Estimating Productivity and Markups Under Imperfect Competition

James Brand*

November 1, 2020

Abstract

This paper revisits the standard production function model and proposes an alternative identification and estimation procedure. Specifically, I argue that some of the assumptions of the standard production function model are inconsistent with the increasingly popular use of production function methods in the estimation of markups. I then show that the seminal nonclassical measurement error result in Hu and Schennach (2008) can be used to nonparametrically identify the production function under alternative assumptions which do not require specifying the demand firms face or any knowledge of firms' input demand functions. I apply the intuition of this result to develop a GMM estimation procedure for the most practically relevant production function models, and explore the performance of the resulting estimates relative to workhorse methods in simulations.

*University of Texas at Austin. email:: jamesbrand@utexas.edu. I am grateful for the advice and support of Dan Ackerberg and Jorge Balat in writing this paper.

1 Introduction

In settings of imperfect competition and differentiated goods, firms' input decisions are complicated functions of the costs and residual demand curves they face in equilibrium. Without assuming the structure of demand or competition, the arguments and functional forms of these input decisions (i.e. input demand functions) are entirely unknown. Since the seminal work of Olley and Pakes (1996), the literature has focused on simpler cases in which input demand functions satisfy two convenient criteria: (i) conditional on fixed inputs (e.g. capital), demand for some variable input is monotonic in a firm's unobservable productivity, and (ii) no other unobservables enter the demand function.[1] I call the combination of these two criteria the "scalar unobservable assumption." In the workhorse production function methods, these assumptions are crucial in identification and estimation, as they ensure that an unobservable productivity term can be written as a function of observables (Levinsohn and Petrin, 2003; Ackerberg, Caves and Frazer, 2015; Gandhi, Navarro and Rivers, 2016).

Relaxing the scalar unobservable assumption is important for two predominant reasons. First, production data is often lacking on many fronts. The most commonly used datasets contain no information on the costs the firm faces (e.g. input prices) nor their competitive environment. The literature has discussed the many issues brought about by these types of missing data. Second, following De Loecker and Warzynski (2012) (DLW), some authors have begun to use estimates from the production function to estimate and discuss nationwide trends in markups and competition. As these objects are so crucially and endogenously linked to unobservable determinants of costs and demand, estimates of the production function should be calculated under the weakest possible assumptions in order to reduce the chance that production function and/or markup estimates inherit the demand structure implied by the scalar unobservable assumption.

Recently an interest has grown in methods which address some of these concerns. Doraszelski and Jaumandreu (2017) and Balat, Brambilla and Sasaki (2016) both maintain a monotonicity assumption analogous to (i) but relax (ii) by permitting firms to have multiple dimensions of productivity (e.g. labor- and capital-augmenting). Relatedly, Li and Sasaki (2017) and Ackerberg and Hahn (2015) both develop models which identify heterogeneous labor and capital productivity indexed by firms' Hicks-neutral productivity. In addition, empirical work in these directions has indicated that labor productivity is quite heterogeneous across firms, meaning that simplistic (e.g. Cobb-Douglas) production functions are misspecified. Jaumandreu (2018), De Loecker et al. (2016), and Blum et al. (2018) have instead focused on permitting firm output to be heterogenous in at least one unobservable dimension, and on allowing these unobservables to affect input use. Each of these are crucial additions to the literature which model what may be the most likely deviations from the scalar unobservable

[1] These assumptions cover simple and intuitive cases, such as (i) perfect competition with homogeneous goods and (ii) monopolistic competition in which all firms face the same residual demand.

assumption.[2]

In this paper I take an alternative approach. Rather than modeling the way in which a firm violates the scalar unobservable assumption, I show that the production function and the distribution of productivity can be identified without any knowledge of the arguments or functional form of firms' input demand functions, as long as three periods of data are available and an independence assumption is satisfied. I present a proof that the Hu and Schennach (2008) (hereafter HS) nonparametric identification result, which they apply to a nonclassical measurement error problem, can be applied to the standard production function model. The key intuition for my approach is that in the production function, conditional on observables (i.e. inputs), a firm's output is an error-ridden signal of its productivity. As this is true in each time period, lagged output can be used as an instrument to distinguish between current productivity and measurement error, as long as the former is persistent and the latter is not. This framing permits the application of the HS identification result directly. I then present a GMM estimator that makes use of this result. Relative to existing methods, the main cost of this approach is that it requires that each firm is observed for three periods, and that any measurement errors in output are fully independent of all inputs. I argue that these costs are well worth the benefits of dropping the scalar unobservable assumption.

Many other authors have applied measurement error methods to identify panel models in general and to the production function specifically. Hu and Shum (2012), Shiu and Hu (2013), and Sasaki (2015) all apply versions of the HS result to prove nonparametric identification of general dynamic panel models under assumptions different (and sometimes more general) than those used here. This paper is perhaps most related to Cunha, Heckman and Schennach (2010), Freyberger (2017), and Arellano and Bonhomme (2016), each of which also apply the HS result to general panel models. Freyberger (2017) considers a factor model with potentially many fixed unobserved factors. Cunha, Heckman and Schennach (2010) consider cases in which there are multiple measurements for each observation in each period, and use these additional measurements to identify unobserved factors in the human capital accumulation processes of children. Arellano and Bonhomme (2016) prove the identification of a non-separable panel model with fixed effects and idiosyncratic errors, and estimate that model via quantile regressions. Though I focus on a simpler subset of models (with additively separable errors), my work here in some ways modifies the approach in Arellano and Bonhomme (2016) to permit a persistent time-varying unobservable.

Many papers in the literature have also applied measurement error methods to the estimation of the production function. Kato and Sasaki (2018) develop uniform confidence bands for the distribution of productivity under the assumption that pro-

[2]A recent paper by Demirer (2019) takes another approach, wherein he permits an additional (labor-augmenting) productivity shock in the production function under the assumption that labor is static. That paper is complimentary to this one, as I cannot (without additional structure) permit input-augmenting productivity shocks but also require weaker assumptions on input decisions to identify the production function.

ductivity levels are independent of innovations to productivity. il Kim, Petrin and Song (2016) permit capital to be mismeasured, and use repeated measurments of capital to identify the production function in this setting.[3] Of particular relevance are Hu, Huang and Sasaki (2017), who estimate a model in which firms' input usage can be mismeasured and/or affected by idiosyncratic errors (e.g. exogenous input price variation). This is one of the only existing papers to permit idiosyncratic unobservables to enter the input demand function, and although their identification proof follows the HS result closely, they propose a GMM estimator for their model. I am, to the best of my knowledge, the first to apply the HS result to identify the production function without assuming any knowledge of the determinants of firms' input use.

Beyond the identification result, the contribution of this paper is a simple GMM estimator for the proposed method which makes use of my identification argument and which is in some ways an extension of the Blundell and Bond (1998) estimation approach to cases in which the unobservable evolves nonlinearly. Although the identification proof motivates an intuitive semi-parametric maximum likelihood estimation approach, the (density) functions which must be estimated in this approach have many arguments, so treating them flexibly presents a computational burden. To ameliorate this issue, I instead propose a GMM estimator in the spirit of Hu, Huang and Sasaki (2017) which covers most commonly estimated production function models. I show in Monte Carlo simulations that, unlike traditional estimators of the production function following the control function approach, the proposed estimator is unaffected by the addition of unobservables into the input demand function.

Interest in the estimation of the production function has grown in part due to the work by De Loecker and Warzynski (2012) , who apply the argument from Hall (1988) to estimate firm-specific markups over marginal cost and have become heavily cited in the literature. The Hall, De Loecker, and Warzynski (HDLW) markup estimation method requires that the researcher estimate the production function for each firm and the distribution of an idiosyncratic, unobservable, component of productivity. These estimates are then combined and weighted appropriately to recover markups. This method has been used to study the impact of trade liberalization on markups (De Loecker et al., 2016), monopsony power in input markets (Morlacco, 2018), and cost efficiencies from mergers (Grieco, Pinkse and Slade, 2018). Most recently De Loecker, Eeckhout and Unger (2018) and De Loecker and Eeckhout (2018) have applied the HDLW method to the United States and more and 100 countries worldwide (respectively) to show evidence of a steady and dramatic increase in average markups.

These papers have sparked a large and growing methodological debate, a significant subset of which has focused on the appropriate methods for estimating the production function specifically for use in studies of markups. In three important recent papers,

[3]Though they do not apply standard measurement error identification results, Collard-Wexler and De Loecker (2016) also address measurement error in capital via a linear IV approach.

Jaumandreu (2018), Blum et al. (2018), and Flynn, Gandhi and Traina (2019) argue that the control function approach as usually implemented cannot identify markups separately of other model parameters. Jaumandreu (2018) solves this issue by explicitly modeling the firm's markup-setting problem, and Flynn, Gandhi and Traina (2019) argue that knowledge of firms' returns to scale can alleviate this identification issue. Each approach is appealing in some cases. With the right data at hand, adding some structure to demand can control for unobservable demand characteristics that might otherwise violate the scalar unobservable assumption. For these cases, Jaumandreu (2018) and Blum et al. (2018) provide two very reasonable approaches.[4] Similarly, many existing estimates of Cobb-Douglas production functions imply constant returns to scale, so in contexts where the researcher has a strong prior on the returns to scale, the approach taken by Flynn, Gandhi and Traina (2019) may be sufficient. However, in many cases we are interested in estimating production functions over many industries (meaning we may not have a good prior on returns to scale in each) with little to no data describing the demand for those products. Further, because the industries I study have many firms, controlling for demand heterogeneity flexibly would require controlling for many prices in estimation. For these contexts dropping the scalar unobservable assumption in the way I propose may have some advantages.

2 Standard Approaches

2.1 Workhorse Production Model

In this section I consider a market of single-product firms and model the production function for firm j in period t of the following form

$$Y_{jt} = F(X_{jt}; \boldsymbol{\beta}) e^{\omega_{jt}} e^{\eta_{jt}}, \tag{1}$$

where X_{jt} is a vector denoting the capital, labor, and other inputs used to produce output Y_{jt}.[5] The variable ω_{jt} represents a Hicks-neutral shock to productivity which is known by the firm in period t but is not observed by the econometrician, and η_{jt} is an i.i.d error term representing either an ex-post (i.e. after input choices have been made) shock to production or measurement error. In the following discussion I will refer to η as "measurement error" or as an "ex-post shock" interchangeably. The rest of the paper does not rely on the interpretation of η as measurement error. Following the literature, the focus of this paper will be the log of equation (1):

$$y_{jt} = f(x_{jt}; \boldsymbol{\beta}) + \omega_{jt} + \eta_{jt} \tag{2}$$

[4] Though both sets of authors treat demand very flexibly, both require specifying the determinants of demand. This is the main drawback of approaches like this, as flexibly specifying demand can require many arguments, and therefore may require estimating many parameters (Berry and Haile, 2014; Compiani, 2018).

[5] Under the appropriate assumptions, Y_{jt} can denote firm revenue or units of output.

I also assume that ω_{jt} evolves according to a first-order Markov process:

$$\omega_{jt} = g(\omega_{jt-1}; \boldsymbol{\rho}) + \xi_{jt} \tag{3}$$

The econometric challenge in estimating equation 2 is that ω_{jt} and η_{jt} both enter linearly and vary across firms and time, and that each firm's input choices will be correlated with (or functionally dependent on) the productivity shock ω_{jt}. Thus, standard regression methods like OLS with fixed effects aimed at recovering unbiased estimates of the production function parameters $\boldsymbol{\beta}$ will fail here. Since the seminal work by Olley and Pakes (1996), researchers have instead relied on inverting one of the firm's first order conditions in order to distinguish between the two unobservables. Under standard assumptions, a short-run profit maximizing firm will choose their flexible input(s) (sometimes called a "proxy" variable) as a monotonically increasing function of the productivity shock ω_{jt}.[6] That is, we can often write

$$v_{jt} = g(x_{jt}, \omega_{jt}, \boldsymbol{c}_{jt}), \tag{4}$$

where v_{jt} is some flexible input used by the firm and $\boldsymbol{c}_{jt}$ is a vector of other observables affecting the firm's input decisions. Because this function is monotonic in ω_{jt}, we can invert it and substitute the inverse into equation (2), yielding

$$y_{jt} = f(x_{jt}; \boldsymbol{\beta}) + g^{-1}(x_{jt}, \boldsymbol{c}_{jt}) + \eta_{jt} \tag{5}$$

$$\equiv \phi(x_{jt}, \boldsymbol{c}_{jt}) + \eta_{jt} \tag{6}$$

Finally, because η_{jt} is i.i.d, it is by construction independent of $x_{jt}$and $\boldsymbol{c}_{jt}$, we can identify η_{jt} as the residual of a nonparametric regression of y_{jt} on $(x_{jt}, \boldsymbol{c}_{jt})$. With η_{jt} then known for all firms, we can calculate $y_{jt} - \eta_{jt} \equiv f(x_{jt}; \boldsymbol{\beta}) + \omega_{jt}$. Under the assumption that ω_{jt} evolves according to a first-order Markov process and the availability of instruments for x_{jt}, the residual productivity shock ξ_{jt} can be inverted and $\boldsymbol{\beta}$ can be estimated by GMM in a second stage.

There are now a series of papers concerning the appropriate variables to include in the vector $\boldsymbol{c}$ in the first step. Klette and Griliches (1996) and De Loecker (2011) note that, when firms' output prices are unobserved (and thus analysis is conducted using industry-deflated revenues), we need to add market- or firm-level dummies to $\boldsymbol{c}$ to control for demand shocks. DLW and De Loecker et al. (2016) demonstrate that the set of variables to be included may be much larger, even when prices are observed. For instance, DLW include a binary variable for firm export status, which allows exporting firms to require more inputs at every level of ω_{jt}. De Loecker et al. (2016) go further and include a vector of market shares and product dummies, motivated as controls for input price variation which is unobserved in their data.

[6] Derivation of this can be found in Levinsohn and Petrin (2003) and Ackerberg, Caves and Frazer (2015).

As argued by Jaumandreu (2018), the set of variables which should be included in $\boldsymbol{c}$ often includes variables which are unavailable to the econometrician. Because researchers studying production functions rarely observe product characteristics, all differences in demand faced by firms is often unobserved. Any demand asymmetries may induce firms to choose different amounts of flexible inputs even conditional on $(x_{jt}, \boldsymbol{c}_{jt}, \omega_{jt})$, thus violating the scalar unobservable assumption. Simple examples of this include brand prestige effects, which imply that even two firms producing the same good will face different residual demand curves, and most product characteristics, which are generally not observed in production data. This makes clear that the scalar unobservable assumption implicitly makes restrictions on the forms of demand and competition which can be supported by this model of the production function. Further, because the appropriate arguments of $\boldsymbol{c}_{jt}$ depend on the nature of demand, it is difficult to (1) know which $\boldsymbol{c}_{jt}$ should be included and (2) sign the bias induced (on the estimates of the production function) by including too many or too few variables in $\boldsymbol{c}_{jt}$.

2.2 Cost Minimization to Recover Markups

There has been a lot of recent interest in the HDLW method of recovering firm-level markups from production function estimates. In this approach, the key assumption is that firms are static cost minimizers with respect to at least one perfectly variable input. Under this condition, firms will optimally set their markups as a function of the elasticitiy of output with respect to this variable input. To see this, let X and V denote the fixed a flexible inputs used in production. As shown by DLW, note that a cost-minimizing firm choosing V to produce at least $\bar{Y}_{jt}$ units of output minimizes

$$\mathcal{L} = P^x_{jt} X_{jt} + P^v_{jt} V_{jt} + \lambda_{jt}(\bar{Y}_{jt} - Y_{jt}(\cdot)) \tag{7}$$

with P^x and P^v denote the prices of their respective inputs. This problem clearly has the associated first-order condition for V

$$\frac{\partial \mathcal{L}_{jt}}{\partial V_{jt}} = P^v_{jt} - \lambda_{jt} \frac{\partial Y_{jt}(\cdot)}{\partial V_{jt}} = 0 \tag{8}$$

Now let P_{jt} denote the price at which output Y_{jt} is sold. Multiplying both sides of equation 8 by $\frac{V_{jt} P_{jt}}{Y_{jt} P_{jt}}$ and rearranging terms gives the following equality

$$\frac{\partial Y_{jt}(\cdot)}{\partial V_{jt}} \frac{V_{jt}}{Y_{jt}} = \frac{P_{jt}}{\lambda_{jt}} \frac{P^v_{jt} V_{jt}}{P_{jt} Y_{jt}} \tag{9}$$

The left side of this equation is the elasticity of output with respect to V_{jt}. The right side is composed of two terms. First, note that λ_{jt}, the shadow cost of an additional unit of output, is the marginal cost of output at $\bar{Y}_{jt}$. Thus, we can define $\mu_{jt} = \frac{P_{jt}}{\lambda_{jt}}$ as the markup over marginal cost. The second term, $\frac{P^v_{jt} V_{jt}}{P_{jt} Y_{jt}}$, is the cost of materials over

firm revenue. Letting s_{jt}^v denote this revenue share of input spending, we can rewrite equation 9 as

$$\theta_{jt}^v = \mu_{jt} s_{jt}^v \tag{10}$$

where θ_{jt}^v is the elasticity of output with respect to V, which can be derived from estimates of the production function. Therefore, once the researcher has estimated the production function, this necessary condition implies that the entire distribution of markups is known.

Note, however, that equation 10 includes measured output Y_{jt} in s_{jt}^v. As has been noted in papers applying these methods, the first-order condition is, by assumption, satisfied by the output firms *expect* to produce, not by the output observed in the data, as the latter is contaminated with measurement error or production shocks by assumption. Thus, in practice, estimated markups $\hat{\mu}_{jt}$ are constructed as

$$\hat{\mu}_{jt} = \frac{\theta_{jt}^v}{s_{jt}^v / exp(\hat{\eta}_{jt})} \tag{11}$$

where $\hat{\eta}_{jt}$ are estimates of η_{jt}. There are two important things to say here. First, the fact that $\hat{\eta}_{jt}$ enters as an exponential term is not innocuous. By Jensen's inequality, $\mathbb{E}[exp(\hat{\eta}_{jt})] > 1$. Therefore, although $\hat{\eta}_{jt}$ is mean zero by construction, $exp(\hat{\eta}_{jt})$ will not be mean 1. Additionally, $\mathbb{E}[exp(\hat{\eta}_{jt})]$ is increasing in the variance of $\hat{\eta}_{jt}$. Thus, the distribution of the estimates $\hat{\eta}_{jt}$ will determine not only the variance but also the *mean* of markups, even conditional on output elasticity estimates $\hat{\theta}_{jt}^v$. Second, some papers in the literature ignore this correction term, presumably because estimates of η in financial data tend to be very small. However, evidence in De Loecker, Eeckhout and Unger (2018) suggests that most of the growth in markups has occurred among the largest 1-5% of firms. If $\hat{\eta}_{jt}$ is largest for these firms, then this correction may reduce average markups substantially. This reveals the second reason that the choice of $\boldsymbol{c}_{jt}$ in the control function is problematic. If the researcher includes too many variables in $\boldsymbol{c}$, she may over fit the data in small samples, thereby artificially reducing the variance of $\hat{\eta}_{jt}$. On the other hand, including too few variables in $\boldsymbol{c}$ may not fully control for ω, meaning that both η and θ^v will be misestimated.

In total, this section has shown that the control function approach implicitly makes assumptions about the nature of the demand and competition firms are facing. This occurs mostly through the choice of variables to include in the control function, which are often ad-hoc and rarely relate to an underlying structural model of demand. Thus, although the scalar unobservable assumption is a convenient assumption when firms face identical demand curves and the determinants of input usage are known, it is ill-suited to many other practical cases. When estimating markups, this can cause two major issues. First, it may bias production function estimates because whatever assumptions are imposed on input choices or output demand may be misspecified.

Second, the distribution of measurement error may be estimated with bias as well due to the fact that likely determinants of input demand (e.g. input prices) are rarely observed in production data.

3 Identification of a General Panel Model

3.1 Model

The model I consider is of a panel of observations of (Y, X), in which X can be endogenous and serially correlated over time. Because this paper is motivated by the study of the production function, I call the cross-sectional unit j a "firm" which uses observable inputs X_{jt} to produce output Y_{jt} each year t. Following convention, I use lowercase letters to denote the log of uppercase variables. Each firm faces a persistent productivity shock which may be correlated with X, as well as an idiosyncratic shock which is realized after all input choices are made and independent of all other model variables. I denote these persistent and idiosyncratic shocks by ω and η, respectively. I assume that the log production function (indexed by year t but omitting the firm index j) is of the form

$$y_t = f_t(x_t) + \omega_t + \eta_t \tag{12}$$

$$f_t(x)\Big|_{x=0} = 0 \tag{13}$$

and that at least three periods of data on all firms are available. The form assumed in equation 12 allows the production function to change arbitrarily over time but not across firms within year. This is common within the literature, as most studies model $f(\cdot)$ as a linear (Cobb-Douglas) or interacted quadratic (translog) function. Equation 13 specifies the location of f and can be exchanged for any assumption specifying the production function at a point. Clearly the mean of ω_t and η_t are not separately identifed in equation 1, so I make the following normalization

Assumption 1. *(Normalization)* $\mathbb{E}[\eta_t | X_t, \omega_t] = 0$

As is standard in the literature, I also assume that ω_t evolves according to a first-order Markov process

$$\omega_t = g(\omega_{t-1}) + \xi_t$$

where the "innovation" to productivity in period t (ξ_t) is assumed to be mean-independent of ω_{t-1} and all values of ω and x before t.

3.2 Identification Intuition

The identification result in this section can be summarized by the following intuition. Start with a simpler model in which $f_t(x) = 0$ for all x, implying that we

can ignore the production function entirely. Then, suppose the researcher had three periods of data available for all (i.e. a very large number of) firms:

$$y_{jt+1} = \omega_{jt+1} + \eta_{jt+1} \equiv g(\omega_{jt}) + \xi_{jt+1} + \eta_{jt+1} \tag{14}$$

$$y_{jt} = \omega_{jt} + \eta_{jt} \tag{15}$$

$$y_{jt-1} = \omega_{jt-1} + \eta_{jt-1} \tag{16}$$

There are three components which contribute to the variation in y_{jt+1} (ω_{jt}, ξ_{jt+1}, and η_{jt+1}), all of which are unobservable. We are interested in recovering the distribution of ω and the Markov process $g(\cdot)$. When $g(\cdot)$ is linear, it can be identified by the regression of y_{jt+1} on y_{jt}. However, when $g(\cdot)$ is nonlinear, clearly this regression is misspecified. The problem, generally speaking, is that some of the variation in y_{jt} is due to η, which serves as a sort of measurement error here. Therefore the covariance between y_{t+1} and y_t clearly differs from that between ω_{t+1} and ω_t. This introduces an opportunity to use our third period of data y_{jt-1}. Because we have assumed that ω_{jt} follows a first-order Markov process, y_{jt-1} is excluded from the regression of y_{jt+1} on y_{jt} conditional on ω_{jt}. Therefore, loosely speaking, we can use y_{jt-1} as an instrument for y_{jt} in this regression to identify variation in ω_{jt}. This then identifies $g(\cdot)$ and the distribution of ξ_{jt+1}.

When we add the production function back into the equation, this verbal argument is more complex. This model permits x_t to be correlated with ω. How, then can we separate ω from x? Begin by noting that I have assumed that $f_t(0) = 0$ for all t. Thus, the preceding argument demonstrates that we can identify the model at the point $x_{t-1} = x_t = x_{t+1} = 0$. Now, consider a slight perturbation from this point by making x_{jt} slightly positive while leaving all inputs in other periods at zero. For now, also suppose that any correlation between ξ_t and x_t does not depend on x_{t-1}. This is only to simplify the argument. Note that in period $t+1$,

$$\mathbb{E}[y_{t+1}|x_{t+1} = x_t = 0] = \mathbb{E}[g(\omega_t)|x_{t+1} = x_t = 0] + \mathbb{E}[\xi_{t+1}|x_{t+1} = x_t = 0] \tag{17}$$

$$\rightarrow \frac{\partial \mathbb{E}[y_{t+1}|x_{t+1} = x_t = 0]}{\partial x_t} = \frac{\partial \mathbb{E}[g(\omega_t)|x_{t+1} = x_t = 0]}{\partial x_t} \tag{18}$$

$$= \int g(\omega_t) \frac{\partial}{\partial x_t} f_{\omega_t|x_{t+1}=x_t=0} d\omega_t \tag{19}$$

and in period t,

$$\mathbb{E}[y_t|x_{t+1}, x_t] = f_t(x_t) + \mathbb{E}[\omega_t|x_{t+1}, x_t] \tag{20}$$

$$\rightarrow \frac{\partial \mathbb{E}[y_t|x_{t+1} = x_t = 0]}{\partial x_t} = \frac{\partial f_t(x_t)}{\partial x_t}\Big|_{x_t=0} + \frac{\partial \mathbb{E}[\omega_t|x_{t+1} = x_t = 0]}{\partial x_t} \tag{21}$$

$$= \frac{\partial f_t(x_t)}{\partial x_t}\Big|_{x_t=0} + \int \omega_t \frac{\partial}{\partial x_t} f_{\omega_t|x_{t+1}=x_t=0} d\omega_t \tag{22}$$

Clearly the observable derivative on the left hand side of 19 provides information regarding the second term in equation 22. As long as the pdf $f_{\omega_t|X_{t+1},X_t}$ is such that the integrals in equations 19 and 22 are one-to-one (once $g(\cdot)$ is known), these equations imply that $\frac{\partial}{\partial x_t} f_t(x_t)$ is identified at zero. Speaking casually, this offers an iterative identification argument. The derivative of the production function at zero identifies the level of $f_t(\cdot)$ in a neighborhood of zero, which can then be used to identify the distribution of ω_t in that neighborhood, and so on. Although the proof does not take this route explicitly, this provides some helpful intuition to understand how $f_t(\cdot)$ and ω_t can be separately identified by three periods of data. I show a more concrete example of this argument in appendix A.1.

To make the point again in words, we are slightly perturbing x_t from zero. As we do this, y_{jt} will move by an amount determined by (i) the derivative of f with respect to x and (ii) the level of endogeneity. To separate these two components, note that (ii) will have an observable effect on y_{jt+1} through $g(\cdot)$. If, conditional on x_{jt+1}, a change in x_t moves y_{jt+1}, this must be through the correlation between x_{jt} and ω_{jt}. Because we know $g(\cdot)$ from the preceding argument at $x_t = 0$ (when the production function is irrelevant), we can invert these observed changes in y_{jt+1} to learn the implied changes in ω_{jt}. In this way, we can separate (i) from (ii) and determine the derivatives of f with respect to x.

3.3 Formal Identification

I now offer a formal identification argument by applying a nonparametric instrumental variables approach following Hu and Schennach (2008). In order to further simply the following assumptions and proof, let

$$Z_t \equiv (X_{t-1}, X_t, X_{t+1})$$

denote a vector of three periods of all observable inputs to production. Note that, conditional on three periods of inputs Z_t, the only variation in firms' output comes from the error term $\omega_t + \eta_t$. Thus, conditional on Z_t, the identification argument for the full model is the same as for the simplistic model without a production function, which is essentially a direct application of the HS result. Next I make the assumptions which are necessary for the main identification result

Assumption 2. *(Exclusion) (i)* $f_{y_{t+1}|Z_t,y_t,\omega_t,y_{t-1}}(y_{t+1}|Z_t,y_t,\omega_t,y_{t-1}) = f_{y_{t+1}|Z_t,\omega_t}(y_{t+1}|Z_t,\omega_t)$ *(ii)* $f_{y_t|Z_t,\omega_t,y_{t-1}}(y_t|Z_t,\omega_t,y_{t-1}) = f_{y_t|Z_t,\omega_t}(y_t|Z_t,\omega_t)$ *for all* $(y_t,\omega_t,y_{t-1}) \in \mathcal{Y}_t \times \Omega_t \times \mathcal{Y}_{t-1}$, *for all* Z_t.

Assumption 3. *(Distinct Eigenvalues) For all* $\omega_t, \omega_t' \in \Omega_t$, *the set* $\{y_{t+1} : f_{y_{t+1}|Z_t,\omega_t}(y_{t+1}|Z_t,\omega_t) \neq f_{y_{t+1}|Z_t,\omega_t'}(y_{t+1}|Z_t,\omega_t')\}$ *has positive probability whenever* $\omega_t \neq \omega_t'$.

Assumption 2, which I call an exclusion restriction, ensures that, conditional on Z_t, (i) y_t and y_{t-1} provide no information regarding y_{t+1} after conditioning on ω_t and

(ii) y_{t-1} provides no information which predicts y_t after conditioning on ω_t. These can be thought of as distributional versions of standard exclusion restrictions applied to instrumental variables, and are implied by workhorse production function models (including the one presented herein) by the first-order Markov assumption as long as η are independent of ω and X. This IV interpretation is central to both the identification argument and the GMM estimation procedure I introduce below. Assumption 3 is a weak assumption which is satisfied if, for example, $g(\cdot)$ is monotonically increasing in ω_{t-1}, or if the conditional variance of ξ_t is monotonic in ω_t. In the proof, which relies on a spectral decomposition argument, this assumption is used to ensure uniqueness of the relevant decomposition. I now define an integral operator $L_{b|a}$, which will be relied on heavily in the proof of the main theorem, and present the final assumption:

Definition. Let a and b denote random variables with supports $\mathcal{A}$ and $\mathcal{B}$. Given two corresponding spaces $\mathcal{G}(\mathcal{A})$ and $\mathcal{G}(\mathcal{B})$ of functions with domains $\mathcal{A}$ and $\mathcal{B}$, respectively, let $L_{b|a}$ denote the operator mapping elements of $\mathcal{G}(\mathcal{A})$ to $\mathcal{G}(\mathcal{B})$ by

$$[L_{b|a}g](b) \equiv \int_{\mathcal{A}} f_{b|a}(b|a)g(a)da,$$

Assumption 4. *(Injectivity) The operators $L_{y_t|Z_t,\omega_t}$ and $L_{y_{t-1}|Z_t,y_t}$ are injective (conditional on Z_t) over the set of $\mathcal{L}^1$ bounded functions in their respective domains for every Z_t*

This assumption is standard in the literature following the Hu and Schennach (2008) result. It is closely related to the bounded completeness assumptions made in the nonparametric IV literature, and more primitive sufficient conditions for injectivity of these operators are discussed that literature, as are families of distributions which satisfy this assumption (Newey and Powell, 2003). Now I claim the following theorem and remark to summarize my identification results

Theorem 3.1. *Under assumptions 1, 2, 3, and 4, the distributions $f_{y_{t+1}|\omega_t}$, $f_{y_t|\omega_t}$, $f_{\omega_t|y_{t-1}}$ are identified, as are the corresponding densities conditional on Z_t.*

Proof. See Appendix A. □

Remark. Given Theorem 3.1, the production function $f_t(X)$ is also nonparametrically identified.

This is shown by the following

$$\mathbb{E}[y_t|x_t,\omega_t] = \int y_t f_{y_t|Z_t,\omega_t}(\cdot|Z_t,\omega_t)dy_t \tag{23}$$

$$= f_t(x_t) + \omega_t \tag{24}$$

where the first line makes explicit that the left hand side is a functional of $f_{y_t|Z_t,\omega_t}(\cdot|Z_t,\omega_t)$, and the second line comes from Assumption 1. Now, we can take derivatives of this function with respect to the inputs to production:

$$\begin{aligned}\frac{\partial \mathbb{E}[y_t|x_t,\omega_t]}{\partial x} &= \int y_{jt}\frac{\partial}{\partial x} f_{y_t|Z_t,\omega_t}(\cdot|Z_t,\omega_t)dy_t \\ &= \frac{\partial}{\partial x} f_t(x_t)\end{aligned}$$

The first equality comes from the definition of the conditional expectation on the left hand side and by the boundedness of the conditional density of y_t. This demonstrates that all derivatives (iterating appropriately for higher-order derivatives) of $f_t(X)$ are functionals of the density $f_{y_t|Z_t,\omega_t}(\cdot|Z_t,\omega_t)$. Thus, we can treat all derivatives of $f_t(\cdot)$ as known. Noting that equation 13 sets the location of $f_t(\cdot)$, the additional knowledge of all derivatives of $f_t(\cdot)$ is sufficient for knowledge of $f_t(\cdot)$ itself. So, knowledge of $f_{y_t|Z_t,\omega_t}(\cdot|Z_t,\omega_t)$ is sufficient under the assumptions of Theorem 3.1 for knowledge of $f_t(\cdot)$.

Remark. Because we observe y_{t-1} (and its distribution), Theorem 3.1 implies that we identify the density f_{ω_t}.

This arises immediately from the Theorem, because we need only integrate y_{t-1} (which is observed) out of the identified distribution $f_{\omega_t|y_{t-1}}$. To summarize, assumptions 1-4 imply that the production function, the distribution of productivity, and the dependence of inputs on productivity are all identified. The latter comes from the fact that each distribution in Theorem 3.1 are identified conditional on the vector of inputs Z_t.

I should note here that the HS identification result is far more general than as applied here. For example, Sasaki (2015), Hu and Shum (2012), and Arellano and Bonhomme (2016) (and others) all consider applications in which the persistent unobservable need not be separable. Sasaki (2015) uses arguments related to HS to demonstrate that, as long as an additional proxy variable is observed (or constructed from additional periods of data), a dynamic model with nonseparable heterogeneity and dynamic selection is identified. In this sense, the model here is less general than may be possible to identify given the assumptions. Said otherwise, perhaps there are weaker identifying assumptions than those used here which could also identify the production function. However, the purpose here is to tailor the identification strategy to the most general case in order to demonstrate that the scalar unobservable (and other substitute assumptions) are unnecessary to identify the standard production function. For this, the HS approach seems well-suited.

3.4 Identification in the Standard Production Function

3.4.1 Timing of Capital Choices

In many ways the assumptions made to prove Theorem 3.1 are too general to interpret easily in a production function context. Some of the standard assumptions of the production function estimation literature are more straightforward, and enforcing them in addition to the preceding identifying assumptions will aid us in choosing instruments for estimation below. Perhaps most importantly is a timing assumption. It is notable, and perhaps surprising, that the preceding argument does not rely on a timing assumption. In principle, each input x_t could be correlated with ω_t (including the innovation ξ_t). Though this flexibility is informative in *identifying* the production function, in practice it makes estimation of the model difficult, as there may be little exogenous variation in x_t in each period which can identify the production function parameters. Separate the input vector x into two components: capital, which is denoted by k, and all other inputs v which are assumed to be flexible. Following the literature, I assume that capital k_{jt} is a function of lagged capital and lagged investment, which implies the following timing assumption

Assumption 5. *k_t is chosen before ξ_{t+1} is realized.*

3.4.2 Flexible Input Elasticity and Prices

To the contrary, I permit all other inputs v to be chosen at period t, meaning that v_t may be correlated with ξ_t. Much attention has been paid in the literature to the identification of the elasticity of output with respect to these flexible inputs. Ackerberg, Caves and Frazer (2015), for example, demonstrate the non-identification of the output elasticity of labor in the Levinsohn and Petrin (2003) estimation approach in a value-added model. Gandhi, Navarro and Rivers (2016), on the other hand, address the identification of the flexible input elasticity in a gross output setting. In each of these cases, the identification of the flexible input elasticity is complicated by the scalar unobservable assumption.[7] For instance, both sets of authors discuss the possibility that variation in input prices, observed or not, could ameliorate their respective problems, and in both papers the authors conclude correctly that any *unobserved* input price variation would violate the scalar unobservable assumption.

Since, in the proposed approach, I make no assumptions regarding the role of unobservables in input demand functions, unobserved (and persistent) input price variation could in principle serve to identify the flexible input elasticity. As input prices likely do vary across firms in practice (e.g. quantity discounts, differing prices from foreign/domestic sources), this provides substantial additional identifying variation for

[7]This is perhaps most clear in Gandhi, Navarro and Rivers (2016), who show that the flexible input elasticity in gross output production functions is not identified by the standard control function approach. This is because the scalar unobservable implies that there are no instruments available to generate variation in this input conditional on productivity and lagged inputs.

input elasticities. In Appendix D I offer some initial thoughts on identifying the distribution of markups when firms have monopsony power in input markets.[8] Most existing work deriving markups from production function estimates assumes that firms are price takers in at least one input market.[9] I show that this can be relaxed whenever two flexible inputs are observed (e.g. labor and materials).

3.5 Comparing Assumptions

Because the production function is identified without making a cost-minimization assumption, one may pause here to wonder whether making such an assumption to recover markups makes the preceding discussion irrelevant. On that topic I make two points. First, this cost-minimization first-order condition can hold even when the firm solves a complex and/or dynamic profit-maximization problem in its other inputs (e.g. capital, labor). Because the production function also leaves these as possibilities (up to assumption 5), we have still left much of the input process (including any influence of demand heterogeneity) free. Second, note that the strength of a cost minimization assumption depends in part on the types and number of available flexible inputs. It is particularly strong in settings using financial data, where the meaning and flexibility of measures of intermediate input spending are unclear. Because we require only that the firm solves a cost minimization problem for the *most* flexible input conditional on all others, additional inputs relax this assumption loosely speaking. Further, additional flexible inputs may allow us to introduce errors into the cost-minimization process. Morlacco (2018) takes this approach to study monopsony power in France, and I discuss some initial thoughts on an extension in Appendix D.

4 Estimation

4.1 Production Function

In this section I build on the foundations laid in Hu, Huang and Sasaki (2017) in translating the identification argument (also closely related to their argument) into a simple estimation procedure. To fix ideas, I rewrite the model of interest in terms of the parameters we wish to estimate (β_0, ρ_0) and provide an additional assumption

$$y_{jt} = f(X_{jt}; \beta_0) + \omega_{jt} + \eta_{jt} \tag{25}$$

$$\omega_{jt} = g(\omega_{jt-1}; \rho_0) + \eta_{jt} \tag{26}$$

Assumption 6. *(Independence) (i)* $\eta_{jt'} \perp \omega_{jt}, Z_{jt}, \xi_{jt}$ *for all* t', t*, and (ii)* η_{jt}, η_{jt+1}*, and* η_{jt-1} *are jointly independent*

[8] The approach therein relies on an application of Kotlarski's Theorem to recover the markup distribution. Recently, Kato, Sasaki and Ura (2018) provide uniform confidence bands on the density of markups derived from this approach.

[9] In fact, most assume that firms are price takers in all flexible input markets. Morlacco (2018) is a notable exception.

Assumption 6 is stronger than the assumptions used for identification, but will ensure that the moments I construct are valid. It also makes clear one of the costs of the estimation approach I propose. Whereas the first stage of the control function approach can be easily adjusted to permit correlation between η_t within firm, over time, this correlation would invalidate the present estimation procedure. When the researcher suspects serial correlation in η, an alternative estimation procedure should be used (e.g. the sieve MLE approach in HS).

In short, the preceding identification argument can be interpreted as demonstrating that, conditional on three periods of inputs, $(\omega_{jt-1} + \eta_{jt-1})$ can be used as an instrument in the regression of $(\omega_{jt+1} + \eta_{jt+1})$ on $(\omega_{jt} + \eta_{jt})$. Thus, I modify the Hu, Huang and Sasaki (2017) procedure by replacing the moments they use to identify input demand functions with moments which instead make use of this instrumental variables argument.[10] The procedure is as follows. For each potential vector of production function parameters β (which yield a guess of a production function $\tilde{f}(\cdot;\beta)$), we can construct

$$\tilde{y}_{jt}(\beta) = y_{jt} - \tilde{f}(X_{jt};\beta) = (\omega_{jt} + \eta_{jt})(\beta) \tag{27}$$

where I emphasize that this constructed residual is dependent on β and that it is a guess of the total residual term $(\omega_{jt} + \eta_{jt})$. Next, I assume that the first-order Markov function $g(\cdot;\rho_0)$ is well-approximated by a $q-$order polynomial with coefficients ρ_0. Note that under this assumption, if there were no measurement error term η_{jt} (i.e. $\eta_{jt} \equiv 0$ for all j, t), we could construct the period-t innovation to productivity as

$$\begin{aligned}\tilde{y}_{jt}(\beta) - \sum_{p=1}^{q} \rho_p \tilde{y}_{jt-1}^p(\beta) &= \omega_{jt}(\beta) - \sum_{p=1}^{q} \rho_p \omega_{jt-1}^p \\ &\equiv \xi_{jt+1}\end{aligned}$$

I suppress the dependence of ξ_{jt} on β, but note that at (β_0, ρ_0) this is the true productivity innovation.[11] Unfortunately, the first equality no longer holds when η_{jt} is reintroduced. Consider the simple case of $q = 2$, and define the feasible residual $\tilde{\xi}$ as follows:

$$\begin{aligned}\tilde{y}_{jt+1}(\beta) - \sum_{p=1}^{q} \rho_p \tilde{y}_{jt}^p(\beta) &= \omega_{jt+1} - \rho_1 \omega_{jt} - \rho_2 \omega_{jt}^2 + \eta_{jt+1} - \rho_1 \eta_{jt} \\ &\quad - \rho_2 \eta_{jt}^2 - 2\rho_2 \omega_{jt} \eta_{jt} \\ &= \xi_{jt+1} + \eta_{jt+1} - \rho_1 \eta_{jt} - \rho_2 \eta_{jt}^2 - 2\rho_2 \omega_{jt} \eta_{jt} \equiv \tilde{\xi}_{jt+1}\end{aligned}$$

[10] Alternatively, researchers interested in estimating the input demand functions HHS identify could stack their moments with mine.

[11] This is a special case of the approach considered by Ackerberg and Hahn (2015).

Note that η_{jt}, η_{jt-1}, and ξ_{jt} are assumed jointly independent by Assumption 6. In this case, the faux-innovations $\tilde{\xi}_{jt+1}$ constructed here differ from the true innovation ξ_{jt+1} by an additional error term which is correlated with $\tilde{y}_{jt}$. Thus, even with knowledge of β_0, the regression of $\tilde{y}_{jt+1}$ on its lag will yield biased estimates of ρ. My identification argument offers the use of $\tilde{y}_{jt-1}$ as an instrument to solve this problem. Because η_t and $\tilde{y}_{t-1}$ are independent, the vector ρ_0 can be identified by the moment condition $\mathbb{E}[\tilde{\xi}_{jt+1}|\tilde{y}_{jt-1}] = 0$. To see this note that

$$\begin{aligned} cov(\tilde{\xi}_{jt+1}, \tilde{y}_{jt-1}) &= cov(\tilde{\xi}_{jt+1}, \omega_{jt-1}) \\ &= cov(\xi_{jt+1} + \eta_{jt+1} - \rho_1\eta_{jt} - \rho_2\eta_{jt}^2 - 2\rho_2\omega_{jt}\eta_{jt}, \omega_{jt-1}) \\ &= 0 \end{aligned}$$

where $cov(\eta_{jt}^2, \omega_{jt-1}) = 0$ by the full independence of ω_{jt-1} and η_{jt}, and $cov(\omega_{jt}\eta_{jt}, \omega_{jt-1}) = 0$ by iterating expectations:

$$\mathbb{E}[\omega_{jt}\eta_{jt}\omega_{jt-1}] = \mathbb{E}[\omega_{jt}\omega_{jt-1}\mathbb{E}[\eta_{jt}|\omega_{jt}\omega_{jt-1}]] = 0 \tag{28}$$

Similar work follows for the covariance of $\tilde{\xi}_{jt+1}$ and $\tilde{y}_{jt-1}^2$. In order to use these covariance restrictions in estimation, we can translate them into moments relating $\tilde{\xi}_{jt+1}$ to $\tilde{y}_{jt-1}$. However, the relevant moment $\mathbb{E}[\tilde{\xi}_{jt+1}\tilde{y}_{jt-1}]$ is equal to a non-zero constant (and similarly for $\tilde{y}_{jt-1}^2$) because although $\tilde{\xi}_{jt+1}$ and $\tilde{y}_{jt-1}$ are uncorrelated, both can have non-zero mean. So, in an abuse of notation, let us redefine $\tilde{\xi}_{jt+1} = \tilde{\xi}_{jt+1} - \mathbb{E}[\tilde{\xi}_{jt+1}]$ in constructing the moments below.[12]

I now combine these moments with those which are standard in the estimation of the production function to give the following set of moments (for the quadratic case)

$$\mathbb{E}\begin{bmatrix} \tilde{\xi}_{jt+1}\tilde{y}_{jt-1} \\ \tilde{\xi}_{jt+1}\tilde{y}_{jt-1}^2 \\ \tilde{\xi}_{jt+1}\boldsymbol{Z}_{jt+1} \end{bmatrix} = 0 \tag{29}$$

where $\boldsymbol{Z}_{jt}$ are standard instruments in production function estimation (e.g. $\boldsymbol{Z}_{jt} = x_{jt-1}$), which are valid under the assumption that ξ_{jt+1} is unpredictable in all periods before $t+1$ and by the full independence of η and Z (Assumption 6).

Before moving on, note that this estimation procedure, while starkly simple to implement, is novel in the literature. Both ACF and Gandhi, Navarro and Rivers (2016) mention the linearity of the Markov process required by dynamic panel methods as one of the main drawbacks of such methods. More recently, papers like Bond et al. (2020) and Shenoy (2020) have made similar critiques of the scalar unobservable assumptions as in this paper and have also only considered linear Markov processes in their discussions of dynamic panel methods. To the contrary, the previous section

[12] I show that this re-centering generates valid moments in Appendix C.

demonstrates that the dynamic panel identification argument can be extended to any first-order Markov process, and this section proves that the quadratic case can be easily handled in a GMM framework. In Appendix C I show that the same is true for cubic Markov processes, although a slightly more involved adjustment must be made to the residual $\tilde{\xi}$, and an additional parameter must be estimated. In practice, most studies of production functions using standard approaches approximate the Markov process $g(\cdot)$ with a quadratic or cubic function, meaning these two cases cover the most common implementations.

4.2 Ex-post Shock

In some cases, in particular when estimating markups using the ratio estimator as in De Loecker, Eeckhout and Unger (2018), researchers may be interested in the distribution of the ex-post shock η. In many other cases, one may be interested in the distribution of productivity ω_t and changes in that distribution over time. Taking the production function estimates from the preceding procedure as given, note that we can now construct estimates of the sum of ω and η, as $\tilde{y}_{jt} = \omega_{jt} + \eta_{jt} = y_{jt} - f(x_{jt}; \beta)$. To estimate the distributions of ω and η separately, we can use the following decomposition which mirrors the key identifying equation of Theorem 3.1:

$$f_{\tilde{y}_{jt+1},\tilde{y}_{jt}|\tilde{y}_{jt-1}} = \int f_{\tilde{y}_{t+1}|\omega_t}(\tilde{y}_{t+1}|\omega_t) f_{\tilde{y}_t|\omega}(\tilde{y}_t|\omega_t) f_{\omega_t|\tilde{y}_{t-1}}(\omega_t|\tilde{y}_{t-1}) d\omega_t \tag{30}$$

This equality holds under exactly the same assumptions needed for Theorem 3.1. As discussed in HS, one could apply this equation to estimation by approximating each distribution by a sieve and estimating the sieve parameters via maximum likelihood. As long as a nonparametric approach is taken in this step, the fact that β is estimated in a prior step is unlikely to substantively affect inference on these distributions.

4.3 Higher Order Markov Processes

Although most empirical applications assume either quadratic or cubic Markov processes, which can be handled by the moments in Section 4.1 or in Appendix C, and although the identification result is nonparametric, one may wish to estimate models with higher order Markov processes. In these cases, there are two reasonable approaches. As long as sufficiently many reasonable instruments are available, one could simple extend the algebra in Appendix C to higher order cases. This approach requires estimating higher-order moments of the measurement error distribution, which may raise numerical issues and which may be difficult to match to appropriate instruments. An easier alternative exists in cases in which the parameters of the Markov process are not of particular interest. In these settings, one can simply include higher orders of $\tilde{y}_{jt-1}$ as instruments to estimate the higher order terms. Estimates of these terms will be biased, but (i) the bias is simple to characterize in terms of moments of the measurement error distribution and (ii) estimates of the production function will be

unbiased. A proof of this will be included in the next draft. See Hu, Huang and Sasaki (2017) for details.

5 Simulations

Part of the appeal of the method I propose is that it makes minimal assumptions regarding the way firms choose the amounts of capital, labor, and materials to use each year. To test the practical performance of this method, I offer simulation examples in which I compare the performance of the estimator developed in the preceding section to the estimation procedure used by DLEU, which applies a procedure like that in Levinsohn and Petrin (2003). As a simple case to study, I consider a Cobb-Douglas production (in logs) in three inputs k, l, and m:

$$y_{jt} = \beta^l l_{jt} + \beta^k k_{jt} + \beta^m m_{jt} + \omega_{jt} + \eta_{jt} \tag{31}$$

with $\beta^l = 0.4$, $\beta^k = 0.3$, and $\beta^m = 0.3$. Following the traditional production function literature, k denotes capital, l represents labor, and m is a perfectly flexible material input. I specify the first-order Markov process for ω_{jt} as a quadratic function in its lag

$$\omega_{jt} = \rho_1 \omega_{jt-1} + \rho_2 \omega_{jt-1}^2 + \xi_{jt} \tag{32}$$

with $\rho_1 = 1$, $\rho_2 = -0.025$, and $\xi_{jt} \sim N(0, 0.05^2)$. I specify reduced-form dynamic processes for capital and labor demand

$$k_{jt} = 0.9 k_{jt-1} + E * \omega_{jt-1} + \gamma_{jt}^k \tag{33}$$

$$l_{jt} = 0.9 l_{jt-1} + E * \omega_{jt-1} + \gamma_{jt}^l \tag{34}$$

where $\gamma^k, \gamma^l \sim N(0, 0.5^2)$ are drawn independently and E controls the extent to which k and l are chosen endogenously. For each simulation sample, I simulate between and 1000 and 4000 firms and compare the method proposed here to the approach taken in DLEU (closest in spirit to Levinsohn and Petrin (2003)), both estimated via just-identified two-step GMM. Following the standard in the literature, I simulate firms for 10 periods and use only the final three periods to estimate the model. In the first set of simulations I specify a reduced-form material input process which satisfies the scalar unobservable assumption:

$$m_{jt} = 0.4 l_{jt}^2 + 0.4 k_{jt}^2 + \omega_{jt} \tag{35}$$

This admittedly arbitrary function was chosen because it should be easily approximated by a polynomial in the first stage control function. In table 1 I present the means and standard deviations of the estimated production function and Markov process coefficients over 1000 simulation samples with E set to 0. Although the variance of DLEU estimates are smaller in general, I find that the proposed method performs approximately as well on average as the DLEU procedure in this case, which is en-

couraging. Next I simulate samples in which DLEU is misspecified by generating an input demand function which does not satisfy the scalar unobservable assumption. I violate this assumption in two ways. In the first set of simulations, materials are chosen according to

$$\tilde{m}_{jt} = m_{jt} + \gamma^m_{jt} \tag{36}$$

where m_{jt} denotes the input demand function in 35, $\gamma^m_{jt} \sim N(0, 0.4^2)$. In table 2 I present the results of both estimation procedures on these simulated samples, which clearly show that the addition of this small error γ^m produces biases in the DLEU estimation procedure, even in large samples and in particular with respect to β^m. The size of this bias is larger than 50% of the true parameter value, meaning markups (which rely directly on β^m) would be significantly overestimated in this setting. To the contrary, the panel estimator remains unbiased for all production function parameters. Next, I make capital and labor endogenous (by setting $E = 0.1$ above) to demonstrate that violations of the scalar unobservable assumption can bias estimates of other parameters (i.e. β^k and β^l). I present results from this exercise in table 3. Though control function estimates of β^k and β^l are persistently and significantly biased, estimates from the method introduced here remain unbiased.

Finally, keeping $E = 0.1$, I violate the scalar unobservable by introducing a more complicated shock into the input demand function:

$$\bar{m}_{jt} = 0.4l^2_{jt} + 0.4k^2_{jt} + \omega_{jt}\tilde{\gamma}^m_{jt} \tag{37}$$

In this equation, $\tilde{\gamma}^m_{jt} \sim N(0, 0.1^2)$ interacts with ω, meaning that the deviation from the scalar unobservable is nonlinear. To the best of my knowledge, no existing studies of the production function are robust to a data-generating of this form, given that (i) the scalar unobservable assumption is violated and (ii) this input demand function need not be optimal (i.e. I make no assumption about the output demand firms face or the prices they face). I report means and standard deviations of estimates of DLEU and panel estimates in table 4, which shows that control function estimates of all parameters are biased.[13] This is in stark contrast to panel estimates, which are nearly unbiased even with a sample of 1000 firms. The simulations in this section, which cover only a small subset of potential deviations from a scalar unobservable, are suggestive of the importance of the proposed approach.

6 Concluding Remarks

In this paper I demonstrate that the traditional production function model is nonparametrically identified without requiring the restrictive scalar unobservable assumption or any knowledge of the demand firms face. The proof I provide is a direct

[13] The similarity of these estimates to those in the previous table are coincidental.

application of the result in Hu and Schennach (2008), and is thus closely related to a number of existing identification results in the literature. Although this nonclassical measurement error result has been applied many times to date, this is the first paper to apply it to the production function model in the way shown herein, and is also the first to use a third period of data to relax the scalar unobservable assumption. I also provide a GMM estimator for the simplest applications of this argument, which makes use of standard assumptions about the timing of input choices and is easy to implement. Monte Carlo simulations indicate that the proposed GMM estimator works well. Though my estimates have higher variance than those from the control function approach when the scalar unobservable assumption is satisfied, the former perform much better than the latter whenever that crucial assumption is violated.

Although this paper is explicitly about estimation of a production function, the approach taken here is a much more general dynamic panel approach. It is increasingly common that researchers have panel data available with more than three time periods for each firm or individual. In many such cases, one may be concerned about persistent unobservables which are correlated with covariates. Examples include studies of the education production function, in which student ability may be time-varying and correlated with inputs given to the student, as well as wage studies in which wages may be correlated with unobserved components of ability.

References

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer. 2015. "Identification properties of recent production function estimators." *Econometrica*, 83(6): 2411–2451.

Ackerberg, Daniel, and Jinyong Hahn. 2015. "Some non-parametric identification results using timing and information set assumptions." Working Paper.

Arellano, Manuel, and Stéphane Bonhomme. 2016. "Nonlinear panel data estimation via quantile regressions." *The Econometrics Journal*, 3(19): C61–C94.

Balat, Jorge, Irene Brambilla, and Yuya Sasaki. 2016. "Heterogeneous Firms: Skilled-Labor Productivity and Export Destinations."

Berry, Steven T, and Philip A Haile. 2014. "Identification in differentiated products markets using market level data." *Econometrica*, 82(5): 1749–1797.

Blum, Bernardo S, Sebastian Claro, Ignatius Horstmann, and David A Rivers. 2018. "The ABCs of Firm Heterogeneity: The Effects of Demand and Cost Differences on Exporting."

Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of econometrics*, 87(1): 115–143.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. 2020. "Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data." National Bureau of Economic Research.

Collard-Wexler, Allan, and Jan De Loecker. 2016. "Production function estimation with measurement error in inputs." National Bureau of Economic Research.

Compiani, Giovanni. 2018. "Nonparametric Demand Estimation in Differentiated Products Markets."

Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica*, 78(3): 883–931.

De Loecker, Jan. 2011. "Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity." *Econometrica*, 79(5): 1407–1451.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and firm-level export status." *American Economic Review*, 102(6): 2437–71.

De Loecker, Jan, and Jan Eeckhout. 2018. "Global market power." National Bureau of Economic Research.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2018. "The Rise of Market Power and the Macroeconomic Implications." UCL mimeo.

De Loecker, Jan, Pinelopi K Goldberg, Amit K Khandelwal, and Nina Pavcnik. 2016. "Prices, markups, and trade reform." *Econometrica*, 84(2): 445–510.

Demirer, Matt. 2019. "Production Function Estimation with Factor-Augmenting Technology: An Application to Markups." Working Paper.

Doraszelski, Ulrich, and Jordi Jaumandreu. 2017. "Measuring the Bias of Technological Change." *Forthcoming, Journal of Political Economy*.

Flynn, Zach, Amit Gandhi, and James Traina. 2019. "Identifying market power from production data." Working paper.

Freyberger, Joachim. 2017. "Non-parametric Panel Data Models with Interactive Fixed Effects." *The Review of Economic Studies*, 85(3): 1824–1851.

Gandhi, Amit, Salvador Navarro, and David Rivers. 2016. "On the Identification of Production Functions: How Heterogeneous is Productivity?"

Grieco, Paul, Joris Pinkse, and Margaret Slade. 2018. "Brewed in North America: Mergers, marginal costs, and efficiency." *International Journal of Industrial Organization*, 59: 24–65.

Hall, Robert E. 1988. "The relation between price and marginal cost in US industry." *Journal of political Economy*, 96(5): 921–947.

Hu, Yingyao, and Matthew Shum. 2012. "Nonparametric identification of dynamic models with unobserved state variables." *Journal of Econometrics*, 171(1): 32–44.

Hu, Yingyao, and Susanne M Schennach. 2008. "Instrumental variable treatment of nonclassical measurement error models." *Econometrica*, 76(1): 195–216.

Hu, Yingyao, Guofang Huang, and Yuya Sasaki. 2017. "Estimating Production Functions with Robustness Against Errors in the Proxy Variables."

il Kim, Kyoo, Amil Petrin, and Suyong Song. 2016. "Estimating production functions with control functions when capital is measured with error." *Journal of Econometrics*, 190(2): 267–279.

Jaumandreu, Jordi. 2018. "Dangerous Shortcuts: Ignoring Marginal Cost Determinants in Markup Estimation." mimeo, Boston University.

Kato, Kengo, and Yuya Sasaki. 2018. "Uniform confidence bands in deconvolution with unknown error distribution." *Journal of Econometrics*, 207(1): 129–161.

Kato, Kengo, Yuya Sasaki, and Takuya Ura. 2018. "Inference based on Kotlarski's Identity." *arXiv preprint arXiv:1808.09375*.

Klette, Tor Jakob, and Zvi Griliches. 1996. "The inconsistency of common scale estimators when output prices are unobserved and endogenous." *Journal of Applied Econometrics*, 343–361.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies*, 70(2): 317–341.

Li, Tong, and Yuya Sasaki. 2017. "Constructive Identification of Heterogeneous Elasticities in the Cobb-Douglas Production Function." *arXiv preprint arXiv:1711.10031*.

Morlacco, Monica. 2018. "Market power in input markets: Theory and evidence from french manufacturing." Working Paper.

Newey, Whitney K, and James L Powell. 2003. "Instrumental variable estimation of nonparametric models." *Econometrica*, 71(5): 1565–1578.

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

Sasaki, Yuya. 2015. "Heterogeneity and selection in dynamic panel data." *Journal of Econometrics*, 188(1): 236–249.

Shenoy, Ajay. 2020. "Estimating the Production Function Under Input Market Frictions." *Review of Economics and Statistics*, 1–45.

Shiu, Ji-Liang, and Yingyao Hu. 2013. "Identification and estimation of nonlinear dynamic panel data models with unobserved covariates." *Journal of Econometrics*, 175(2): 116–131.

Table 1: Panel and DLEU Estimates: Baseline DGP

		β^k	β^l	β^m	ρ_1	ρ_2
Truth:		0.4	0.3	0.3	1	-0.025
DLEU	$N = 1000$	0.4	0.299	0.295		
		(0.042)	(0.043)	(0.027)		
	$N = 2000$	0.4	0.299	0.296		
		(0.029)	(0.03)	(0.024)		
	$N = 4000$	0.401	0.3	0.299		
		(0.021)	(0.021)	(0.013)		
Panel	$N = 1000$	0.402	0.295	0.291	1.001	-0.024
		(0.125)	(0.126)	(0.106)	(0.017)	(0.006)
	$N = 2000$	0.401	0.299	0.296	1	-0.025
		(0.085)	(0.084)	(0.094)	(0.012)	(0.004)
	$N = 4000$	0.4	0.3	0.299	1	-0.025
		(0.059)	(0.057)	(0.047)	(0.008)	(0.003)

Note: Cobb-Douglas production function estimates from DLEU and panel approaches. Data simulated such that DLEU is correctly specified with reduced-form input choices as described in text. Standard deviations of estimates reported in parentheses.

Table 2: Panel and DLEU Estimates: Additive Noise

		β^k	β^l	β^m	ρ_1	ρ_2
Truth:		0.4	0.3	0.3	1	-0.025
DLEU	$N = 1000$	0.401	0.302	0.543		
		(0.048)	(0.048)	(0.347)		
	$N = 2000$	0.403	0.301	0.573		
		(0.034)	(0.035)	(0.357)		
	$N = 4000$	0.401	0.301	0.593		
		(0.03)	(0.03)	(0.366)		
Panel	$N = 1000$	0.399	0.305	0.27	1.002	-0.024
		(0.145)	(0.142)	(0.284)	(0.02)	(0.008)
	$N = 2000$	0.399	0.298	0.285	1.001	-0.025
		(0.086)	(0.089)	(0.187)	(0.013)	(0.005)
	$N = 4000$	0.403	0.298	0.289	1	-0.025
		(0.083)	(0.077)	(0.175)	(0.011)	(0.004)

Note: Cobb-Douglas production function estimates from DLEU and panel approaches. Data simulated with an additive idiosyncratic error in choice of m. Standard deviations of estimates reported in parentheses.

Table 3: Panel and DLEU Estimates: Endogenous k, l

		β^k	β^l	β^m	ρ_1	ρ_2
Truth:		0.4	0.3	0.3	1	-0.025
DLEU	$N = 1000$	0.612	0.516	0.523		
		(0.064)	(0.065)	(0.155)		
	$N = 2000$	0.615	0.515	0.524		
		(0.045)	(0.046)	(0.118)		
	$N = 4000$	0.613	0.513	0.527		
		(0.032)	(0.032)	(0.075)		
Panel	$N = 1000$	0.415	0.325	0.291	0.997	-0.025
		(0.177)	(0.182)	(0.174)	(0.027)	(0.01)
	$N = 2000$	0.408	0.299	0.297	1	-0.025
		(0.113)	(0.106)	(0.088)	(0.016)	(0.006)
	$N = 4000$	0.4	0.302	0.3	1	-0.025
		(0.063)	(0.064)	(0.057)	(0.01)	(0.004)

Note: Cobb-Douglas production function estimates from DLEU and panel approaches. Data simulated with an idiosyncratic error in choice of m and with endogenous k and l. Standard deviations of estimates reported in parentheses.

Table 4: Panel and DLEU Estimates: Nonseparable Noise

		β^k	β^l	β^m	ρ_1	ρ_2
Truth:		0.4	0.3	0.3	1	-0.025
DLEU	$N = 1000$	0.615	0.514	0.519		
		(0.067)	(0.064)	(0.192)		
	$N = 2000$	0.614	0.515	0.525		
		(0.046)	(0.046)	(0.109)		
	$N = 4000$	0.614	0.513	0.527		
		(0.032)	(0.032)	(0.075)		
Panel	$N = 1000$	0.431	0.324	0.3	0.996	-0.025
		(0.247)	(0.215)	(0.203)	(0.029)	(0.01)
	$N = 2000$	0.404	0.307	0.299	0.998	-0.025
		(0.11)	(0.1)	(0.087)	(0.017)	(0.006)
	$N = 4000$	0.401	0.302	0.299	1	-0.025
		(0.063)	(0.062)	(0.056)	(0.01)	(0.004)

Note: Cobb-Douglas production function estimates from DLEU and panel approaches. Data simulated with a multiplicative (i.e. nonseparable) idiosyncratic error. Standard deviations of estimates reported in parentheses.

Appendices

A Proof of Theorem 3.1

Proof. As Theorem 3.1 and the assumptions which provide it are a special case of HS, here I reproduce their argument in my particular setting and refer the reader there for more detail. In order to apply their result as originally stated, we require a known functional to calculate ω from y conditional on ω and x for every x. As discussed by Arellano and Bonhomme (2016), this mapping need not be known as long as it is a known transformation of the data. To generate such a transformation, let us assume a single dimensional x to ease notation and note that the assumed location of $f_t(\cdot)$ implies that

$$\mathbb{E}[y_{jt}|\omega_{jt}, x_{jt} = 0] = \omega_{jt}$$

Thus, at $x_{jt} = 0$, the conditional expectation above is sufficient. Then note that changes in x_{jt} conditional on ω_{jt} only move the production function:

$$\frac{\partial \mathbb{E}[y_{jt}|\omega_{jt}, x_{jt}]}{\partial x_{jt}} = \frac{\partial f_t(x_{jt})}{\partial x_{jt}}$$

this implies that, at each $\tilde{x}$, ω_{jt},

$$\begin{aligned}\omega_{jt} &\equiv \mathbb{E}[y_{jt}|\omega_{jt}, \tilde{x}] - f_t(\tilde{x}) \\ &= \mathbb{E}[y_{jt}|\omega_{jt}, \tilde{x}] - \int_0^{\tilde{x}} \frac{\partial \mathbb{E}[y_{jt}|\omega_{jt}, x]}{\partial x} dx\end{aligned}$$

This can clearly be calculated iteratively (i.e. increasing x from zero) and for each dimension of a multivariate x. The rest of the proof follows directly from HS. By the definition of conditional densities and Assumption 2,

$$\begin{aligned}f_{y_{t+1},y_t|Z_t,y_{t-1}} &= \int f_{y_{t+1},y_t,\omega_t|Z_t,y_{t-1}}(y_{t+1}, y_t, \omega_t|Z_t, y_{t-1}) d\omega_t \\ &= \int f_{y_{t+1}|Z_t,y_t,y_{t-1},\omega_t}(y_{t+1}|Z_t, y_t, y_{t-1}, \omega_t) f_{y_t,\omega_t|Z_t,y_{t-1}}(y_t, \omega_t|Z_t, y_{t-1}) d\omega_t \\ &= \int f_{y_{t+1}|Z_t,\omega_t}(y_{t+1}|Z_t, \omega_t) f_{y_t,\omega_t|Z_t,y_{t-1}}(y_t, \omega_t|Z_t, y_{t-1}) d\omega_t \\ &= \int f_{y_{t+1}|Z_t,\omega_t}(y_{t+1}|Z_t, \omega_t) f_{y_t|Z_t,\omega_t,y_{t-1}}(y_t|Z_t, \omega_t) f_{\omega_t|Z_t,y_{t-1}}(\omega_t|Z_t, y_{t-1}) d\omega_t \\ &= \int f_{y_{t+1}|Z_t,\omega_t}(y_{t+1}|Z_t, \omega_t) f_{y_t|Z_t,\omega}(y_t|Z_t, \omega_t) f_{\omega_t|Z_t,y_{t-1}}(\omega_t|Z_t, y_{t-1}) d\omega_t\end{aligned}$$

This demonstrates that the observed distribution $f_{y_{t+1},y_t|Z_t,y_{t-1}}$ can be decomposed into the distributions of interest in Theorem 3.1. Though I explicitly condition on Z_t in

the above equations in order to connect the above to Theorem 3.1, the remaining work for the rest of the proof is unnecessarily burdened by conditioning every distribution on Z_t. To ease this burden, I suppress this conditioning, though the entire argument should be interpreted as for a fixed Z_t. To begin, I rewrite the above decomposition for a fixed Z_t, suppressing that notation, for reference:

$$f_{y_{t+1},y_t|y_{t-1}} = \int f_{y_{t+1}|\omega_t}(y_{t+1}|\omega_t) f_{y_t|\omega}(y_t|\omega) f_{\omega_t|y_{t-1}}(\omega_t|y_{t-1}) d\omega_t \tag{38}$$

Now, it remains to show that this decomposition is unique. Toward that end, I define the following two additional operators. Let

$$\begin{aligned} L_{y_{t+1};y_t|y_{t-1}} &\equiv \int f_{y_{t+1}y_t|y_{t-1}}(y_{t+1}y_t|y_{t-1}) g(y_{t-1}) dy_{t-1} \\ \Delta_{y_{t+1}|\omega_t} g &\equiv f_{y_{t+1}|\omega_t}(y_{t+1}|\cdot) g(\cdot) \end{aligned}$$

where $L_{y_{t+1};y_t|y_{t-1}}$ maps functions of y_{t-1} to functions of y_t, and $\Delta_{y_{t+1}|\omega_t}$ maps functions of ω to functions of ω. Now, we can rewrite $L_{y_{t+1};y_t|y_{t-1}}$ in terms of $\Delta_{y_{t+1}|\omega_t}$ and the operators from Assumption 4:

$$\begin{aligned} \left[L_{y_{t+1};y_t|y_{t-1}} g\right](y_t) &= \int f_{y_{t+1}y_t|y_{t-1}}(y_{t+1}; y_t|y_{t-1}) g(y_{t-1}) dy_{t-1} \\ &= \int \int f_{y_{t+1};y_t,\omega_t|y_{t-1}}(y_{t+1}y_t, \omega_t|y_{t-1}) d\omega_t g(y_{t-1}) dy_{t-1} \\ &= \int \int f_{y_t|\omega_t}(y_t|\omega_t) f_{y_{t+1}|\omega_t} f_{\omega_t|y_{t-1}}(\omega_t|y_{t-1}) g(y_{t-1}) dy_{t-1} d\omega_t \\ &= \int f_{y_t|\omega_t}(y_t|\omega_t) f_{y_{t+1}|\omega_t} \int f_{\omega_t|y_{t-1}}(\omega_t|y_{t-1}) g(y_{t-1}) dy_{t-1} d\omega_t \\ &= \int f_{y_t|\omega_t}(y_t|\omega_t) f_{y_{t+1}|\omega_t} \left[L_{\omega_t|y_{t-1}}\right](\omega_t) d\omega_t \\ &= \int f_{y_t|\omega_t}(y_t|\omega_t) \left[\Delta_{y_{t+1}|\omega_t} L_{\omega_t|y_{t-1}} g\right](\omega_t) d\omega_t \\ &= \left[L_{y_t|\omega_t} \Delta_{y_{t+1}|\omega_t} L_{\omega_t|y_{t-1}} g\right](y_t) \end{aligned}$$

This gives us the equivalence of the operators:

$$L_{y_{t+1};y_t|y_{t-1}} = L_{y_t|\omega_t} \Delta_{y_{t+1};\omega_t} L_{\omega_t|y_{t-1}} \tag{39}$$

integrating over y_{t+1} on both sides of this equation gives

$$L_{y_t|y_{t-1}} = L_{y_t|\omega_t} L_{\omega_t|y_{t-1}}$$

Next, we can invert $L_{y_t|\omega_t}$ to give us

$$L_{\omega_t|y_{t-1}} = L^{-1}_{y_t|\omega_t} L_{y_t|y_{t-1}}$$

The existence of this inverse is assumed directly in Assumption 4. Substituting this into equation 39,

$$L_{y_{t+1};y_t|y_{t-1}} = L_{y_t|\omega_t} \Delta_{y_{t+1};\omega_t} L^{-1}_{y_t|\omega_t} L_{y_t|y_{t-1}}$$

Finally, applying $L^{-1}_{y_t|y_{t-1}}$ from the right on both sides of the equation yields

$$L_{y_{t+1};y_t|y_{t-1}} L^{-1}_{y_t|y_{t-1}} = L_{y_t|\omega_t} \Delta_{y_{t+1};\omega_t} L^{-1}_{y_t|\omega_t}$$

This inverse operator exists by Assumption 4, as shown in Lemma 1 in HS. As is standard in the proofs in this literature, I note that the right hand side of this equation is in the form of an eigenvalue-eigenfunction decomposition with eigenvalues corresponding to $f_{y_{t+1}|\omega_t}$ and eigenfunctions corresponding to $f_{y_t|\omega_t}$. Following the argument made in HS, Assumptions 1 and 3 ensure that this decomposition is unique, which concludes the outline of the proof. See HS for a detailed discussion of the spectral decomposition argument. □

A.1 Identifying Endogeneity

In section 3.2 I offer some intuition for how observing a future period of data can aid the identification of the production function. In particular, I argue that knowledge of the Markov process $g(\omega)$ permits inverting the derivative of y_{t+1} with respect to x_t to determine the correlation between ω_t and x_t. Knowledge of this correlation is clearly essential for identifying the production function (to un-do the omitted variable bias induced by the unobservable ω). Here I demonstrate this argument more formally for a simplified class of distributions of productivity. Suppose the researcher knew that ω was distributed (conditional on x) $Uniform[0, A - x]$ for an unknown scalar $A > 0$. Clearly this class of distributions permits endogeneity of x, as it implies that ω will be increasing (on average) in x. Recall equation 19:

$$\rightarrow \frac{\partial \mathbb{E}[y_{t+1}|x_{t+1} = x_t = 0]}{\partial x_t} = \int g(\omega_t) \frac{\partial}{\partial x_t} f_{\omega_t|x_{t+1}=x_t=0} d\omega_t$$

With the assumed class of conditional productivity distributions, we can substitute for $f_{\omega_t|x_{t+1}=x_t=0}$

$$\rightarrow \frac{\partial \mathbb{E}[y_{t+1}|x_{t+1}=x_t=0]}{\partial x_t} = \int g(\omega_t)\frac{\partial}{\partial x_t}\frac{1}{A-x_t}\Big|_{x_t=0} d\omega_t \tag{40}$$

$$= \frac{-1}{A^2}\int g(\omega_t)d\omega_t \tag{41}$$

Because $g(\cdot)$ is known (from the argument made in the text at $x = 0$) and the left side of the equation is observable, this implies that A (and the conditional distribution of ω_t at every x_t) is known. Now recall equation 22:

$$\rightarrow \frac{\partial \mathbb{E}[y_t|x_{t+1}=x_t=0]}{\partial x_t} = \frac{\partial f_t(x_t)}{\partial x_t}\Big|_{x_t=0} + \int \omega_t \frac{\partial}{\partial x_t} f_{\omega_t|x_{t+1}=x_t=0} d\omega_t$$

Given knowledge of $f_{\omega_t|x_t}$, the second term on the right side of the equation is now known, meaning that the difference between the (observable) left hand side and this known term identifies the derivative of f_t at $x_t = 0$. If f_t is Cobb-Douglas, this identifies the *entire* production function. When f_t is analytic, identifying all higher order derivatives by repeatedly differentiating equation 22 also identifies the entire production function. Though this is a trivially simple example, it demonstrates the importance of being able to identify the Markov process at $Z_t = 0$ and of the assumptions concerning the invertibility of integral operators (which are much more general than the class of uniform distributions here).

B Estimation Details

B.1 Control Function

In producing production function estimates via the control function approach, I estimate industry-specific, time-invariant, translog production functions in two steps. First, I estimate the control function as a quadratic function in k and v (little changes upon the addition of year fixed effects). After subtracting estimated ex-post shocks from output, I estimate the production function via two-step GMM. In each step, I begin the Nelder-Mead search from more than 100 different starting positions and choose among the resulting estimates those which produce the smallest objective value subject to the constraint that the linear terms are between zero and 1. The instruments I use are k_{t-1} and v_{t-1}, the interaction between the two, and a quadratic term for each.

B.2 Panel

For the panel GMM estimator I again use two-step efficient GMM to estimate an industry-specific, time-invariant, translog production function. I begin the optimization at many initial points and use the same criteria to choose the best estimates as I

do in the control function approach. I find that the some industries are well-identified with a just-identified system, while others require additional instruments. To avoid making somewhat arbitrary instrument inclusion decisions industry-by-industry, I begin by estimating the production function for each industry with instruments k_t, v_t, their interaction, and quadratic terms in each, as well as linear terms in k_{t-1} and v_{t-1}. I use the residual $\omega_{t-1} + \eta_{t-1}$ implied by the production function parameters (at each objective function evaluation) as instruments for the Markov process, as described in the text. Note that in the panel estimator I use three periods of data to identify the production function at t: t, $t-1$, and $t+1$. Thus, k_t and v_t are still lags of inputs relative to the productivity innovation used in estimation (ξ_{t+1}). Next, I test for over identification in each industry. For any industry for which I can reject the null, I re-estimate the production function, this time dropping k_{t-1} and v_{t-1} (thus making the system just-identified).

C Validity of GMM Approach

C.1 Quadratic Case

In this section I show explicitly that the moments I construct in the quadratic case are valid. I assume that the first-order Markov process for productivity is quadratic, i.e. that

$$\omega_{jt} = \rho_1 \omega_{jt-1} + \rho_2 \omega_{jt-1}^2 + \xi_{jt} \tag{42}$$

for each firm j and year t. Under this assumption, my (just-identified) estimation procedure is as follows:

1. Guess β
2. Construct $\tilde{y}_{jt} = y_{jt} - f(k_{jt}, v_{jt}, \beta)$ for all t
3. Construct residual $\tilde{\xi}_{jt} = \tilde{y}_{jt} - \rho_1 \tilde{y}_{jt-1} - \rho_2 y_{jt-1}^2$ for all t
4. Center $\tilde{\xi}_{jt+1} = \tilde{\xi}_{jt+1} - \mathbb{E}[\tilde{\xi}_{jt+1}]$
5. Interact $\tilde{\xi}_{jt+1}$ with $k_{jt}, v_{jt-1}, \tilde{y}_{jt-1}$, and $\tilde{y}_{jt-1}^2$ to construct four moments
6. Loop over 1-4 to minimize moments

Note that in this model,

$$\begin{aligned} \tilde{\xi}_{jt+1} &= \xi_{jt+1} + \eta_{jt+1} - \rho_1 \eta_{jt} - 2\rho_2 \omega_{jt} \eta_{jt} - \rho_2 \eta_{jt}^2 \\ \mathbb{E}[\tilde{\xi}_{jt+1}] &= -\rho_2 \mathbb{E}[\eta_{jt}^2] \end{aligned}$$

By the full independence of η_{jt} with inputs, this means that $\tilde{\xi}_{jt+1}$ is mean-independent of my instruments. I now show this more explicitly. Note that at the true $\beta = \beta^0$

$$\begin{aligned}
\mathbb{E}[\tilde{\xi}_{jt+1}\tilde{y}_{j-1}] &= \mathbb{E}[\tilde{\xi}_{jt+1}(\omega_{jt-1} + \eta_{jt-1})] \\
&= \mathbb{E}[\tilde{\xi}_{jt+1}\omega_{jt-1}] \\
&= \mathbb{E}[(\xi_{jt+1} + \eta_{jt+1} - \rho_1\eta_t)\omega_{jt-1}] - \mathbb{E}[2\rho_2\omega_{jt}\eta_{jt}\omega_{jt-1}] - \mathbb{E}[\rho_2\eta_{jt}^2\omega_{jt-1}] \qquad (43)
\end{aligned}$$

where the second line comes from the fact that η_{jt-1} is mean independent of ξ_{jt+1}, and the third comes from substitution of the definition of $\tilde{\xi}_{jt+1}$. Taking each term of equation 43 separately, note first that the first term is zero because each term in the parenthesis are mean zero and mean independent of ω_{jt-1}. This follows because, for any ϵ meeting both of these criteria,

$$\begin{aligned}
\mathbb{E}[\epsilon\omega_{jt-1}] &= \mathbb{E}[\mathbb{E}[\epsilon\omega_{jt-1}|\omega_{jt-1}]] \\
&= \mathbb{E}[\omega_{jt-1}\mathbb{E}[\epsilon|\omega_{jt-1}]] \\
&= \mathbb{E}[\omega_{jt-1}]\mathbb{E}[\epsilon] \\
&= 0
\end{aligned}$$

by the Tower property. Note that iterating expectations also eliminates the second term of equation 43:

$$\begin{aligned}
\mathbb{E}[\rho_2\omega_{jt}\eta_{jt}\omega_{jt-1}] &= \mathbb{E}[\rho_2\mathbb{E}[\omega_{jt}\eta_{jt}\omega_{jt-1}|\omega_{jt-1}, \omega_{jt}]] \\
&= \rho_2\mathbb{E}[\omega_{jt}\omega_{jt-1}\mathbb{E}[\eta_{jt}|\omega_{jt}, \omega_{jt-1}]] \\
&= \rho_2\mathbb{E}[\omega_{jt}\omega_{jt-1}]\mathbb{E}[\eta_{jt}|\omega_{jt}\omega_{jt-1}] \\
&= 0
\end{aligned}$$

Finally, consider the third term of equation 43. Applying the law of iterated expectations once more gives:

$$\mathbb{E}[\rho_2\eta_{jt}^2\omega_{jt-1}] = \rho_2\mathbb{E}[\eta_{jt}^2]\mathbb{E}[\omega_{jt-1}]$$

These three terms, combined with the unconditional mean $\mathbb{E}[\tilde{\xi}_{jt+1}]$ shown above, imply that

$$\mathbb{E}[\tilde{\xi}_{jt+1}\tilde{y}_{jt-1}] = -\mathbb{E}[\eta_{jt}^2]\mathbb{E}[\omega_{jt-1}] = \mathbb{E}[\tilde{\xi}_{jt+1}]\mathbb{E}[\tilde{y}_{jt-1}] \qquad (44)$$

This implies that $cov(\tilde{\xi}_{jt+1}, \tilde{y}_{jt-1}) = 0$. Thus, after demeaning $\tilde{\xi}_{jt+1}$, $\mathbb{E}[\tilde{\xi}_{jt+1}\tilde{y}_{jt-1}] = 0$. Nearly identical work demonstrates that k_{jt} and v_{jt-1} are valid instruments. Longer, but identical in spirit, work shows the same for $\tilde{y}_{jt-1}^2$.

C.2 Cubic Case

In this section I discuss the modification necessary for the estimation of a model with a cubic first-order Markov process, i.e. when productivity evolves according to

$$\omega_{jt} = \rho_1\omega_{jt-1} + \rho_2\omega_{jt-1}^2 + \rho_3\omega_{jt-1}^3 + \xi_{jt} \tag{45}$$

In this case, the feasible residual $\tilde{\xi}_{jt+1}$ is now

$$\tilde{\xi}_{jt+1} = \xi_{jt+1} + \eta_{jt+1} - \rho_1\eta_{jt} - \rho_2\eta_{jt}^2 - \rho_3\eta_{jt}^3 - 2\rho_2\omega_{jt}\eta_{jt} - 3\rho_3\omega_{jt}^2\eta_{jt} - 3\rho_3\omega_{jt}\eta_{jt}^2 \tag{46}$$

Note that the mean of this residual is now

$$\mathbb{E}[\tilde{\xi}_{j+1}] = -\rho_2\mathbb{E}[\eta_{jt}^2] - \rho_3\mathbb{E}[\eta_{jt}^3] - 3\rho_3\mathbb{E}[\eta_{jt}^2]\mathbb{E}[\omega_{jt}] \tag{47}$$

which clearly implies that $\tilde{\xi}_{jt}$ is no longer mean independent of $\tilde{y}_{jt-1}$ (because $\tilde{y}_{jt-1}$ is correlated with ω_{jt}).

$$\mathbb{E}[\tilde{\xi}_{jt+1}|\tilde{y}_{jt-1}] = -\rho_2\mathbb{E}[\eta_{jt}^2|\tilde{y}_{jt-1}] - \rho_3\mathbb{E}[\eta_{jt}^3|\tilde{y}_{jt-1}] - 3\rho_3\mathbb{E}[\eta_{jt}^2|\tilde{y}_{jt-1}]\mathbb{E}[\omega_{jt}|\tilde{y}_{jt-1}]$$

What is required in this case is an additional adjustment to $\tilde{\xi}_{jt+1}$, and the estimation of an additional moment. Note that the mean-independence of $\tilde{\xi}_{jt+1}$ is violated by the third term. Thus, the appropriate adjustment to $\tilde{\xi}_{jt+1}$ is one which eliminates the influence of this term on the objective function. One option to do this is to estimate that term. Let us redefine

$$\tilde{\xi}_{jt+1} \equiv \omega_{jt+1} - \sum_{p=1}^{3}\rho_p\omega_{jt}^p + 3\rho_3\tilde{y}_{jt}\sigma_{\eta_t} \tag{48}$$

where σ_{η_t} denotes the variance of measurement error in period t. Because $\sigma_{\eta_t} \equiv \mathbb{E}[\eta_{jt}^2]$ and $\mathbb{E}[\tilde{y}_{jt}] = \mathbb{E}[\omega_{jt}]$,

$$\mathbb{E}[\omega_{jt}\eta_{jt} - \tilde{y}_{jt}\sigma_{\eta_t}] = 0$$

This also holds when $\tilde{\xi}_{jt+1}$ is interacted with instruments. Now, clearly, this approach relies on researcher knowledge, or joint estimation, of σ_{η_t}. For the latter approach, k_{jt+1} and/or k_{jt+1}^2 can be used as instruments for σ_η. These are valid instruments because (i) k_{jt+1} is independent of ξ_{jt+1} and all measurement errors by assumption (because it is determined in the previous period) and (ii) at any $\hat{\sigma}_{\eta_t} \neq \sigma_{\eta_t}$, some variation in ω_{jt} will persist in $\tilde{\xi}_{jt+1}$, meaning $\tilde{\xi}_{jt+1}$ and k_{jt+1} will be correlated (through the correlation between investment and productivity).

D Some Thoughts on Markups in Monopsonies

In this appendix I consider the estimation of the distribution of markups (i) without assuming the scalar unobservable assumption, as in the main text and (ii) under the

additional presence of monopsony power in setting prices. Substituting the standard corrected markup into the first order condition and taking logs of both sides yields

$$\log(\theta^m_{jt}) - log(s^m_{jt}) = log(\mu_{jt}) + \eta_{jt} \tag{49}$$

Note now that the left side is observed, but both terms to the right of the equality are unobserved. Further, η_{jt} and μ_{jt} vary at the same (firm-year) level, meaning there is no way to separate them without further assumptions. Equation 49 is only valid if firms have no monopsony power in the market for m. In many markets, this is an unrealistic assumption. Instead suppose that firms have some monopsony power in *all* input markets, meaning the first order condition is now

$$\mu_{jt} * \phi_{jt} = \frac{\theta^m_{jt}}{s^m_{jt} exp(\eta_{jt})} \tag{50}$$

where ϕ_{jt} is a function of the (input) supply function faced by the firm and represents a wedge relative to the competitive input market case. Suppose that the researcher observed multiple variable inputs, which is often the case in practice. The standard approach has been to choose, loosely speaking, the "most variable" input and use the markups implied by the output elasticity θ for that input. One alternative approach is to think of ϕ_{jt} as an error in our estimates of the markup distribution and apply Kotlasrki's Theorem to remove it. Denoting the two flexible inputs as m and x, 50 (in logs) becomes

$$log\left(\frac{\theta^m_{jt}}{s^m_{jt}}\right) = log(\mu_{jt}) + \eta_{jt} + log(\phi^m_{jt})$$
$$log\left(\frac{\theta^x_{jt}}{s^x_{jt}}\right) = log(\mu_{jt}) + \eta_{jt} + log(\phi^x_{jt})$$

Now, we have two equations which both contain $log(\mu_{jt}) + \eta_{jt}$ plus an additive error. In order to estimate the distributions of ϕ^m_{jt} and ϕ^x_{jt}, I propose following a recent paper by Kato, Sasaki and Ura (2018), who demonstrate under general assumptions how to estimate and conduct inference on the distribution of $log(\mu) + \eta$. I refer the reader to their paper for econometric assumptions, but it may be informative to discuss the assumptions we must make concerning the relationship between the various unobservables on the right hand side (Assumption 1 in Kato, Sasaki and Ura (2018)):

Assumption 7. *(i) μ, η, ϕ^m, and ϕ^x are continuously distributed with finite first moments and either $log(\phi^m)$ or $log(\phi^x)$ has mean zero.*

(ii) μ, η, ϕ^m, and ϕ^x are mutually independent.

These assumptions are clearly weaker than standard approaches, which require

that $\phi^m = \phi^x = 1$, meaning that $log(\phi^m) = log(\phi^x) = 0$ identically. The requirement that μ and the wedges ϕ^x and ϕ^m are mutually independent may seem restrictive, as (i) firms with monopsony power in one market may have it in another, and (ii) large firms may have large markups and more monopsony power, but again these assumptions are weaker than those made in most applications. The drawback of relaxing these assumptions is that the distribution which is identified is that of $\log(\mu) + \eta$, not $log(\mu)$ alone. The extent of this drawback depends on the setting. As I discuss in the main text, it is unclear what prior we should have on the extent of measurement error in commonly used datasets. In settings where η is believed to be negligible or non-existent (i.e. Ackerberg and Hahn (2015)), the identified distribution is simple $log(\mu)$. In other cases, where measurement error cannot be ignored, deriving the distribution of markups is less clear. I argue in the main text that the distribution of η alone is identified. With this distribution and that of $log(\mu) + \eta$ known, perhaps an additional deconvolution could identify f_μ alone.