega^0). \quad (8)$$

Essentially, this pseudo-solution is simply reversing the moments (which is possible because of the similar structure in terms of observables). The first moment equals zero at the pseudo-parameters, because those pseudo-parameters make it equivalent to the second moment at the true parameters. And similarly, the second moment equals zero at the pseudo-parameters because those pseudo-parameters make it equivalent to the first moment at the true parameters.

While this shows that simply adding a second moment condition based on the input choice equation does not in itself solve the under-identification problem, examination of the pseudo-solution (8) does lead to two important conclusions that we rely on in our proposed solutions. First, note that at the pseudo-solution, coefficient θ is negated, i.e. $\theta^P = -\theta^0$. In other words, if in the true DGP, firm choice of input level x_{it} depends positively on productivity shock ω_{it} , then in the pseudo-DGP described by the pseudo-parameters, firm choice of input level x_{it} depends *negatively* on productivity shock ω_{it} . This observation can be helpful at distinguishing true parameters from pseudo-parameters as, in many cases, one may be willing to *a priori* sign (e.g. in the production function context, firm profit maximization would likely imply $\theta^0 > 0$; in other econometric contexts, one might be willing to sign the endogeneity problem (e.g. Manski and Pepper (2000))). Second, note that pseudo-parameter $\beta^P = \beta^0 + 1/\theta^0$. This also implies that, if we know the sign of θ^0 , we know whether β^P is higher or lower than β^0 .

3.1 Discussion

Before discussing some additional restrictions of the model, we highlight two additional points. First, while we focus on a sign restriction as a crucial part of trying to address the under-identification problem, there are other possible approaches. In particular, an economic model might make functional restrictions between the six parameters $(\alpha, \beta, \pi, \theta, \rho_\omega, \rho_x)$. Production functions are a particularly good example of this, as if perfectly competitive firms choose x_{it} to statically profit maximize (where κ_{it} is proportional the (log) cost of a unit of the input), then firm FOC implies that $\theta = \frac{1}{1-\beta}$. Imposing this restriction would resolve the under-identification problem in the joint moment conditions $e_1(\Phi)$ and $e_2(\Phi)$. This is directly related to work in the production function literature that imposes FOC from profit maximization to make inferences on production function parameters, e.g. Gandhi, Navarro, and Rivers (2020). We do not pursue this approach here as we want to instead investigate broader approaches to addressing the under-identification problem. In many examples where dynamic panel methods are used, one might be unwilling to make such cross-equation restrictions. Even in the production function case, one might be unwilling to assume perfectly competitive static profit maximization (or one might want to allow the objective function to depend on other factors that imply $\theta \neq \frac{1}{1-\beta}$).⁴

Second, note that Φ^P is actually not the only pseudo-solution to the joint moment conditions $e_1(\Phi)$ and $e_2(\Phi)$. In particular, consider

$$\Phi^{\infty 1} = (\alpha^0, \beta^0, \textit{anything}, \infty, \rho_\omega^0, \rho_\omega^0)$$

and

$$\Phi^{\infty 2} = (\alpha^0 - \pi^0/\theta^0, \beta^0 + 1/\theta^0, \textit{anything}, \infty, \rho_x^0, \rho_x^0).$$

Note that these additional spurious solutions are both obtained by setting $\theta = \infty$. In doing this, $e_1(\Phi)$ and $e_2(\Phi)$ essentially become the same equation, since the coefficient on the x terms are the same, i.e. $\beta^0 = \beta^0 + 1/\theta^0$. Intuitively, the spurious solutions $\Phi^{\infty 1}$ and $\Phi^{\infty 2}$ are not giving the two solutions we need to distinguish between – instead, they are finding the same solution twice. Furthermore, it is not clear whether this solution is the truth or Φ^P . This issue is what leads us to study additional moment restrictions implied by the model.

3.2 Additional Moment Restrictions

To further aid in addressing the under-identification problem, we derive some additional equations generated by our model. First, we further work with input demand function (6) to derive

⁴For example, if productivity shock ω_{it} is allowed to impact the unit cost of input, then cross-equation restriction $\theta = \frac{1}{1-\beta}$ would not hold.

a reduced form process by which x_{it} evolves over time. Given that we can invert out the unobserved productivity term as $\omega_t = \frac{x_t - \kappa_t - \pi^0}{\theta^0}$, we can use the AR(1) process for ω_{it} and (3), by replacing ω_t with $\frac{x_t - \kappa_t - \pi^0}{\theta^0}$. By rearranging terms, we obtain

$$\begin{aligned}
x_{it} &= \pi^0 + \theta^0 \omega_{it} + \kappa_{it} \\
&= \pi^0 + \theta^0 (\rho_\omega^0 \omega_{i,t-1} + \xi_{it}) + \rho_x^0 \kappa_{i,t-1} + u_{it} \\
&= \pi^0 + \theta^0 \rho_\omega^0 \left(\frac{x_{i,t-1} - \kappa_{i,t-1} - \pi^0}{\theta^0} \right) + \theta^0 \xi_{it} + \rho_x^0 \kappa_{i,t-1} + u_{it} \\
&= (1 - \rho_\omega^0) \pi^0 + \rho_\omega^0 x_{i,t-1} + (\rho_x^0 - \rho_\omega^0) \kappa_{i,t-1} + u_{it} + \theta^0 \xi_{it}, \tag{9}
\end{aligned}$$

where for the second equality, we utilize the AR(1) process of κ_t , i.e., (4). We then apply the ρ_x -differencing again to remove $\kappa_{i,t-1}$ as follows.

$$\begin{aligned}
x_{it} &= (1 - \rho_\omega^0) \pi^0 + \rho_\omega^0 x_{i,t-1} + (\rho_x^0 - \rho_\omega^0) (\rho_x^0 \kappa_{i,t-2} + u_{i,t-1}) + u_{it} + \theta^0 \xi_{it} \\
&= (1 - \rho_\omega^0) \pi^0 + \rho_\omega^0 x_{i,t-1} + u_{it} + \theta^0 \xi_{it} \\
&\quad + (\rho_x^0 - \rho_\omega^0) \left\{ \rho_x^0 \frac{1}{(\rho_x^0 - \rho_\omega^0)} (x_{i,t-1} - (1 - \rho_\omega^0) \pi^0 - \rho_\omega^0 x_{i,t-2} - u_{i,t-1} - \theta^0 \xi_{i,t-1}) + u_{i,t-1} \right\} \\
&= (1 - \rho_\omega^0) \pi^0 + \rho_\omega^0 x_{i,t-1} \\
&\quad + \rho_x^0 (x_{i,t-1} - (1 - \rho_\omega^0) \pi^0 - \rho_\omega^0 x_{i,t-2} - u_{i,t-1} - \theta^0 \xi_{i,t-1}) + (\rho_x^0 - \rho_\omega^0) u_{i,t-1} + u_{it} + \theta^0 \xi_{it} \\
&= (1 - \rho_\omega^0) (1 - \rho_x^0) \pi^0 + (\rho_\omega^0 + \rho_x^0) x_{i,t-1} - \rho_x^0 \rho_\omega^0 x_{i,t-2} + u_{it} - \rho_\omega^0 u_{i,t-1} - \rho_x^0 \theta^0 \xi_{i,t-1} + \theta^0 \xi_{it}, \tag{10}
\end{aligned}$$

where in the first equality we replace $\kappa_{i,t-1}$ with $\rho_x^0 \kappa_{i,t-2} + u_{i,t-1}$ using (4), and in the second equality, we replace $\kappa_{i,t-2}$ in terms of x using (9). This result shows that x_t indeed follows a type of ARIMA(2,1) process, with two moving average terms, u_t and ξ_t .

Because we know the properties of the unobserved terms in (10), we can consider estimating the parameters. While our model implies that x_{t-1} is correlated with both u_{t-1} and ξ_{t-1} , one could consider using an IV estimation strategy using x_{t-2} and x_{t-3} as instruments. This regression would identify π^0 , and two composite parameters, $\psi_1^0 = \rho_\omega^0 + \rho_x^0$ and $\psi_2^0 = -\rho_x^0 \rho_\omega^0$ that are functions of ρ_x^0 and ρ_ω^0 . However, despite there being these two composite parameters in two unknowns, they do not uniquely identify ρ_x^0 and ρ_ω^0 . Essentially, this is another representation of the under-identification problem from above.

Interestingly, however, this new equation (10) does add something when combined with moments $e_1(\Phi)$ and $e_2(\Phi)$ studied above. Recall that these two moments have two additional pseudo-solutions, $\Phi^{\infty 1}$ and $\Phi^{\infty 2}$, where $\Phi^{\infty 1}$ involves $\rho_\omega = \rho_x = \rho_\omega^0$, and $\Phi^{\infty 2}$ involves $\rho_\omega = \rho_x = \rho_x^0$. We now show that enforcing the additional information in (10) *rules out both of these pseudo-solutions*, i.e. given that (10) identifies ψ_1^0 and ψ_2^0 (asymptotically), the two-equation

system

$$\begin{aligned}\psi_1^0 &= \rho_\omega + \rho_x \\ \psi_2^0 &= -\rho_x \rho_\omega\end{aligned}$$

has *no solutions* where the two ρ parameters are equal.

First, recall that we already assume that $\rho_\omega^0 \neq \rho_x^0$ (as explained in the previous section, an absence of this assumption would give us even more severe identification problems⁵). Next, note that if the two unknowns are equal in the above system, i.e., $\rho_x = \rho_\omega$, we would need $\psi_2^0 = -[\frac{\psi_1^0}{2}]^2$. On the other hand, the true parameters satisfy $\psi_1^0 = \rho_x^0 + \rho_\omega^0$ and $\psi_2^0 = -\rho_x^0 \rho_\omega^0$. Combining these two conditions gives $-\rho_x^0 \rho_\omega^0 = -[\frac{\rho_x^0 + \rho_\omega^0}{2}]^2$, which is equivalent to $(\rho_x^0 - \rho_\omega^0)^2 = 0$, a contradiction to our maintained assumption. Therefore, forcing the two coefficients in (10) to equal $\rho_\omega + \rho_x$ and $-\rho_x \rho_\omega$, respectively, rules out $\Phi^{\infty 1}$ (for which we should have $\rho_\omega = \rho_x = \rho_\omega^0$) and $\Phi^{\infty 2}$ (for which we should have $\rho_\omega = \rho_x = \rho_x^0$).

While (10) is useful for understanding the under-identification and excluding $\Phi^{\infty 1}$ and $\Phi^{\infty 2}$, using it in the estimation would make directly estimating this ARIMA(2,1) equation impractical. Due to the endogeneity of regressors in the ARIMA(2,1) process, we need to use instruments further lagged than $I_{i,t-1}$, such as x_{t-2} and x_{t-3} . This may lead to imprecise estimates because further lagged instruments are typically less informative.⁶ To this end we do not implement the additional moment based on the above. Instead, we form a similar moment as follows. First, invert (9) to obtain

$$\kappa_{i,t-1} = \frac{1}{(\rho_x^0 - \rho_\omega^0)} (x_{it} - (1 - \rho_\omega^0)\pi^0 - \rho_\omega^0 x_{i,t-1} - u_{it} - \theta^0 \xi_{it}),$$

which implies

$$\kappa_{it} = \rho_x^0 \kappa_{i,t-1} + u_{it} = \frac{\rho_x^0}{(\rho_x^0 - \rho_\omega^0)} [x_{it} - (1 - \rho_\omega^0)\pi^0 - \rho_\omega^0 x_{i,t-1} - u_{it} - \theta^0 \xi_{it}] + u_{it}. \quad (11)$$

Also we plug ω_t , inverted from the input demand equation (2), into the production function and obtain

$$y_{it} = \alpha^0 + \beta^0 x_{it} + \frac{x_{it} - \pi^0 - \kappa_{it}}{\theta^0} + \eta_{it} = (\alpha^0 - \pi^0/\theta^0) + (\beta^0 + 1/\theta^0)x_{it} - \kappa_{it}/\theta^0 + \eta_{it}. \quad (12)$$

⁵Relatedly, when $\rho_\omega^0 = \rho_x^0$, (9) shows that x_{it} follows only an AR(1) process and hence one (asymptotically) obtains a zero coefficient on term $x_{i,t-2}$ if one estimates the ARIMA(2,1) equation (10). Therefore, at least in theory, one could test whether $\rho_\omega^0 = \rho_x^0$ holds in the true DGP. In other words, the very fact that one (asymptotically) obtains a non-zero coefficient on $x_{i,t-2}$ rules out the case $\rho_\omega^0 = \rho_x^0$.

⁶Utilizing the ARIMA(2,1) equation (10), our Monte Carlo experiments show poor estimates of the ρ parameters. These results are available upon request.

Then, combining (11) and (12), we obtain

$$\begin{aligned}
y_{it} &= (\alpha^0 - \pi^0/\theta^0) + (\beta^0 + 1/\theta^0)x_{it} \\
&\quad - \left\{ \frac{\rho_x^0}{\theta^0(\rho_x^0 - \rho_\omega^0)} [x_{it} - (1 - \rho_\omega^0)\pi^0 - \rho_\omega^0 x_{i,t-1} - u_{it} - \theta^0 \xi_{it}] + u_{it}/\theta^0 \right\} + \eta_{it} \\
&= \gamma_0^0 + \gamma_1^0 x_{it} + \gamma_2^0 x_{i,t-1} + e_{it},
\end{aligned} \tag{13}$$

where $\gamma_0^0 = \alpha^0 + \frac{\pi^0 \rho_\omega^0 (1 - \rho_x^0)}{\theta^0 (\rho_x^0 - \rho_\omega^0)}$, $\gamma_1^0 = [(\beta^0 + 1/\theta^0) - \frac{\rho_x^0}{\theta^0 (\rho_x^0 - \rho_\omega^0)}] = \beta^0 - \frac{\rho_\omega^0}{\theta^0 (\rho_x^0 - \rho_\omega^0)}$, $\gamma_2^0 = \frac{\rho_x^0 \rho_\omega^0}{\theta^0 (\rho_x^0 - \rho_\omega^0)}$ and $e_{it} = \frac{\rho_\omega^0 u_{it} + \theta^0 \rho_x^0 \xi_{it}}{\theta^0 (\rho_x^0 - \rho_\omega^0)} + \eta_{it}$. Since the unobservables in e_t are all mean independent of $I_{i,t-1}$, we can form the following moment condition

$$e_3(\Phi) \equiv E[y_{it} - \gamma_0 - \gamma_1 x_{it} - \gamma_2 x_{i,t-1} | I_{i,t-1}] = 0.$$

In summary, we have derived three moments, $e_1(\Phi)$, $e_2(\Phi)$, and $e_3(\Phi)$ – one based on the production function, one based on the input demand function, and one based on a higher order time series properties of these functions. Note how estimation using the additional moment $e_3(\Phi)$ implicitly helps impose $\rho_\omega \neq \rho_x$; when ρ_ω approaches ρ_x , the composite coefficients γ_0 , γ_1 , and γ_2 become unbounded. As we have discussed previously, imposing $\rho_\omega \neq \rho_x$ also naturally rules out pseudo-solutions $\Phi^{\infty 1}$ and $\Phi^{\infty 2}$. Then, combined with a sign restriction on θ^0 , this should resolve the under-identification problem, and we implement this computationally in the next section.

4 Joint Estimation

In this section, we propose a joint estimation approach that estimates all parameters in the benchmark model described in Section 2. It is worth noting that this estimator relies on the specifics in this model and builds on our identification result in Section 3. Specifically, it exploits the known distance between the two solutions, the known sign of endogeneity, i.e. θ , and the additional moment restrictions derived from a specific input choice equation. Later on, in Section 7, we discuss some different ways in which one might take the intuition from this section (as well as previous sections) to more general models. In the simple model, we start with an initial estimate of β from the dynamic panel approach, which could be either the true or the spurious solution in finite samples, and then finalize the estimate by two simple searches along the direction of θ . To facilitate our global minimization, we concentrate out $(\rho_\omega, \rho_x, \alpha, \pi)$ and estimate (θ, β) jointly.

We first illustrate our estimation strategy in Figure 1, and then formally develop our estimation procedure below. In this illustration, we assume only β is unknown. The red solid

line represents the GMM objective function using the usual dynamic panel approach. The blue dashed line represents our proposed GMM objective function by augmenting the moment conditions generated from the input demand process. In the first step, we use the dynamic panel approach (red solid line) to find an initial estimate of β , say $\tilde{\beta}$. As suggested in Section 2, estimate $\tilde{\beta}$ can be either true parameter β (point A in the figure), or spurious solution $\beta + \frac{1}{\theta}$ (point B). Since we cannot immediately tell whether $\tilde{\beta}$ equals β or $\beta + \frac{1}{\theta}$, we hypothesize it as point A (true minimum) or B (spurious minimum) and try to minimize the proposed GMM objective function in the direction that is suggested by the A or B hypotheses. We have two possible cases.

In the first case, the dynamic panel approach yields point A, i.e., $\tilde{\beta} = \beta$. 1) If we hypothesize it as A (the true solution), we would fix the β estimate to be $\tilde{\beta}$ and search for θ . In this case, we seek no further improvement in the direction of β and stay at the minimum of the blue line, point A. 2) If we hypothesize it as B (the spurious solution), we would fix the $\beta + \frac{1}{\theta}$ estimate to be $\tilde{\beta}$ and search for $\theta > 0$. Or equivalently, we search for β on the left-hand side of point A. In this case, the objective function is strictly higher than point A for any $\theta > 0$. Either way, our approach yields point A if the dynamic panel approach does.

In the second case, the dynamic panel approach yields point B, i.e., $\tilde{\beta} = \beta + \frac{1}{\theta}$. 1) If we hypothesize it as A (the true solution), we would fix the β estimate to be $\tilde{\beta}$ and search for θ . In this case, we seek no further improvement in the direction of β and stay at point B' in the blue line, straight above point B. 2) If we hypothesize it as B (the spurious solution), we would fix the $\beta + \frac{1}{\theta}$ estimate to be $\tilde{\beta}$ and search for $\theta > 0$. Or equivalently, we search for β on the left-hand side of point B. In this case, the proposed objective function (blue dashed line) is minimized at point A. By comparing the two hypotheses (points B' and A), we again arrive at point A. Therefore, our approach yields point A even if the dynamic panel approach yields point B (the spurious minimum).

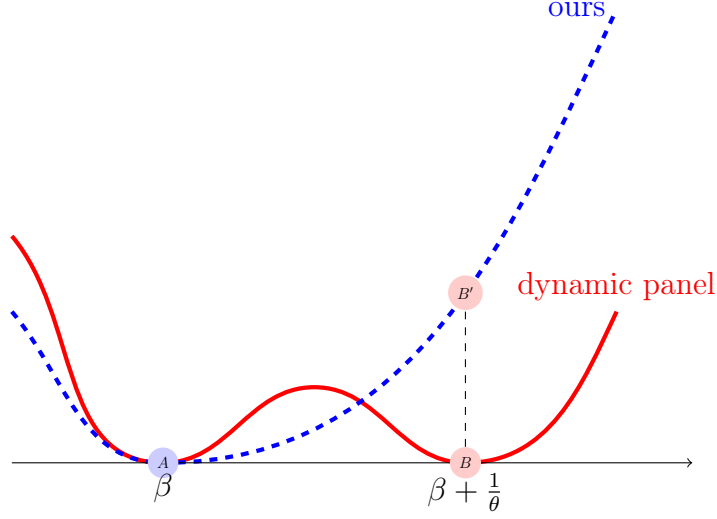


Figure 1: Illustration of the Proposed Estimator

Now, we formally describe the above estimation procedure in four steps.

Step 1: Obtain an initial estimate of β from the dynamic panel approach. Take the ρ_ω differencing of y_t and y_{t-1} to get

$$y_t - \beta x_t = \rho_\omega(y_{t-1} - \beta x_{t-1}) + (1 - \rho_\omega)\alpha + \xi_t + \eta_t - \rho_\omega\eta_{t-1}.$$

Fix a trial value of β , and construct both $y_t - \beta x_t$ and $y_{t-1} - \beta x_{t-1}$ to run the above regression. While y_{t-1} and η_{t-1} are correlated, parameter ρ_ω , profiled by β (i.e. for a given β), can be consistently estimated using IV regression. For example, we can use $y_{t-2} - \beta x_{t-2}$ as the instrument for $y_{t-1} - \beta x_{t-1}$ to obtain consistent estimates $\widehat{\rho}_\omega$ and $\widehat{\alpha} = \frac{(1 - \widehat{\rho}_\omega)\alpha}{1 - \widehat{\rho}_\omega}$, profiled by β . To obtain a moment condition for β from these profiled estimates, we can construct residual e_t^1 as

$$e_t^1 = \xi_t + \eta_t - \rho_\omega\eta_{t-1} = y_t - \beta x_t - \widehat{\rho}_\omega(y_{t-1} - \beta x_{t-1}) - (1 - \widehat{\rho}_\omega)\widehat{\alpha}, \quad (14)$$

and construct moment condition (i) as:

$$E[e_t^1 | I_{t-1}] = 0,$$

where we can take, e.g., $I_{t-1} = x_{t-1}$ or $I_{t-1} = (x_{t-1}, x_{t-2})$. Then, search over β to minimize the GMM objective function based on moment condition (i). This gives us $\widetilde{\beta}$, which is either the true solution or the spurious solution. Step 1 can be regarded as a concentrated dynamic panel approach.

Step 2: Estimate θ by hypothesizing the initial estimate of β as the true solution. Assume that $\widetilde{\beta}$ is the true parameter, i.e. $\widetilde{\beta} = \beta^0$. Fix $\beta = \widetilde{\beta}$, take a trial $\theta > 0$, and stack up

the three moments constructed following Step 2(a) and 2(b) below, i.e.,

$$E[(e_t^1, e_t^2, e_t^3)|I_{t-1}] = 0, \quad (15)$$

where we can again take $I_{t-1} = x_{t-1}$ or include further lagged instruments. Minimize the proposed GMM objective function, based on the moments (15), by searching over θ . The starting values of θ can be a grid search on R_+ , i.e. imposing a sign restriction. We then obtain estimates $(\widehat{\beta}_A = \widetilde{\beta}, \widehat{\theta}_A)$ under (true minimum) hypothesis A.

2(a) Given θ , construct both $y_t - (\widetilde{\beta} + 1/\theta)x_t$ and $y_{t-1} - (\widetilde{\beta} + 1/\theta)x_{t-1}$ to run another regression. In particular, take the ρ_x differencing of y_t and y_{t-1} and obtain

$$y_t - (\widetilde{\beta} + 1/\theta)x_t = \rho_x(y_{t-1} - (\widetilde{\beta} + 1/\theta)x_{t-1}) + (1 - \rho_x)(\alpha - \pi/\theta) - u_t/\theta + \eta_t - \rho_x\eta_{t-1}.$$

Parameter ρ_x , profiled by θ , can be consistently estimated with an IV regression by using $y_{t-2} - (\widetilde{\beta} + 1/\theta)x_{t-2}$ as the instrument for $y_{t-1} - (\widetilde{\beta} + 1/\theta)x_{t-1}$. As in Step 1, we obtain $\widehat{\rho}_x$, $\widehat{\pi} = \theta(\widehat{\alpha} - \frac{(1-\widehat{\rho}_x)(\alpha - \pi/\theta)}{1-\widehat{\rho}_x})$, and residual e_t^2 as

$$e_t^2 = -u_t/\theta + \eta_t - \widehat{\rho}_x\eta_{t-1} = y_t - (\widetilde{\beta} + 1/\theta)x_t - \widehat{\rho}_x(y_{t-1} - (\widetilde{\beta} + 1/\theta)x_{t-1}) - (1 - \widehat{\rho}_x)(\widehat{\alpha} - \widehat{\pi}/\theta),$$

and construct moment condition (ii) for θ as:

$$E[e_t^2|I_{t-1}] = 0.$$

2(b) Again, for a given θ , construct residual e_t^3 from (13) as

$$e_t^3 = \frac{\rho_\omega u_t + \theta \rho_x \xi_t}{\theta(\rho_x - \rho_\omega)} + \eta_t = y_t - \gamma_0 - \gamma_1 x_t - \gamma_2 x_{t-1},$$

where $\gamma_0 = \alpha + \frac{\pi \rho_\omega (1 - \rho_x)}{\theta(\rho_x - \rho_\omega)}$, $\gamma_1 = \beta - \frac{\rho_\omega}{\theta(\rho_x - \rho_\omega)}$, and $\gamma_2 = \frac{\rho_x \rho_\omega}{\theta(\rho_x - \rho_\omega)}$. Plugging in the profiled estimates $\widetilde{\beta}$, $\widehat{\rho}_\omega$, $\widehat{\rho}_x$, $\widehat{\alpha}$, $\widehat{\pi}$ from Step 1 and 2(a) to obtain e_t^3 , we can construct our moment condition (iii) as a function of θ :

$$E[e_t^3|I_{t-1}] = 0.$$

Step 3: Estimate θ by hypothesizing the initial estimate β as the spurious solution.

Assume that $\widetilde{\beta}$ is the spurious solution, i.e. $\widetilde{\beta} = \beta^0 + 1/\theta^0$. Fix $\beta = \widetilde{\beta} - 1/\theta$ profiled by θ . Repeat Step 2 (with $\widetilde{\beta} - 1/\theta$ replacing $\widetilde{\beta}$ in the above description) and obtain another set of estimates $(\widehat{\beta}_B = \widetilde{\beta} - 1/\widehat{\theta}_B, \widehat{\theta}_B)$ under hypothesis B.

Step 4: Final Estimate

Compare the two values of the GMM objective function obtained in Step 2 and 3, respectively: the estimates associated with the smaller objective function value is our final estimate.

5 Monte Carlo Simulation

Our design of the simulations is twofold. First, we conduct experiments to demonstrate the identification issues of the original dynamic panel approach. Second, we adopt our joint estimation approach as discussed in Section 4 and demonstrate its performance. Our DGP is based on the dynamic panel setup described in Section 2.

For the true parameter values, we set $\beta^0 = 0.6$, $\theta^0 = 1$, $\rho_x^0 = 0.5$, $\rho_\omega^0 = 0.7$, and $\alpha^0 = 0$, $\pi^0 = 0$. The i.i.d shocks ξ_{it} , u_{it} and η_{it} are normally distributed: $\xi_{it} \sim \mathcal{N}(0, 1)$, $u_{it} \sim \mathcal{N}(0, 1)$ and $\eta_{it} \sim \mathcal{N}(0, 0.1^2)$. To construct the productivity series and the persistent input factor from the AR(1) processes (3) and (4), we set the initial values $\omega_{i0} = 0$ and $\kappa_{i0} = 0$. Based on the setup above, we simulate 1,000 panel datasets, each with 1,000 firms and 10 time periods.⁷

5.1 Original Dynamic Panel Procedure

We follow the dynamic panel literature (Blundell and Bond (2000), Akerberg, Caves, and Frazer (2015)) and adopt the concentrated moment approach for estimation described by Step 1 in Section 4. From residual function (14), we use the following moment condition:

$$E[e_{it}^1 \times I_{i,t-1}] = 0$$

for GMM estimation, where $I_{i,t-1}$ denotes information available at time $t - 1$. We use two lags of x as our instruments, i.e., $I_{i,t-1} = (x_{i,t-1}, x_{i,t-2})'$.⁸

Our main point is to illustrate how the dynamic panel procedure sometimes ends up near the true parameters and sometimes ends up near the pseudo-parameter. For each simulated dataset we randomly pick a starting value of β from the uniform distribution on $[0.5, 1.5]$ and estimate β by minimizing the objective function using *fminunc* in Matlab. Figure 2 clearly illustrates the two distinctive modes of the dynamic panel – one is located around the true value of 0.6, and the other is around the spurious minimum of 1.6. Specifically, 31.8% of the

⁷To minimize the influence of the random initial values of ω_{i0} and κ_{i0} , we generate 40 periods of data and select the last 10 periods of data for our experiments.

⁸In both the dynamic approach and our joint estimation approach, we use $x_{i,t-1}$ and $x_{i,t-2}$ as “instruments” to form unconditional moments, as the overidentified model appeared to help with numeric stability - important when doing many Monte Carlo replications. Because of its simplicity, and because the scale of our moments is similar, we simply use the identity matrix to weight the moments. One important note in practice is that, in our joint estimation procedure, the same weighting matrix must be used for both Step 2 and Step 3 in order for comparison in Step 4 to be valid.

Table 1: Dynamic Panel and Joint Estimation Approaches

Panel A: Dynamic Panel Approach					
Coefficient	# of Replications	Mean	Std Dev	Min	Max
β	1,000	0.918	0.475	0.226	1.836
Percentage of spurious minimum					
$\beta > 1$				31.8%	
Panel B: Joint Estimation Approach					
Coefficient	# of Replications	Mean	Std Dev	Min	Max
β	1,000	0.597	0.075	0.226	0.765

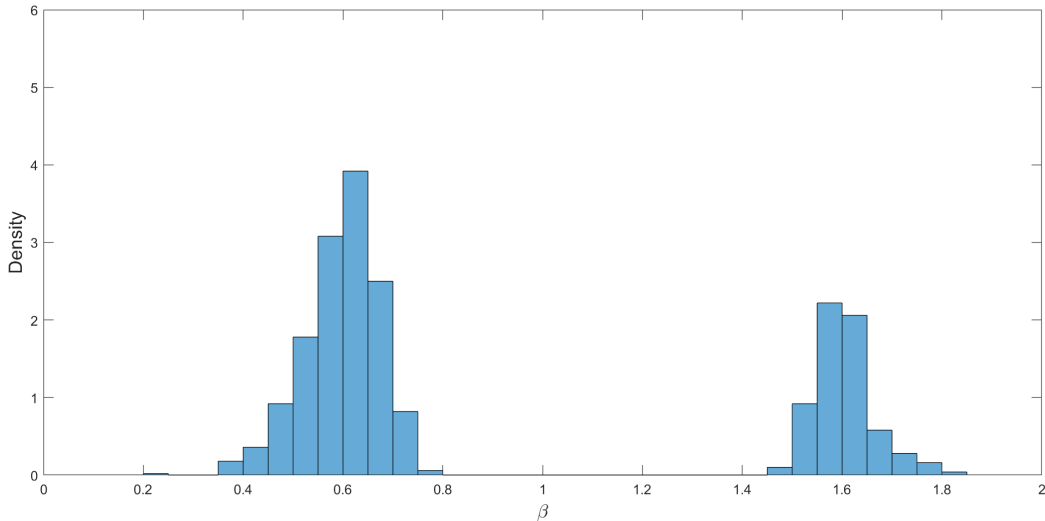


Figure 2: Dynamic Panel Approach

replications converge to the spurious minimum. As a consequence, the mean of the 1,000-simulation estimate of β is biased upward, i.e. 0.918, and the standard deviation is equal to 0.475 (see Panel A of Table 1).

5.2 Experiments with the Joint Estimation Approach

Next, we adopt our joint estimation approach to jointly estimate (β, θ) , as described in Section 4. We first obtain an initial estimate of β from the dynamic panel approach, as in the prior section with a starting value randomly selected from $[0.5, 1.5]$. This means that the initial estimate of β is sometimes near the true parameter, and sometimes near the pseudo-parameter. In each of Stages 2 and 3, we search over θ using a grid of starting values between 0 and 2 (in increments of 0.2). We impose a positivity constraint on θ by searching over $\gamma = -\log(\theta)$ ⁹

In this experiment, we see clearly from Figure 3 that the distribution of the estimated parameter β only has a single mode, near the true minimum of 0.6. Panel B of Table 1 shows that the experiment yields a close-to-unbiased estimate of β : the mean across the 1,000 replications is 0.597. Compared to the estimates from the dynamic panel approach (Panel A of Table 1), the standard deviation of the estimated β by this new approach also declines significantly to 0.075. Clearly, our proposed joint estimation approach outperforms the original dynamic panel approach in that it appears to avoid the spurious minimum.

⁹Note that if we used a grid for the starting point in the original DP procedure, we would typically find multiple zeros (or “close to” zeros) - one near the true parameter, and one near the pseudo-parameter. But since the DP objective function cannot distinguish between them, this would not help in and of itself.

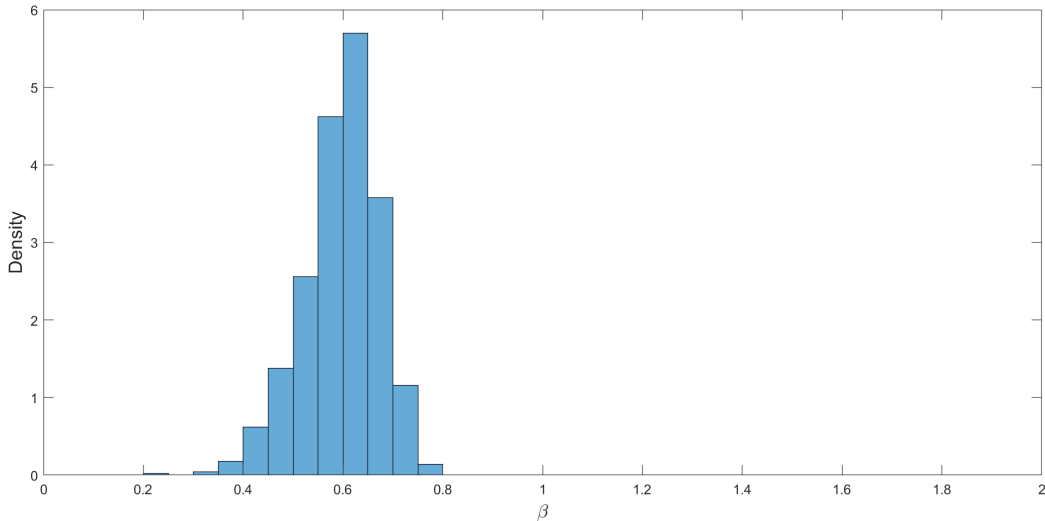


Figure 3: Joint Estimation Approach

5.3 DGPs with varying parameters

5.3.1 Varying ρ_x^0

In this experiment, we modify the DGPs by varying the values of ρ_x^0 from 0.3 to 0.6 with a step size of 0.1.¹⁰ For each DGP, we again simulate 1,000 panel datasets of 1,000 firms and 10 time periods. The remaining parameters are set as in the baseline DGP. We experiment with both the dynamic panel and our proposed joint estimation approach, again with x_{t-1} and x_{t-2} as the instruments for the moment conditions. Table 2 shows that our proposed approach outperforms the dynamic panel approach by avoiding the spurious solution, which continues to be problematic for the dynamic panel approach (in 26.0% to 37.1% of the 1,000 simulation runs).

5.3.2 Varying θ^0

In this experiment, we modify the DGPs by varying the values of θ^0 from 0.6 to 1 with a step size of 0.1. Examining (2), note that the degree of endogeneity grows as θ^0 gets larger. Other than θ^0 , the setup is the same as above. The results again show that the proposed approach consistently outperforms the dynamic panel approach by avoiding the spurious minimum. For

¹⁰We do not experiment with $\rho_x^0 = 0.7 = \rho_\omega^0$ for which the dynamic panel moment has more severe non-identification with infinitely many solutions, as discussed in Section 2. One would expect deteriorating performances of both the dynamic panel approach and the augmented moment approach as ρ_x^0 becomes closer to ρ_ω^0 . In other words, if we were in (or near) this knife-edge case, one would obtain very large standard errors regardless of the estimation procedure used.

Table 2: Simulation Results for Varying ρ_x^0

		Estimated β ($\beta^0 = 0.6$)				
		Dynamic Panel			Joint Estimation	
ρ_x^0	# of Replications	Mean	Std Dev	Percentage of $\beta > 1$	Mean	Std Dev
0.3	1,000	0.862	0.449	26.0%	0.604	0.079
0.4	1,000	0.876	0.457	28.1%	0.592	0.059
0.5	1,000	0.954	0.487	35.6%	0.594	0.077
0.6	1,000	0.963	0.514	37.1%	0.582	0.167

Table 3: Simulation Results for Varying θ^0

		Estimated β ($\beta^0 = 0.6$)				
		Dynamic Panel			Joint Estimation	
θ^0	# of Replications	Mean	Std Dev	Percentage of $\beta > 1$	Mean	Std Dev
0.6	1,000	0.847	0.597	14.7%	0.602	0.061
0.7	1,000	0.915	0.606	22.0%	0.597	0.062
0.8	1,000	0.901	0.542	24.2%	0.600	0.067
0.9	1,000	0.957	0.526	32.3%	0.596	0.073
1	1,000	0.918	0.475	31.8%	0.597	0.075

this experiment as well, the spurious solutions continue to exist for the dynamic panel approach (in 14.7% to 32.3% of the 1,000 simulation runs). However, the effect of varying θ^0 on the performance of the dynamic panel approach is non-trivial since θ^0 influences both the degree of endogeneity and the location of the spurious minimum.

6 Extensions

As noted earlier, our above model is a very simplified version of the traditional dynamic panel model. We now show how the under-identification problem also exists in some more general models.

6.1 Model with Fixed Effects

In particular, traditional dynamic panel models include fixed effects in the main equation, e.g.

$$y_{it} = \alpha_i + \beta^0 x_{it} + \omega_{it} + \eta_{it}. \quad (16)$$

Here, we show that this model also has the under-identification problem.

In model (16), forming usable moments requires an additional difference to remove the fixed effect α_i . More specifically, the ρ_ω^0 -difference

$$y_{it} - \rho_\omega^0 y_{i,t-1} = \alpha_i(1 - \rho_\omega^0) + \beta^0(x_{it} - \rho_\omega^0 x_{i,t-1}) + \xi_{it} + \eta_{it} - \rho_\omega^0 \eta_{i,t-1}.$$

contains α_i , so an additional first difference is typically done to obtain

$$\begin{aligned} & (y_{it} - \rho_\omega^0 y_{i,t-1}) - (y_{i,t-1} - \rho_\omega^0 y_{i,t-2}) \\ &= \beta^0 ((x_{it} - \rho_\omega^0 x_{i,t-1}) - (x_{i,t-1} - \rho_\omega^0 x_{i,t-2})) + \xi_{it} - \xi_{i,t-1} + (\eta_{it} - \rho_\omega^0 \eta_{i,t-1}) - (\eta_{i,t-1} - \rho_\omega^0 \eta_{i,t-2}) \end{aligned}$$

Since this eliminates α_i , one can proceed to estimate β^0 with a moment condition based on this, e.g.

$$E[(y_{it} - \rho_\omega y_{i,t-1}) - (y_{i,t-1} - \rho_\omega y_{i,t-2}) - \beta((x_{it} - \rho_\omega x_{i,t-1}) - (x_{i,t-1} - \rho_\omega x_{i,t-2})) | I_{i,t-2}] = 0. \quad (17)$$

To show that this moment condition may also have the under-identification problem, again consider a DGP for input x_{it} . Since the DGP for y_{it} includes a fixed effect, it is natural to include one here as well, e.g.

$$x_{it} = \pi_i + \theta^0 \omega_{it} + \kappa_{it}.$$

Following the previous argument, substitute the inverted production function into

$$\begin{aligned} x_{it} &= \pi_i + \theta^0 (y_{it} - \alpha_i - \beta^0 x_{it} - \eta_{it}) + \kappa_{it} \\ &= \left(\frac{1}{\beta^0 \theta^0 + 1} \right) (\pi_i - \theta^0 \alpha_i + \theta^0 y_{it} - \theta^0 \eta_{it} + \kappa_{it}) \end{aligned}$$

and ρ_x^0 -difference to obtain

$$x_{it} - \rho_x^0 x_{i,t-1} = \left(\frac{1}{\beta^0 \theta^0 + 1} \right) ((1 - \rho_x^0) (\pi_i - \theta^0 \alpha_i) + \theta^0 (y_{it} - \rho_x^0 y_{i,t-1}) - \theta^0 (\eta_{it} - \rho_x^0 \eta_{i,t-1}) + u_{it}).$$

First differencing this equation results in

$$\begin{aligned} & (x_{it} - \rho_x^0 x_{i,t-1}) - (x_{i,t-1} - \rho_x^0 x_{i,t-2}) \\ &= \left(\frac{\theta^0}{\beta^0 \theta^0 + 1} \right) ((y_{it} - \rho_x^0 y_{i,t-1}) - (y_{i,t-1} - \rho_x^0 y_{i,t-2})) \\ & \quad - \left(\frac{\theta^0}{\beta^0 \theta^0 + 1} \right) ((\eta_{it} - \rho_x^0 \eta_{i,t-1}) - (\eta_{i,t-1} - \rho_x^0 \eta_{i,t-2})) + \left(\frac{1}{\beta^0 \theta^0 + 1} \right) (u_{it} - u_{i,t-1}), \end{aligned}$$

and given that the unobservable terms in this equation are assumed to be mean independent of $I_{i,t-2}$, this implies

$$E[(y_{it} - \rho_x^0 y_{i,t-1}) - (y_{i,t-1} - \rho_x^0 y_{i,t-2}) - \left(\beta^0 + \frac{1}{\theta^0} \right) ((x_{it} - \rho_x^0 x_{i,t-1}) - (x_{i,t-1} - \rho_x^0 x_{i,t-2})) | I_{i,t-2}] = 0. \quad (18)$$

Comparing (18) to (17), it is clear that there will be a pseudo-solution to moment condition (17) in the same location as the model without a fixed effect, i.e. where $\rho_\omega = \rho_x^0$ and $\beta = \beta^0 + \frac{1}{\theta^0}$. Note that this argument does not depend on the existence of a fixed effect in the input demand equation, i.e. there is a pseudo-solution even if $\pi_i = \pi$.

6.2 Multiple Endogenous Variables

We next examine the under-identification problem where we have multiple variable inputs in the production function context or, more generally, multiple endogenous regressors. Consider a panel regression model with two endogenous regressors x and z ,

$$y_{it} = \alpha^0 + \beta^0 x_{it} + \gamma^0 z_{it} + \omega_{it} + \eta_{it}, \quad (19)$$

where

$$\begin{aligned} x_{it} &= \pi_x^0 + \theta_\omega^0 \omega_{it} + \theta_\kappa^0 \kappa_{it} + \theta_\varphi^0 \varphi_{it}, \\ z_{it} &= \pi_z^0 + \delta_\omega^0 \omega_{it} + \delta_\kappa^0 \kappa_{it} + \delta_\varphi^0 \varphi_{it}. \end{aligned}$$

Here, \wp_{it} and κ_{it} denote two persistent market factors – such as input prices – and they follow autoregressive processes $\wp_{it} = \rho_z^0 \wp_{i,t-1} + v_{it}$ and $\kappa_{it} = \rho_x^0 \kappa_{i,t-1} + u_{it}$, respectively. Note that both inputs depend on both factors – this is typically the case in practice as the optimal choice of one variable input typically depends on the prices of all variable inputs.

On the one hand, exploiting the AR(1) process of productivity $\omega_{it} = \rho_\omega^0 \omega_{i,t-1} + \xi_{it}$, and ρ_ω^0 -differencing leads to

$$y_{it} - \rho_\omega^0 y_{i,t-1} = \alpha^0(1 - \rho_\omega^0) + \beta^0(x_{it} - \rho_\omega^0 x_{i,t-1}) + \gamma^0(z_{it} - \rho_\omega^0 z_{i,t-1}) + \xi_{it} + (\eta_{it} - \rho_\omega^0 \eta_{i,t-1}),$$

which can be used for constructing the following moment condition, as a generalization of moment condition (5),

$$E[y_{it} - \rho_\omega y_{i,t-1} - \alpha(1 - \rho_\omega) - \beta(x_{it} - \rho_\omega x_{i,t-1}) - \gamma(z_{it} - \rho_\omega z_{i,t-1}) | I_{i,t-1}] = 0. \quad (20)$$

On the other hand, consider a linear combination of the two endogenous inputs

$$\delta_\kappa^0 x_{it} - \theta_\kappa^0 z_{it} = (\delta_\kappa^0 \pi_x^0 - \theta_\kappa^0 \pi_z^0) + (\delta_\kappa^0 \theta_\omega^0 - \theta_\kappa^0 \delta_\omega^0) \omega_{it} + (\delta_\kappa^0 \theta_\wp^0 - \theta_\kappa^0 \delta_\wp^0) \wp_{it},$$

where the expression on the right-hand side follows from the two input functions. This shows that we can write a linear combination of two inputs as a linear function of two sources of persistence (one for productivity and one for market factor) – similar in structure to the one input in the benchmark model. Solving for ω_{it} and plugging it into the output equation (19) leads to a similar structure of the output function, except that ω_{it} is replaced by \wp_{it} and the coefficients on inputs are different from the true production function parameters. Notice that this is similar in structure to the alternative expression of the output function like the equation we derived for the one input case (12) .

Therefore, following similar arguments as before, we can derive a spurious solution of moment function (20) by taking ρ_z^0 -differencing due to the autoregressive process of \wp_{it} , instead of the productivity. As the result, this spurious solution is at the dynamic parameter $\rho_\omega = \rho_z^0$ of the persistence market factor \wp_{it} , not the productivity. Similarly, a different linear combination $\delta_\wp^0 x_{it} - \theta_\wp^0 z_{it}$ leads to another spurious solution of moment (20). This solution is at the dynamic parameter $\rho_\omega = \rho_x^0$ of the persistence market factor κ_{it} .

In summary, we conclude that when the production function has multiple inputs, the dynamic panel moments yield the same number of spurious solutions as the number of flexible inputs that have persistent shocks.

6.3 Fixed Input

The under-identification problem still exists when we have both fixed and flexible inputs. Consider a panel regression model with a flexible input x and a fixed input z like capital as follows

$$y_{it} = \alpha^0 + \beta^0 x_{it} + \gamma^0 z_{it} + \omega_{it} + \eta_{it}, \quad (21)$$

where the flexible input function is given by

$$x_{it} = \pi^0 + \theta_\omega^0 \omega_{it} + \theta_z^0 z_{it} + \kappa_{it},$$

and persistent market factor κ_{it} follows the AR(1) process $\kappa_{it} = \rho_x^0 \kappa_{i,t-1} + u_{it}$ as before.

Solving for ω_{it} in the input function and plugging it into (21) leads to a similar structure of the alternative expression of the output function, except that ω_{it} is replaced by κ_{it} and the coefficients on inputs are different from the true production function. Following similar arguments as before, we can derive a spurious solution of the dynamic panel moment due to this similar structure where the spurious solution is at the dynamic parameter $\rho_\omega = \rho_x^0$ of persistence market factor κ_{it} .

6.4 Beyond Linearity and AR(1)

In the above three extensions to our basic model, we were able to analytically show that the pseudo-solution (or pseudo-solutions) continue to be a potential problem. In this section, we consider some additional extensions. While we are unable to analytically show the existence of pseudo-solutions in these extensions, we are able to use simulations to show that they continue to exist, at least in neighborhoods of our benchmark model. In particular, we consider relaxing the assumed linearity of the input demand function and the AR(1) process for unobserved shock κ_{it} in that input demand function.

It is important to note that we continue to assume that ω_{it} follows an AR(1) process. This means that our basic dynamic panel moment condition, i.e. $e_1(\Phi)$, continues to be zero at the true parameter value. What we are focused on is not the zero at the true parameter value, but whether there continues to be a pseudo-solution (or pseudo-solutions), and the location of such pseudo-solutions. To do this, we start with the same basic DGP of Section 5 as our benchmark case and consider perturbations of this model. There is no need to do multiple Monte Carlo replications to study this – it is easier to simply boost the number of observations (to 200,000) to simulate the “asymptotic” objective function and examine the location of zeros. In the benchmark case we know there are “asymptotic” zeros at 0.6 and 1.6. In the alternative models, we know there is an “asymptotic” zero at 0.6. The question is whether there are other

pseudo-solutions as well.

First, we consider an alternative model where the input function is a non-linear function of the productivity process, i.e.:

$$x_{it} = \pi^0 + \theta_1^0 \omega_{it} + \theta_2^0 \omega_{it}^2 + \kappa_{it},$$

where $\theta_1^0 = 1$, as in the benchmark model. Figure 4 shows the basic dynamic panel objective function (based on $e_1(\Phi)$), for both the benchmark model (without nonlinearity, i.e. $\theta_2^0 = 0$) and this alternative model where we set $\theta_2^0 = 0.5$ and 1.¹¹ For easy visualization, we display the objective function as a function of just β by concentrating out the constant term and AR(1) parameter.¹² The objective functions for the alternative model also have pseudo-solutions, at β quite close to the pseudo-solution in the benchmark model. Interestingly, to the right of the pseudo-solution, the objective function becomes flatter than the benchmark case, especially when $\theta_2^0 = 1$. However, the main finding is that the pseudo-solution does not appear to disappear when the x_{it} process is not linear in ω_{it} .

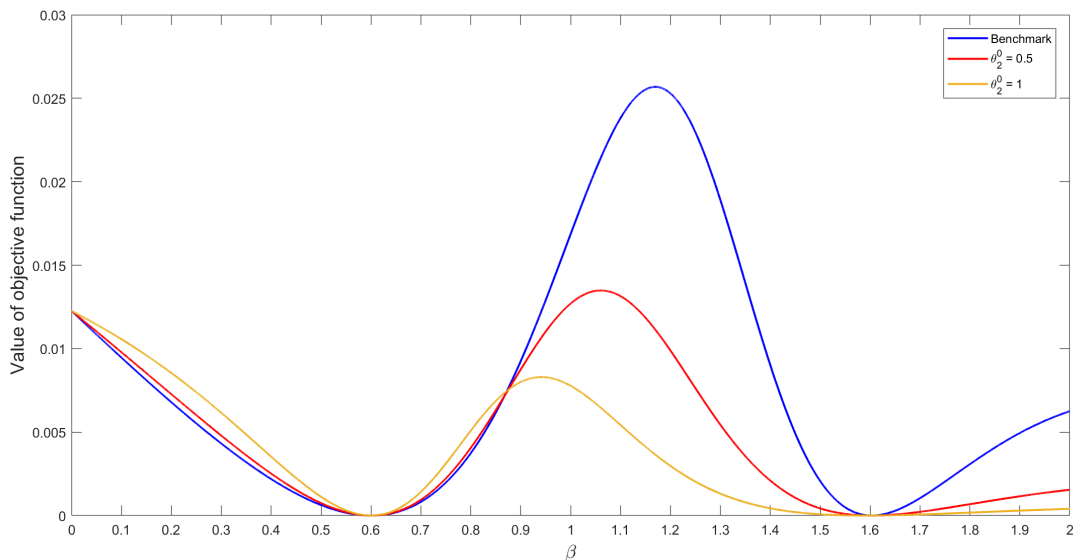


Figure 4: Nonlinearity in the Productivity Process

Second, we consider a model where there is a nonlinearity in the first-order Markov process

¹¹For ease of comparison, we rescale objective functions such that they are equal at $\beta = 0$ in Figures 4-7.

¹²To concentrate out ρ_ω and the constant, for a candidate $\beta = \tilde{\beta}$, we estimate ρ_ω and the constant with the linear equation

$$\left(y_{it} - \tilde{\beta} x_{it} \right) = \alpha(1 - \rho_\omega) + \rho_\omega \left(y_{i,t-1} - \tilde{\beta} x_{i,t-1} \right) + \xi_{it} + \eta_{it} - \rho_\omega \eta_{i,t-1}$$

using instrumental variables with $\left(y_{i,t-2} - \tilde{\beta} x_{i,t-2} \right)$ as an instrument for $\left(y_{i,t-1} - \tilde{\beta} x_{i,t-1} \right)$. So that this objective function illustrates exact zeros (rather than “close to zeros”), we only interact moment $e_1(\Phi)$ with $x_{i,t-1}$.

for κ_{it} . Specifically, we consider:

$$\kappa_{it} = \frac{e^{\theta_2^0 |\kappa_{i,t-1}|}}{1 + e^{\theta_2^0 |\kappa_{i,t-1}|}} \kappa_{i,t-1} + u_{it}.$$

When we set $\theta_2^0 = 0$, $\frac{e^{\theta_2^0 |\kappa_{i,t-1}|}}{1 + e^{\theta_2^0 |\kappa_{i,t-1}|}} = 0.5 \forall \kappa_{i,t-1}$, so this model is equivalent to the benchmark case (where the AR(1) parameter = 0.5). When $\theta_2^0 > 0$, the Markov-process is non-linear. One interpretation of this is that the extent to which $\kappa_{i,t-1}$ depreciates depends on the level of $|\kappa_{i,t-1}|$. Note that since $|\kappa_{i,t-1}|$ is positive, the heterogeneous depreciation rate in this model is between 0.5 and 1. Figure 5 shows the (concentrated) objective function for this model when $\theta_2^0 = 0.5$ and 0.75. Again, the pseudo-solution persists, but moves down relative to the benchmark model. In these examples, the pseudo-solutions are at approximately 1.46 and 1.51 for $\theta_2^0 = 0.5$ and 0.75, respectively. Interestingly, the effect of θ_2^0 on the location of the pseudo-parameter is non-monotonic. This appears to be because changing θ_2^0 affects multiple things in the model - e.g., it simultaneously changes both the average level of depreciation and its heterogeneity.

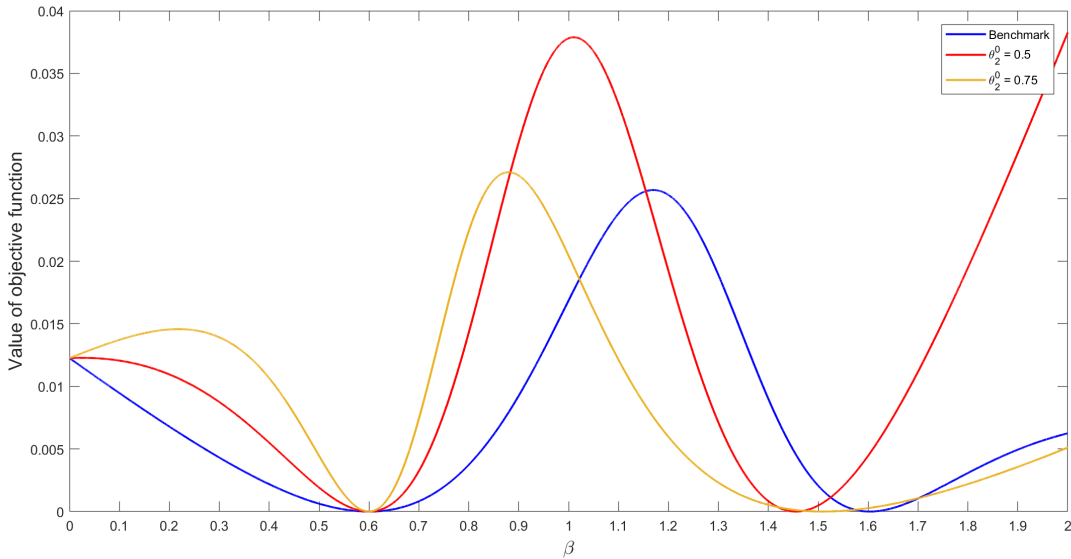


Figure 5: Nonlinearity in the Input Market Persistence Process

Third, we change the κ_{it} process in a different way - keeping linearity but allowing it to follow an AR(2) process instead of an AR(1) process, i.e.

$$\kappa_{it} = \rho_1^x \kappa_{i,t-1} + \rho_2^x \kappa_{i,t-2} + u_{it}.$$

As in the benchmark case, we set ρ_1^x to 0.5 and compare the benchmark, i.e. $\rho_2^x = 0$, to alternative models where $\rho_2^x = 0.01$ or 0.02. Figure 6 shows the (concentrated) objective

functions. Like in the other alternative models, there is still a pseudo-solution in this model. However, in this case the pseudo-solution is at a higher β , at approximately 1.67 and 1.82 for $\rho_2^x = 0.01$ and 0.02, respectively.

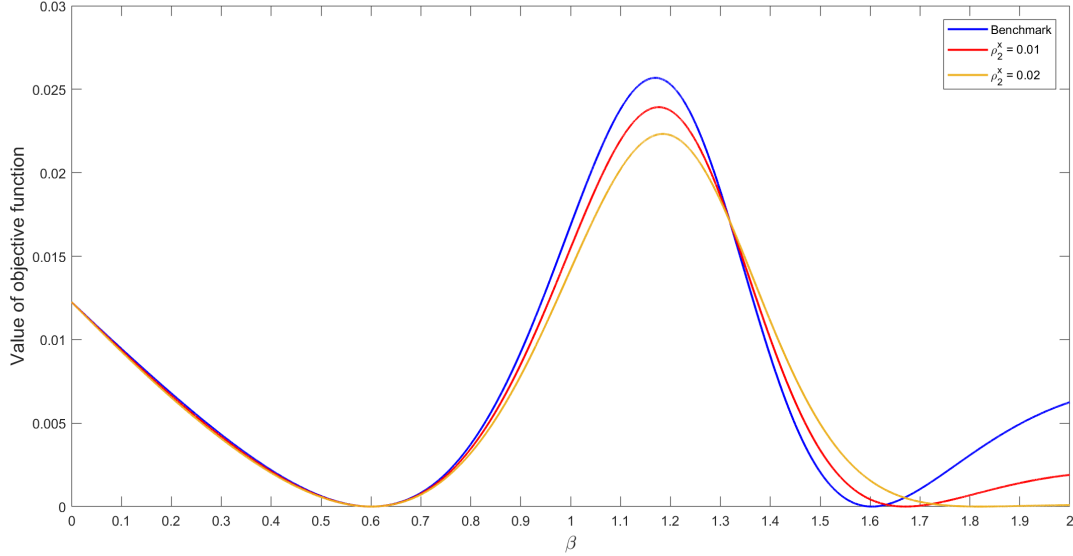


Figure 6: AR(2)

In summary, these models suggest that the existence of pseudo-solutions is not a knife-edge result that relies on the very specific input demand equation in our benchmark model. While the exact locations of these pseudo-solutions differ from the benchmark model, the general patterns are similar - in particular, the pseudo-solutions are at a larger β than the true parameter β^0 ($=0.6$). Obviously, this is only a small set of alternative models and there could very well be alternative models without pseudo-solutions, especially as one moves parameters further away from the benchmark model. In other experiments, we have found that many things can happen - for example, in the following model (where the curvature of the non-linearity in κ_{it} is reversed),

$$\kappa_{it} = \left(1 - \frac{e^{|\kappa_{i,t-1}|}}{1 + e^{|\kappa_{i,t-1}|}} \right) \kappa_{i,t-1} + u_{it}$$

we see multiple pseudo-solutions, one slightly above 1.6, and one at approximately $\beta = 1.95$ (Figure 7). However, the results hint that this is a general issue - the ρ -differencing moment (5) of the dynamic panel approach (and likely of related quasi-differencing approaches, e.g. the proxy variable approaches of OP/LP/ACF) creates a quadratic form indeterminacy that seems to exist relatively generally. As such, we believe that this is an issue that researchers should be aware of when using these techniques.

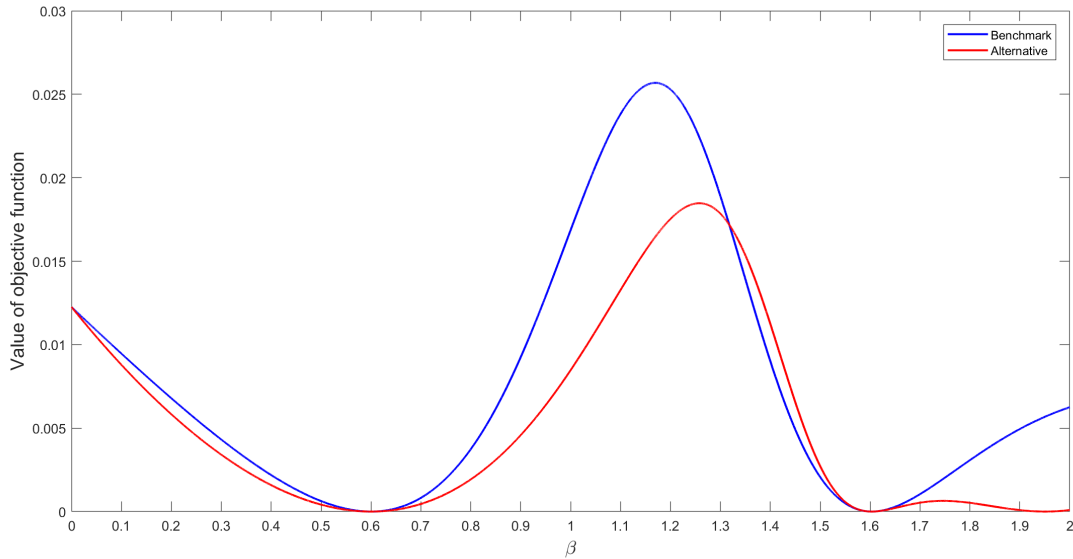


Figure 7: Alternative Model Where the Curvature of the Non-linearity in κ_{it} is Reversed

6.5 Stronger Timing/Information Set Assumptions

Interestingly, we have found one class of modeling assumptions that does appear to remove the pseudo-solution. Recall Assumption 1, the timing and information set assumption that the moment conditions of our benchmark model are predicated on. Again, this can be interpreted as following from an economic model where 1) firms do not observe ω_{it} and κ_{it} until t , and 2) firms choose x_{it} at t and hence implying (with some additional assumptions) that future innovations in ω_{it} and κ_{it} , i.e. $\xi_{i\tau}$ and $u_{i\tau}$ for $\tau > t$, are uncorrelated with x_{it} .

Some work in this literature has been willing to make such timing (and information set) assumptions stronger. For example, in OP/LP/ACF procedures, it is often assumed that firms need to commit to their capital choice at $t - 1$. In demand models considering endogenous product characteristics, it is sometimes assumed that those product design decisions are made one or more periods before the product is sold. If one assumes that x_{it} is chosen at $t - 1$ (and one continues to make the assumption that firms do not observe ω_{it} and κ_{it} until t) then information set I_{it-1} in Assumption 1 additionally includes x_{it} . This means that (5) implies an additional orthogonality condition that can be used for estimation (since ξ_{it} is now mean independent of x_{it}). Intuitively, since x_{it} is chosen at $t - 1$, innovations that are not observed until after that choice, e.g. ξ_{it} are orthogonal to it.

In the dynamic panel literature, this assumption is often referred to as x_{it} being “predetermined”. Akerberg (2020) details how these timing and information set assumptions can be strengthened or relaxed – an important point is that it is not only the timing of when x_{it} is

chosen that matters but also what the firm knows at that time. For example, if x_{it} is chosen at $t - 1$ but one assumes that firms observe ω_{it} and κ_{it} also in $t - 1$ (one period earlier), then one does not obtain additional moment restrictions.

So, consider an alternative model where firms make input decisions at time $t - 1$ as a function of their information set $I_{i,t-1}$, e.g.

$$x_{it} = \pi^0 + \theta^0 \rho_\omega^0 \omega_{i,t-1} + \kappa_{i,t-1}. \quad (22)$$

Note that $\rho_\omega^0 \omega_{i,t-1}$ is a conditional expectation of ω_{it} at time $t - 1$. In a production function context, it makes sense that a firm would base their x_{it} decision on their expectation of ω_{it} , but this assumption is not necessary as our derivation is for arbitrary θ^0 . Since x_{it} is chosen at $t - 1$, it is now a function of $\kappa_{i,t-1}$ instead of κ_{it} (because, like ω_{it} , κ_{it} has not been observed yet).¹³

Following the arguments of the benchmark case, we can consider the basic dynamic panel moment evaluated at the pseudo-parameters $(\beta, \rho_\omega, \alpha) = (\beta^0 + \frac{1}{\theta^0}, \rho_x^0, \alpha^0 - \frac{\pi^0}{\theta^0})$. The first part of the derivation from Section 2 stays the same:

$$\begin{aligned} & E[(y_{it} - \rho_\omega y_{i,t-1}) - \alpha(1 - \rho_\omega) - \beta(x_{it} - \rho_\omega x_{i,t-1}) | I_{i,t-1}] \\ &= E[(y_{it} - \rho_x^0 y_{i,t-1}) - \left(\alpha^0 - \frac{\pi^0}{\theta^0}\right) (1 - \rho_x^0) - \left(\beta^0 + \frac{1}{\theta^0}\right) (x_{it} - \rho_x^0 x_{i,t-1}) | I_{i,t-1}] \\ &= E[\alpha^0(1 - \rho_x^0) + \beta^0(x_{it} - \rho_x^0 x_{i,t-1}) + \omega_{it} - \rho_x^0 \omega_{i,t-1} + \eta_{it} - \rho_x^0 \eta_{i,t-1} \\ &\quad - \left(\alpha^0 - \frac{\pi^0}{\theta^0}\right) (1 - \rho_x^0) - \left(\beta^0 + \frac{1}{\theta^0}\right) (x_{it} - \rho_x^0 x_{i,t-1}) | I_{i,t-1}] \\ &= E\left[\frac{\pi^0}{\theta^0} (1 - \rho_x^0) - \frac{1}{\theta^0} (x_{it} - \rho_x^0 x_{i,t-1}) + \omega_{it} - \rho_x^0 \omega_{i,t-1} + \eta_{it} - \rho_x^0 \eta_{i,t-1} | I_{i,t-1}\right]. \end{aligned} \quad (23)$$

However, $(x_{it} - \rho_x^0 x_{i,t-1})$ is now different. Using (22), we have

$$(x_{it} - \rho_x^0 x_{i,t-1}) = (1 - \rho_x^0) \pi^0 + \theta^0 [\rho_\omega^0 \omega_{i,t-1} - \rho_x^0 \rho_\omega^0 \omega_{i,t-2}] + (\kappa_{i,t-1} - \rho_x^0 \kappa_{i,t-2}).$$

and note also that we can write

$$\omega_{it} - \rho_x^0 \omega_{i,t-1} = \rho_\omega^0 \omega_{i,t-1} - \rho_x^0 \rho_\omega^0 \omega_{i,t-2} + \xi_{it} - \rho_x^0 \xi_{i,t-1}.$$

¹³Working with κ_{it} instead of $\kappa_{i,t-1}$ does not change our result (i.e. firms knowing κ_{it} at time $t - 1$). What matters for our result is firm i knowing only $\omega_{i,t-1}$ at time $t - 1$, not ω_{it} . It does not matter for κ_{it} . In some sense it is just a relabeling of κ_{it} . More generally, what is important for a stronger timing/information assumption to rule out the pseudo-solution is the unobserved factor that enters both the productivity function and the input function, not for the unobserved factor(s) that only enters the input function.

Substituting these in to (23), we get

$$\begin{aligned}
& E\left[\frac{\pi^0}{\theta^0}(1 - \rho_x^0) - \frac{1}{\theta^0}(x_{it} - \rho_x^0 x_{i,t-1}) + \omega_{it} - \rho_x^0 \omega_{i,t-1} + \eta_{it} - \rho_x^0 \eta_{i,t-1} | I_{i,t-1}\right] \\
= & E\left[\frac{\pi^0}{\theta^0}(1 - \rho_x^0) - \frac{1}{\theta^0} \left[(1 - \rho_x^0) \pi^0 + \theta^0 [\rho_\omega^0 \omega_{i,t-1} - \rho_x^0 \rho_\omega^0 \omega_{i,t-2}] + (\kappa_{i,t-1} - \rho_x^0 \kappa_{i,t-2}) \right] \right. \\
& \left. + \rho_\omega^0 \omega_{i,t-1} - \rho_x^0 \rho_\omega^0 \omega_{i,t-2} + \xi_{it} - \rho_x^0 \xi_{i,t-1} + \eta_{it} - \rho_x^0 \eta_{i,t-1} | I_{i,t-1}\right] \\
= & E\left[-\frac{1}{\theta^0} u_{i,t-1} + \xi_{it} - \rho_x^0 \xi_{i,t-1} + \eta_{it} - \rho_x^0 \eta_{i,t-1} | I_{i,t-1}\right]
\end{aligned}$$

Importantly, because x_{it} is chosen at $t - 1$, it is in $I_{i,t-1}$. And because x_{it} is chosen as a function of $\omega_{i,t-1}$ and $\kappa_{i,t-1}$, it is clearly correlated with $u_{i,t-1}$ and $\xi_{i,t-1}$. Hence, as long as x_{it} is utilized in the formation of moment conditions, the pseudo-solution can be ruled out.¹⁴ Of course, at the true parameters, the dynamic panel moment condition is

$$\begin{aligned}
& E[(y_{it} - \rho_\omega^0 y_{i,t-1}) - \alpha^0(1 - \rho_\omega^0) - \beta^0(x_{it} - \rho_\omega^0 x_{i,t-1}) | I_{i,t-1}] \\
= & E[(y_{it} - \alpha^0 - \beta^0 x_{it}) - \rho_\omega^0(y_{i,t-1} - \alpha^0 - \beta^0 x_{i,t-1}) | I_{i,t-1}] \\
= & E[\omega_{it} - \rho_\omega^0 \omega_{i,t-1} + \eta_{it} - \rho_\omega^0 \eta_{i,t-1} | I_{i,t-1}] = E[\xi_{it} + \eta_{it} - \rho_\omega^0 \eta_{i,t-1} | I_{i,t-1}] = 0,
\end{aligned}$$

and since ξ_{it} is not in $I_{i,t-1}$ and hence not correlated with x_{it} , this continues to hold.

In conclusion, if one is willing to make stronger timing assumptions, it appears that the pseudo-solution we consider is ruled out. Presumably this is also the case if one is willing to make even stronger timing assumptions, e.g. where x_{it} is chosen at some $t - \Delta$ where $\Delta > 1$. Of course, and as evidenced by the empirical literature, these are often assumptions that researchers are unwilling to make, such as in the case of variable inputs in the production function context.

7 More General Lessons

Sections 3 and 4 suggested specific moment equations and procedures to avoid the pseudo-solution in our simple model. However, the additional moments we derive rely on the specific DGP for x_{it} given by (2) and (4). Given that the previous section suggests that, with more general x_{it} DGPs, these dynamic panel and proxy variable approaches have similar identification issues, a natural question is what to do in those cases. In particular, suppose a researcher is *not* willing to impose (2) and (4) and instead wants to allow for (or model) a more general x_{it} process. Obviously, it is difficult for us to give exact advice given the many possible models that researcher may construct. However, our illustration of the simple case provides

¹⁴Of course, this does not preclude other possible pseudo-solutions, but we have not found any in simulations.

some constructive advice on how researchers might reduce the likelihood of ending up at a pseudo-solution in more complex models.

One possibility would be to specify the DGP for x_{it} up to parameters and follow the logic of Section 3 and/or 4 to derive analogous estimating equations. This could involve use of two moments (analogous to e_t^1 and e_t^2) or three moments (adding e_t^3), and given the intuition above one would likely want to impose assumptions on the direction of the relationship between x_{it} and ω_{it} , analogous to the sign restriction on θ we use above. In numeric optimization, one might also be on the lookout for the possibility of ending up at pseudo-solutions similar to $\Phi^{\infty 1}$ and $\Phi^{\infty 2}$.

However, suppose one is unwilling to commit to assuming a specific model for x_{it} . This is perhaps more natural given this literature - one nice aspect of both the dynamic panel and proxy variable methods is that they do *not* rely on fully specifying a parametric DGP for x_{it} (though often they do make some assumptions on these DGPs, e.g. the scalar unobservable assumption used in the proxy variable literature). We also believe our findings lead to constructive practical advice in this situation. In particular, in our simple model (and in “nearby” more complex models), the pseudo-solution for β is greater than the true β^0 (this is based on $\theta^0 > 0$, the reverse is true when $\theta^0 < 0$). This means that we know the direction in which we might err. In particular, if one has found a candidate estimate $\hat{\beta}$ (and believes $\theta^0 > 0$), one might want to be extra diligent in continuing to search for minima at parameters where $\beta < \hat{\beta}$. This is because one is erroneously at the spurious zero, the non-pseudo-solution should be below $\hat{\beta}$. Of course, while one should always be diligent in numeric search to avoid local minima, the results here suggest one should be extra concerned about local minima in one specific direction relative to a candidate estimate $\hat{\beta}$.

A related possibility is to estimate the model based on only e_t^1 , i.e. without specifying a DGP for x_{it} , but to then do some ex-post testing of that solution. In particular, given a candidate estimate $(\hat{\alpha}, \hat{\beta}, \hat{\rho}_\omega)$ based on e_t^1 , one can construct the residuals

$$\widehat{\omega_{it} + \eta_{it}} = y_{it} - \hat{\alpha} - \hat{\beta}x_{it}.$$

Again, assuming one is willing to sign θ , one could examine the relationship between x_{it} and this residual, e.g. through a regression or correlation analysis. Although this regression cannot be interpreted causally because of the existence of η_{it} in the constructed residual, the measurement error tends to attenuate the coefficient.¹⁵ So, at least in models “close to” our simple model, one would expect that the sign of this relationship at the pseudo-solution would be the reverse of what one would expect. In other words, if one believes that $\theta > 0$, an indication of a possible

¹⁵Note that this would be easier in the proxy variable methods, as a by-product of the assumptions guaranteeing invertibility in those models allows one to separately identify $\widehat{\omega_{it}}$ and $\widehat{\eta_{it}}$.

pseudo-solution would be finding a negative correlation between $\widehat{\omega_{it} + \eta_{it}}$ and x_{it} .¹⁶

Lastly, one might consider starting by estimating a model that is more robust to the pseudo-solution problem but is not necessarily an appropriate model. Initial estimates from this misspecified model can be used as starting values for estimating the correct model to reduce the likelihood of ending up at a pseudo-solution in the correct model. Our above analyses suggest at least two possibilities for this. First, even if one does not believe the stronger timing assumption of Section 6.5, one might estimate that model anyway. Then one could use the estimates as starting values for a model with the (correct) weaker timing assumption. The hope is that the stronger timing assumption does not overly bias the parameters, and thus keeps the numeric search procedure from straying to a pseudo-solution. Second, one could assume input choice (2) and use our augmented procedure, even if one does not believe that the input choice equation is exactly specified. One could again use the resulting estimates as starting parameters for an estimation procedure that does not impose a specific input choice equation. Again, the hope is that while those starting parameters may be inconsistent, they may be close enough to the truth to be helpful at getting the numeric search procedure to the true solution instead of the pseudo-solution.

Of course, these informal suggestions are not foolproof in avoiding pseudo-solutions, and in some datasets with even more complex x_{it} processes, there may not even be pseudo-solutions. However, we feel that being aware of the possibility of such pseudo-solutions and this guidance for avoiding them can be useful for applied researchers.

8 Conclusion

We have studied a potential identification problem in methods that use timing and information set assumptions to resolve endogeneity problems in structural models. These methods have been applied widely, both in the production function context and elsewhere. Prior work calls attention to this identification problem in a fairly narrow context. We show that the identification problem appears to be endemic across a broad set of models using these types of assumptions, and demonstrate how to resolve it using a sign restriction and additional modelling. Our Monte Carlo simulations illustrate that our procedure works in a simple example, and we suggest ways of extracting intuition from that solution to more complex empirical settings.

¹⁶Note that this is related to the practice of comparing these coefficient estimates to those of standard OLS, i.e. assessing signing the bias of OLS. This is because the sign of the OLS bias is related to θ . So the possibility of pseudo-solutions provides another rationale for doing this, i.e. if one is at the pseudo-solution, the sign of the OLS bias may be the opposite of what one expects it to be.

References

- [1] Akerberg, D. (2020), "Timing Assumptions and Efficiency: Empirical Evidence in the Production Function Context", mimeo, UT Austin
- [2] Akerberg, D., K. Caves, and G. Frazer (2015), "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83, 2411-2451.
- [3] Alonso-Borrego, C. and M. Arellano (1999) "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data," *Journal of Business & Economic Statistics* 17, 36-49
- [4] Arellano, M. and S.R. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277-297.
- [5] Bajari, P., J. Fruehwirth, K. Kim, and C. Timmins (2012), "A Rational Expectations Approach to Hedonic Price Regressions with Time-Varying Unobserved Product Attributes: The Price of Pollution," *American Economic Review*, 102(5), 1898-1926.
- [6] Bishop, K. and A. Murphy (2019), "Valuing Time-varying Attributes Using the Hedonic Model: When Is a Dynamic Approach Necessary?," *Review of Economics and Statistics*, 101, 134-145.
- [7] Blundell, R., and S. Bond (2000), "GMM Estimation With Persistent Panel Data: An Application to Production Functions," *Econometric Reviews*, 19, 321-340.
- [8] Gandhi, A, S. Navarro, and D. Rivers (2020), "On the Identification of Gross Output Production Functions," *Journal of Political Economy*, 128, 2973-3016.
- [9] Grieco, P, S. Li, and H. Zhang (2016), "Production Function Estimation with Unobserved Input Price Dispersion," *International Economic Review*, 57(2), 665-690.
- [10] Grennan, M. (2013), "Price Discrimination and Bargaining: Empirical Evidence from Medical Devices," *American Economic Review*, 103(1), 145-177.
- [11] Kim, K., Y. Luo, and Y. Su (2019), "A Robust Approach to Estimating Production Functions: Replication of the ACF procedure," *Journal of Applied Econometrics*, 34(4), 612-619.
- [12] Lee, R.S. (2013), "Vertical Integration and Exclusivity in Platform and Two-Sided Markets," *American Economic Review*, 103(7), 2960-3000.

- [13] Manski, C.F. and J.V. Pepper (2000), "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997-1010.
- [14] Levinsohn, J. and A. Petrin (2003), "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317-342.
- [15] Olley, S. and A. Pakes (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263-1298.
- [16] Sullivan (2017), "The Ice Cream Split: Empirically Distinguishing Price and Product Space Collusion", mimeo, U Wisconsin
- [17] Sweeting, A. (2009), "The Strategic Timing Incentives of Commercial Radio Stations: An Empirical Analysis Using Multiple Equilibria," *Rand Journal of Economics*, 40(4), 710-742.