IO Field Exam August 2022

This exam has three possible parts you can complete corresponding to 220A (Handel), 220B(Kawai), and the half-semester course in the Haas Ph.D. program (Backus).

Please answer any two out of three sections of the exam. Note that those of you who did not prepare for the Backus portion, either because you did not take the course or were not aware of this part, will not be penalized in any way for that. This means that exams will be graded independently of one another and not graded relatively to one another. Good Luck!

IO Field Exam: ECON 220A

This section has several questions related to the papers discussed in class. Please answer all of them in detail.

Question 1: Selection Markets (100 Points)

This will be a multi-part question asking about selection markets.

A. (10 points) In Einav Finkelstein and Cullen (2010) the authors set up a simple framework to study adverse selection in competitive insurance markets. Draw a graph related to their framework that describes a competitive market with adverse selection. Label the key objects of interest, including the deadweight loss from adverse selection.

B. (10 points) Handel, Hendel and Whinston also describes equilibria in competitive health insurance markets. Describe **in detail** the key differences in the underlying models in EFC and HWW. Illustrate the difference in an EFC style graph, similar to what you drew in the first part of this question.

C. (10 points) Describe the central tradeoff studied in HHW and what the authors find empirically regarding this tradeoff.

D. (10 points) In Handel (2013) there are two primary sources of descriptive evidence for inertia. Please describe these two sources of evidence and describe which one you think better identifies inertia.

E. (20 points) Write down the demand model from Handel (2013) in detail. Describe in detail (i) how risk preferences are identified and (ii) how inertia (modeled as a switching cost) is identified.

F. (20 points) Now, imagine that the mechanism underlying inertia is not a “switching cost” but is instead some other micro-founded model for inertia, such as rational inattention or naïve inattention (or some other foundation for inertia!). Write down a version of this alternative model that you could estimate, i.e. modify the model in part E. to have this new micro-foundation for inertia.

G. (10 points) Describe in depth how you might empirically test whether your model in F. is a better model than the switching cost model set up in Handel (2013).

H. (10 points) Describe in depth (i) if you think your model in F. would have different implications for adverse selection than the switching cost model in E. and, if so (ii) what would those implications be?

Question 2: Vertical Markets (50 Points)

A. (20 points) The introduction of the Ho and Lee paper we covered in class has a figure that summarizes the key counterfactual results in the paper. Describe that figure in detail (feel free to draw it out if you want) with a specific focus on the key comparative statics they investigate. Note: you do not need to know the exact numbers in the figure, just the broad economic tradeoffs and implications.

B. (10 points) In the Crawford and Yurukoglu paper we discussed in class the authors use moment inequalities as part of their model and estimation. Describe the reason why they use moment inequalities and write down an example showing how they can help identify parameters of interest.

C. (20 points) In the Crawford and Yurukoglu paper we discussed in class channels and distributors bargain over rates / input prices. Write down the bargaining protocol / model they use in as much detail as you can, using their notation when possible (10 points). Then, describe why having the bargaining model, as opposed to just a distributor marginal input cost that doesn't change, is important for the key comparative statics and conclusions in the paper.

Question 3: Short Questions (30 Points)

A. (10 points) What are the major innovations in the Nevo *Econometrica* paper on breakfast cereals, relative to BLP (1995)? Describe innovations in (i) demand estimation and (ii) dealing with endogeneity. What are the main results Nevo finds in his paper?

B. (10 points) What are the primary innovations made in BLP (1995) relative to Bresnahan (1987)? Describe innovations in both demand estimation and identification.

C. (10 points) The Berry, Gaynor and Scott Morton (2019) paper we discussed in week one discusses several key issues with recent studies that show a correlation between concentration and high markups. Clearly state <u>three reasons</u> why studies they discuss are problematic, including at least two reasons that do not have to do with issues of causality.

Read Sections II and III from Gandhi Navarro and Rivers (2020) to answer the following questions. I suggest you skim the questions before reading GNR.

(1) On page 2979, 3rd full paragraph, they say "Without loss of generality, we can normalize $E[\varepsilon_{jt}|\mathcal{I}] = E[\varepsilon_{jt}] = 0$". Why is this a normalization?

(2) Show how to get expression (5) from expression (4).

(3) Focus on the part of proof of Theorem 1 that starts with "Next, given the definition of $(\tilde{f}, \tilde{h})$ and noting that $d_t = d\ \forall t$, we have $\cdots$". Show how to get from the first line of the expression to the second line of the expression.

(4) Again, focus on the proof of Theorem 1. In the last part of the proof, why does the conditioning on Γ_{jt} appear only in the term for $\tilde{f}(k_{jt}, l_{jt}, m_{jt})$ and drop from all other terms?

(5) Gandhi Navarro and Rivers (2020) say that if researchers had access to firm-varying flexible input and output prices, this would help in estimating the production function using the orthogonality condition for $\eta + \varepsilon$. Discuss why these data would be useful and how one could incorporate these data in the estimation.

show that the gross output production function and productivity can be nonparametrically identified.

This identification strategy—regressing revenue shares on inputs to identify the flexible input elasticity, solving the partial differential equation, and integrating this into the dynamic panel/proxy variable structure to identify the remainder of the production function—gives rise to a natural two-step nonparametric sieve estimator in which different components of the production function are estimated via polynomial series in each stage. We present a computationally straightforward implementation of this estimator. Furthermore, as the numerical equivalence result in Hahn, Liao, and Ridder (2018) shows, our estimator has the additional advantage that inference on functionals of interest can be performed using standard two-step parametric results. This gives us a straightforward approach to inference.

We validate the performance of our empirical strategy on simulated data generated under three different production functions (Cobb-Douglas, constant elasticity of substitution [CES], and translog). We find that our nonparametric estimator performs quite well in all cases. We also show that our procedure is robust to misspecification arising from the presence of adjustment costs in the flexible input. We then apply our estimator, as well as several extensions of it, to plant-level data from Colombia and Chile. We show that our estimates correct for transmission bias present in ordinary least squares (OLS). Consistent with the presence of transmission bias, OLS overestimates the flexible intermediate-input elasticities and underestimates the elasticities of capital and labor. OLS estimates also tend to understate the degree of productivity heterogeneity compared with our estimates. Finally, we show that our estimates are robust to allowing for fixed effects, alternative flexible inputs, or some additional unobservables in the flexible-input demand.

The rest of the paper is organized as follows. In section II, we describe the model and set up the firm's problem. In section III, we examine the extent to which the proxy variable/dynamic panel methods can be applied to identify the gross output production function. Section IV presents our nonparametric identification strategy. In section V, we describe our estimation strategy. Section VI compares our approach with the related literature. In section VII, we present estimates from our procedure applied to Monte Carlo simulated data as well as plant-level data from Colombia and Chile. Section VIII concludes.

II. The Model

In this section, we describe the economic model of the firm that we study. We focus attention in the main text on the classic case of perfect competition in the intermediate-input and output markets. We discuss the case of monopolistic competition with unobserved output prices in appendix O6 (apps. O1–O8 are available online).

A. *Data and Definitions*

We observe a panel consisting of firms $j = 1, \ldots, J$ over periods $t = 1, \ldots, T$. A generic firm's output, capital, labor, and intermediate inputs are denoted by $(Y_{jt}, K_{jt}, L_{jt}, M_{jt})$, respectively, and their log values are denoted in lowercase by $(y_{jt}, k_{jt}, l_{jt}, m_{jt})$. Firms are sampled from an underlying population, and the asymptotic dimension of the data is to let the number of firms $J \to \infty$ for a fixed T; that is, the data take a short panel form. From this data, the econometrician can directly recover the joint distribution of $\{(y_{jt}, k_{jt}, l_{jt}, m_{jt})\}_{t=1}^{T}$.

Firms have access to information in period t, which we model as a set of random variables $\mathcal{I}_{jt}$.[6] The information set $\mathcal{I}_{jt}$ contains the information that the firm can use to solve its period t decision problem. Let $x_{jt} \in \{k_{jt}, l_{jt}, m_{jt}\}$ denote a generic input. If an input is such that $x_{jt} \in \mathcal{I}_{jt}$—that is, the amount of the input employed in period t is in the firm's information set for that period—then we say that the input is *predetermined* in period t. Thus, a predetermined input is a function of the information set of a prior period, $x_{jt} = \mathbb{X}(\mathcal{I}_{jt-1}) \in \mathcal{I}_{jt}$. If an input's optimal period t choices are affected by lagged values of that same input, then we say that the input is *dynamic*. If an input is neither predetermined nor dynamic, then we say that it is *flexible*. We refer to inputs that are predetermined, dynamic, or both as *nonflexible*.

B. *The Production Function and Productivity*

We assume that the relationship between output and inputs is determined by an underlying production function F and a Hicks neutral productivity shock ν_{jt}.

ASSUMPTION 1. The relationship between output and the inputs takes the form

$$\begin{aligned} Y_{jt} &= F(k_{jt}, l_{jt}, m_{jt})e^{\nu_{jt}} \Leftrightarrow \\ y_{jt} &= f(k_{jt}, l_{jt}, m_{jt}) + \nu_{jt}. \end{aligned} \tag{1}$$

The production function f is differentiable at all $(k, l, m) \in \mathbb{R}^3_{++}$ and strictly concave in m.

Following the proxy variable literature, the Hicks neutral productivity shock ν_{jt} is decomposed as $\nu_{jt} = \omega_{jt} + \varepsilon_{jt}$. The distinction between ω_{jt} and

[6] Formally, the firm's information set is the σ-algebra $\sigma(\mathcal{I}_{jt})$ spanned by these random variables $\mathcal{I}_{jt}$. For simplicity, we refer to $\mathcal{I}_{jt}$ as the information set.

ε_{jt} is that ω_{jt} is known to the firm before making its period t decisions, whereas ε_{jt} is an ex post productivity shock realized only after period t decisions are made. The stochastic behavior of both of these components is explained next.

ASSUMPTION 2. $\omega_{jt} \in \mathcal{I}_{jt}$ is known to the firm at the time of making its period t decisions, whereas $\varepsilon_{jt} \notin \mathcal{I}_{jt}$ is not. Furthermore, ω_{jt} is Markovian so that its distribution can be written as $P_\omega(\omega_{jt} \mid \mathcal{I}_{jt-1}) = P_\omega(\omega_{jt} \mid \omega_{jt-1})$. The function $h(\omega_{jt-1}) = E[\omega_{jt} \mid \omega_{jt-1}]$ is continuous. The shock ε_{jt}, on the other hand, is independent of the within-period variation in information sets, $P_\varepsilon(\varepsilon_{jt} \mid \mathcal{I}_{jt}) = P_\varepsilon(\varepsilon_{jt})$.

Given that $\omega_{jt} \in \mathcal{I}_{jt}$ but ε_{jt} is completely unanticipated on the basis of $\mathcal{I}_{jt}$, we will refer to ω_{jt} as persistent productivity, ε_{jt} as ex post productivity, and $\nu_{jt} = \omega_{jt} + \varepsilon_{jt}$ as total productivity. Observe that we can express $\omega_{jt} = h(\omega_{jt-1}) + \eta_{jt}$, where η_{jt} satisfies $E[\eta_{jt} \mid \mathcal{I}_{jt-1}] = 0$. The random variable η_{jt} can be interpreted as the (unanticipated at period $t-1$) "innovation" to the firm's persistent productivity ω_{jt} in period t.[7]

Without loss of generality, we can normalize $E[\varepsilon_{jt} \mid \mathcal{I}_{jt}] = E[\varepsilon_{jt}] = 0$, which is in units of log output. However, the expectation of the ex post shock, in units of the level of output, becomes a free parameter that we denote as $\mathcal{E} \equiv E[e^{\varepsilon_{jt}} \mid \mathcal{I}_{jt}] = E[e^{\varepsilon_{jt}}]$.[8] As opposed to the independence assumption on ε_{jt} in assumption 2, much of the previous literature assumes only mean independence $E[\varepsilon_{jt} \mid \mathcal{I}_{jt}] = 0$ explicitly (although stronger implicit assumptions are imposed, as we discuss below). This distinction would be important if more capital-intensive firms faced less volatile ex post productivity shocks due to automation, for example. In terms of our analysis, the only role that full independence plays (relative to mean independence) is allowing us to treat $\mathcal{E} \equiv E[e^{\varepsilon_{jt}}]$ as a constant, which makes the analysis more transparent.[9] If only mean independence is assumed, we would have $\mathcal{E}(\mathcal{I}_{jt}) \equiv E[e^{\varepsilon_{jt}} \mid \mathcal{I}_{jt}]$. We discuss the implications of this distinction below in our discussion of assumption 4 for proxy variable methods and after theorem 2 for our proposed identification strategy.

Our interest is in the case in which the production function contains both flexible and nonflexible inputs. For simplicity, we mainly focus on the case of a single flexible input in the model (but see app. O6)—namely, intermediate inputs m—and treat capital k and labor l as predetermined in the model (hence, $k_{jt}, l_{jt} \in \mathcal{I}_{jt}$). We could have also generalized the

[7] It is straightforward to allow the distribution of $P_\omega(\omega_{jt} \mid \mathcal{I}_{jt-1})$ to depend on other elements of $\mathcal{I}_{jt-1}$, such as firm export or import status, R&D, etc. In these cases, ω_{jt} becomes a controlled Markov process from the firm's point of view. See Kasahara and Rodrigue (2008) and Doraszelski and Jaumandreu (2013) for examples.

[8] See Goldberger (1968) for an early discussion of the implicit reinterpretation of results that arises from ignoring $\mathcal{E}$ (i.e., setting $\mathcal{E} \equiv E[e^{\varepsilon_{jt}}] = 1$ while simultaneously setting $E[\varepsilon_{jt}] = 0$) in the context of Cobb-Douglas production functions.

[9] While independence is sufficient, we could replace this assumption with mean independence and the high-level assumption that $\mathcal{E} \equiv E[e^{\varepsilon_{jt}}]$ is a constant.

model to allow it to vary with time t (e.g., f_t, h_t). For the most part, we do not use this more general form of the model in the analysis to follow because the added notational burden distracts from the main ideas of the paper. However, we revisit this idea below when it is particularly relevant for our analysis.

C. The Firm's Problem

The proxy variable literature of Levinsohn and Petrin (2003), Wooldridge (2009), and Ackerberg, Caves, and Frazer (2015) uses a flexible input demand—intermediate inputs—to proxy for the unobserved persistent productivity ω.[10] To do so, they assume that the demand for intermediate inputs can be written as a function of a single unobservable (ω), the so-called *scalar unobservability* assumption,[11] and that the input demand is *strictly monotone* in ω (see, e.g., assumptions 4 and 5 in Ackerberg, Caves, and Frazer 2015). We formalize this in the following assumption.

Assumption 3. The scalar unobservability and strict monotonicity assumptions of Levinsohn and Petrin (2003), Wooldridge (2009), and Ackerberg, Caves, and Frazer (2015) place the following restriction on the flexible input demand:

$$m_{jt} = \mathbb{M}_t\left(k_{jt}, l_{jt}, \omega_{jt}\right). \tag{2}$$

The intermediate-input demand $\mathbb{M}$ is assumed to be strictly monotone in a single unobservable ω_{jt}.

We follow the same setup used by both Levinsohn and Petrin (2003) and Ackerberg, Caves, and Frazer (2015) to justify assumption 3.[12] In particular, we write down the same problem of a profit-maximizing firm under perfect competition. From this, we derive the explicit intermediate-input demand equation underlying assumption 3. The following assumption formalizes the environment in which firms operate.

Assumption 4. Firms are price takers in the output and intermediate-input markets, with ρ_t denoting the common intermediate-input price and P_t denoting the common output price facing all firms in period t. Firms maximize expected discounted profits.

Under assumptions 1, 2, and 4, the firm's profit-maximization problem with respect to intermediate inputs is

$$\max_{M_{jt}} P_t E\left[F\left(k_{jt}, l_{jt}, m_{jt}\right) e^{\omega_{jt}+\varepsilon_{jt}} \mid \mathcal{I}_{jt}\right] - \rho_t M_{jt}, \tag{3}$$

[10] See Heckman and Robb (1985) for an early exposition (and Heckman and Vytlacil 2007 for a general discussion) of the replacement function approach of using observables to perfectly proxy for unobservables.

[11] Olley and Pakes (1996) do not include intermediate inputs in their model.

[12] See app. A in Levinsohn and Petrin (2003) and p. 2429 in Ackerberg, Caves, and Frazer (2015).

which follows because M_{jt} does not have any dynamic implications and thus affects only current-period profits. The first-order condition of problem (3) is

$$P_t \frac{\partial}{\partial M_{jt}} F\left(k_{jt}, l_{jt}, m_{jt}\right) e^{\omega_{jt}} \mathcal{E} = \rho_t . \tag{4}$$

This equation can then be used to solve for the demand for intermediate inputs

$$m_{jt} = \mathbb{M}\left(k_{jt}, l_{jt}, \omega_{jt} - d_t\right) = \mathbb{M}_t\left(k_{jt}, l_{jt}, \omega_{jt}\right), \tag{5}$$

where $d_t \equiv \ln(\rho_t/P_t) - \ln \mathcal{E}$. It can also be inverted to solve for productivity, ω.

Equations (4) and (5) are derived under the assumption that ε_{jt} is independent of the firm's information set ($P_\varepsilon(\varepsilon_{jt} \mid \mathcal{I}_{jt}) = P_\varepsilon(\varepsilon_{jt})$). If instead only mean independence of ε_{jt} were assumed ($E[\varepsilon_{jt} \mid \mathcal{I}_{jt}] = 0$), we would have $P_t(\partial F(k_{jt}, l_{jt}, m_{jt})/\partial M_{jt}) e^{\omega_{jt}} \mathcal{E}(\mathcal{I}_{jt}) = \rho_t$, and hence $m_{jt} = \mathbb{M}_t(k_{jt}, l_{jt}, \omega_{jt}, \mathcal{I}_{jt})$. Assumption 3 is therefore implicitly imposing that if $\mathcal{E}(\mathcal{I}_{jt})$ is not constant, then it is at most a function of the variables already included in equation (2). In theory, this can be relaxed by allowing the proxy equation to also depend on the other elements of the firm's information set, as long as this is done in a way that does not violate scalar unobservability/monotonicity.

Given the structure of the production function, we can formally state the problem of transmission bias in the nonparametric setting. Transmission bias classically refers to the bias in Cobb-Douglas production function parameter estimates from an OLS regression of output on inputs (see Marschak and Andrews 1944; Griliches and Mairesse 1998). In the nonparametric setting, we can see transmission bias more generally as the empirical problem of regressing output y_{jt} on inputs (k_{jt}, l_{jt}, m_{jt}), which yields

$$E\left[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}\right] = f\left(k_{jt}, l_{jt}, m_{jt}\right) + E\left[\omega_{jt} \mid k_{jt}, l_{jt}, m_{jt}\right],$$

and hence the elasticity of the regression in the data with respect to an input $x_{jt} \in \{k_{jt}, l_{jt}, m_{jt}\}$,

$$\frac{\partial}{\partial x_{jt}} E\left[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}\right] = \frac{\partial}{\partial x_{jt}} f\left(k_{jt}, l_{jt}, m_{jt}\right) + \frac{\partial}{\partial x_{jt}} E\left[\omega_{jt} \mid k_{jt}, l_{jt}, m_{jt}\right],$$

is a biased estimate of the true production elasticity $(\partial f(k_{jt}, l_{jt}, m_{jt})/\partial x_{jt})$.

III. The Proxy Variable Framework and Gross Output

Both the dynamic panel literature and the proxy literature of Olley and Pakes (1996), Levinsohn and Petrin (2003), Wooldridge (2009), and

Ackerberg, Caves, and Frazer (2015) have mainly focused on estimating value-added models of production, in which intermediate inputs do not enter the estimated production function.[13] One exception is Levinsohn and Petrin (2003), which employs a gross output specification. However, previous work by Bond and Söderbom (2005) and Ackerberg, Caves, and Frazer (2015) has identified an identification problem with the Levinsohn and Petrin (2003) procedure. Therefore, in this section we examine whether the modified proxy variable approach developed by Ackerberg, Caves, and Frazer (2015) for value-added production functions can be extended to identify gross output production functions under the setup described in the previous section.[14]

Under the proxy variable structure, the inverted proxy equation, $\omega_{jt} = \mathbb{M}^{-1}(k_{jt}, l_{jt}, m_{jt}) + d_t$, is used to replace for productivity. Here transmission bias takes a very specific form:

$$\begin{aligned} E\left[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}, d_t\right] &= f\left(k_{jt}, l_{jt}, m_{jt}\right) + \mathbb{M}^{-1}\left(k_{jt}, l_{jt}, m_{jt}\right) + d_t \\ &\equiv \phi\left(k_{jt}, l_{jt}, m_{jt}\right) + d_t, \end{aligned} \tag{6}$$

where d_t represents a time fixed effect. Clearly, no structural elasticities can be identified from this regression (the "first stage")—in particular, the flexible input elasticity, $(\partial f(k_{jt}, l_{jt}, m_{jt})/\partial m_{jt})$. Instead, all the information from the first stage is summarized by the identification of the random variable $\phi(k_{jt}, l_{jt}, m_{jt})$ and, as a consequence, the ex post productivity shock $\varepsilon_{jt} = y_{jt} - E[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}, d_t]$.

The question then becomes whether the part of $\phi(k_{jt}, l_{jt}, m_{jt})$ that is due to $f(k_{jt}, l_{jt}, m_{jt})$ versus the part that is due to ω_{jt} can be separately identified using the second-stage restrictions of the model. This second stage is formed by adopting a key insight from the dynamic panel data literature (Arellano and Bond 1991; Blundell and Bond 1998, 2000)—namely, that given an assumed time-series process for the unobservables (in this case, the Markovian process for ω in assumption 2), appropriately lagged input decisions can be used as instruments. That is, we can rewrite the production function as

[13] Intermediate inputs, however, may still be used as the proxy variable for productivity (see Ackerberg, Caves, and Frazer 2015).

[14] We restrict our attention in the main text to the use of intermediate inputs as a proxy vs. the original proxy variable strategy of Olley and Pakes (1996) that uses investment. As Levinsohn and Petrin (2003) argued, the fact that investment is often zero in plant-level data leads to practical challenges in using the Olley and Pakes (1996) approach, and as a result, using intermediate inputs as a proxy has become the preferred strategy in applied work. In app. O1, we show that our results extend to the case of using investment instead, as well as to the use of dynamic panel methods.

$$\begin{aligned} y_{jt} &= f\left(k_{jt}, l_{jt}, m_{jt}\right) + \omega_{jt} + \varepsilon_{jt} \\ &= f\left(k_{jt}, l_{jt}, m_{jt}\right) + h\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d_{t-1} - f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \quad (7) \\ &\quad + \eta_{jt} + \varepsilon_{jt} \end{aligned}$$

to form the second-stage equation. Assumption 2 implies that for any transformation $\Gamma_{jt} = \Gamma(\mathcal{I}_{jt-1})$ of the lagged-period information set $\mathcal{I}_{jt-1}$ we have the orthogonality $E[\eta_{jt} + \varepsilon_{jt} \mid \Gamma_{jt}] = 0$.[15] We focus on transformations that are observable by the econometrician, in which case Γ_{jt} will serve as the instrumental variables for the problem.[16]

One challenge in using equation (7) for identification is the presence of an endogenous variable m_{jt} in the model that is correlated with η_{jt}. However, all lagged output/input values, as well as the current values of the predetermined inputs k_{jt} and l_{jt}, are transformations of $\mathcal{I}_{jt-1}$.[17] Therefore, the full vector of potential instrumental variables given the data described in section II.A is given by $\Gamma_{jt} = (k_{jt}, l_{jt}, d_{t-1}, y_{jt-1}, k_{jt-1}, l_{jt-1}, m_{jt-1}, \ldots, d_1, y_{j1}, k_{j1}, l_{j1}, m_{j1})$.[18]

A. *Identification*

Despite the apparent abundance of available instruments for the flexible input m_{jt}, notice that by replacing for ω_{jt} in the intermediate-input demand equation (5), we obtain

$$m_{jt} = \mathbb{M}\left(k_{jt}, l_{jt}, h\left(\mathbb{M}^{-1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d_{t-1}\right) + \eta_{jt} - d_t\right). \quad (8)$$

This implies that the only sources of variation left in m_{jt} after conditioning on $(k_{jt}, l_{jt}, d_{t-1}, k_{jt-1}, l_{jt-1}, m_{jt-1}) \in \Gamma_{jt}$ (which are used as instruments for themselves) are the unobservable η_{jt} itself and d_t. Therefore, for all of the remaining elements in Γ_{jt}, their only power as instruments is via their dependence on d_t.

[15] Notice that since ε_{jt} is recoverable from the first stage, one could instead use the orthogonality $E[\eta_{jt} \mid \Gamma_{jt}] = 0$. However, this can be formed only for observations in which the proxy variable—intermediate-input demand (or investment in Olley and Pakes 1996)—is strictly positive. Observations that violate the strict monotonicity of the proxy equation need to be dropped from the first stage, which implies that ε_{jt} cannot be recovered. This introduces a selection bias since $E[\eta_{jt} \mid \Gamma_{jt}, \iota_{jt} > 0] \neq E[\eta_{jt} \mid \Gamma_{jt}]$, where ι_{jt} is the proxy variable. The reason is that firms that receive lower draws of η_{jt} are more likely to choose nonpositive values of the proxy, and this probability is a function of the other state variables of the firm.

[16] The idea that one can use expectations conditional on lagged information sets to exploit the property that the innovation should be uncorrelated with lagged variables goes back to at least the work on rational expectations models; see, e.g., Sargent (1978) and Hansen and Sargent (1980).

[17] If k_{jt} and/or l_{jt} are dynamic but not predetermined, then only lagged values enter Γ_{jt}.

[18] Following Doraszelski and Jaumandreu (2013), we exclude d_t from the instruments, as current prices and the innovation to productivity are determined contemporaneously and hence may be correlated (see also Ackerberg et al. 2007).

Identification of the production function f by instrumental variables is based on projecting output y_{jt} onto the exogenous variables Γ_{jt} (see, e.g., Newey and Powell 2003). This generates a restriction between (f, h) and the distribution of the data that takes the form

$$\begin{aligned} E\left[y_{jt} \mid \Gamma_{jt}\right] &= E\left[f\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}\right] + E\left[\omega_{jt} \mid \Gamma_{jt}\right] \\ &= E\left[f\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}\right] + h\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right. \\ &\quad \left. + d_{t-1} - f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right). \end{aligned} \tag{9}$$

The unknown functions underlying equation (9) are given by (f, h), since $\phi(k_{jt-1}, l_{jt-1}, m_{jt-1}) + d_{t-1}$ is known from the first-stage equation (6). The true (f^0, h^0) are identified if no other $(\tilde{f}, \tilde{h})$ among all possible alternatives also satisfy the functional restriction (9) given the distribution of the observables.[19]

In theorem 1, we first show that in the absence of time-series variation in prices, $d_t = d \; \forall \; t$, the proxy variable structure does not suffice to identify the gross output production function.[20] Specifically, we show that the application of instrumental variables (via the orthogonality restriction $E[\eta_{jt} + \varepsilon_{jt} \mid \Gamma_{jt}] = 0$) to equation (7) is insufficient for identifying the production function f (and the Markovian process h). Intuitively, if d_t does not vary over time in equation (8), then the only remaining source of variation in m_{jt} is the innovation η_{jt}, which is by construction orthogonal to the remaining elements of Γ_{jt}.

Theorem 1. In the absence of time-series variation in relative prices, $d_t = d \; \forall \; t$, under the model defined by assumptions 1–4, there exists a continuum of alternative $(\tilde{f}, \tilde{h})$ defined by

$$\begin{aligned} \tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) &\equiv (1 - a)f^0\left(k_{jt}, l_{jt}, m_{jt}\right) + a\phi\left(k_{jt}, l_{jt}, m_{jt}\right), \\ \tilde{h}(x) &\equiv ad + (1 - a)h^0\left(\frac{1}{(1 - a)}(x - ad)\right) \end{aligned}$$

for any $a \in (0, 1)$, that satisfies the same functional restriction (9) as the true (f^0, h^0).

Proof. We begin by noting that from the definition of ϕ, it follows that $E[y_{jt} \mid \Gamma_{jt}] = E[\phi(k_{jt}, l_{jt}, m_{jt}) + d_t \mid \Gamma_{jt}]$. Hence, for any (f, h) that satisfy (9), it must be the case that

[19] Some researchers may not be interested in recovering h. In our results below, regardless of whether h is identified, the production function f is not (except in the degenerate case in which there are no differences in ω across firms, so $\phi(k_{jt}, l_{jt}, m_{jt}) = f(k_{jt}, l_{jt}, m_{jt})$).

[20] In app. O1, we show that a similar result holds for the case of investment as the proxy variable and for the use of dynamic panel techniques under this same structure.

$$\begin{aligned} &E\left[\phi\left(k_{jt}, l_{jt}, m_{jt}\right) + d_t - f\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}\right] \\ &\quad = h\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d_{t-1} - f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right). \end{aligned} \tag{10}$$

Next, given the definition of $(\tilde{f}, \tilde{h})$ and noting that $d_t = d \ \forall \ t$, we have

$$\begin{aligned} &\tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) + \tilde{h}\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - \tilde{f}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \\ &\quad = f^0\left(k_{jt}, l_{jt}, m_{jt}\right) + a\left(\phi\left(k_{jt}, l_{jt}, m_{jt}\right) - f^0\left(k_{jt}, l_{jt}, m_{jt}\right)\right) + ad \\ &\qquad + (1-a)h^0\left(\frac{(1-a)\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right)}{1-a}\right) \\ &\quad = f^0\left(k_{jt}, l_{jt}, m_{jt}\right) + a\left(\phi\left(k_{jt}, l_{jt}, m_{jt}\right) + d - f^0\left(k_{jt}, l_{jt}, m_{jt}\right)\right. \\ &\qquad + (1-a)h^0\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right). \end{aligned}$$

Now, take the conditional expectation of the above (with respect to Γ_{jt}):

$$\begin{aligned} &E\left[\tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}\right] + \tilde{h}\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - \tilde{f}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \\ &\quad = E\left[f^0\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}\right] + ah^0\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right. \\ &\qquad \left. + d - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \\ &\qquad + (1-a)h^0\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \\ &\quad = E\left[f^0\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}\right] + h^0\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right. \\ &\qquad \left. + d - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right), \end{aligned}$$

where the first equality uses the relation in equation (10). Thus, (f^0, h^0) and $(\tilde{f}, \tilde{h})$ satisfy the functional restriction (9) and cannot be distinguished via instrumental variables. QED

We now provide two corollaries to our main theorem to describe the extent to which time-series variation (via d_t) can be used to identify the model. (In app. O2, we provide an illustration of these results in the context of the commonly employed Cobb-Douglas parametric form.)

In corollary 1, we show that if $T = 2$ (the minimum number of periods required by these procedures), the model cannot be identified, even if d_t varies. Intuitively, since the second stage already conditions on d_1, the only remaining potential source of variation is in d_2, which of course does not vary.

Corollary 1. For $T = 2$, under the model defined by assumptions 1–4, there exists a continuum of alternative $(\tilde{f}, \tilde{h})$ defined by

$$\begin{aligned} \tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) &\equiv (1-a)f^0\left(k_{jt}, l_{jt}, m_{jt}\right) + a\phi\left(k_{jt}, l_{jt}, m_{jt}\right), \\ \tilde{h}(x) &\equiv ad_2 + (1-a)h^0\left(\frac{1}{(1-a)}(x - ad_1)\right) \end{aligned}$$

for $t = 1, 2$ and for any $a \in (0, 1)$ that satisfies the same functional restriction (9) as the true (f^0, h^0).

Proof. The proof follows from the same steps in the proof of theorem 1. QED

In corollary 2, we show that when one relaxes the assumption of time homogeneity in either the production function or the Markov process for productivity, the model similarly cannot be identified, even with $T > 2$. Intuitively, once the model varies with time, time-series variation is no longer helpful.

COROLLARY 2. Under the model defined by assumptions 1–4,

i) if the production function is time varying, f_t^0, there exists a continuum of alternative $(\tilde{f}_t, \tilde{h})$ defined by[21]

$$\tilde{f}_t(k_{jt}, l_{jt}, m_{jt}) \equiv (1-a)f_t^0(k_{jt}, l_{jt}, m_{jt}) + a\phi_t(k_{jt}, l_{jt}, m_{jt}) + ad_t,$$
$$\tilde{h}(x) \equiv (1-a)h^0\left(\frac{1}{(1-a)}x\right),$$

or

ii) if the process for productivity is time varying, h_t^0, there exists a continuum of alternative $(\tilde{f}, \tilde{h}_t)$ defined by

$$\tilde{f}(k_{jt}, l_{jt}, m_{jt}) \equiv (1-a)f^0(k_{jt}, l_{jt}, m_{jt}) + a\phi(k_{jt}, l_{jt}, m_{jt}),$$
$$\tilde{h}_t(x) \equiv ad_t + (1-a)h_t^0\left(\frac{1}{(1-a)}(x - ad_{t-1})\right),$$

such that for any $a \in (0, 1)$, these alternative functions satisfy the functional restriction (9).

Proof. The proof follows from the same steps in the proof of theorem 1. QED

The result in theorem 1 and its two corollaries is a useful benchmark, as it directly relates to the econometric approach used in the proxy variable literature. However, this instrumental variables approach does not necessarily exhaust the sources of identification inherent in the proxy variable structure. First, since the instrumental variables approach is based only on conditional expectations, it does not employ the entire distribution of the data $(y_{jt}, m_{jt}, \Gamma_{jt})$. Second, it does not directly account for the fact that assumption 3 also imposes restrictions (scalar unobservability and monotonicity) on the determination of the endogenous variable m_{jt} via $\mathbb{M}(\cdot)$. Therefore, the proxy variable structure imposes restrictions on a simultaneous system of equations because, in addition to the

[21] Notice that when the production function is allowed to be time varying, the first-stage estimates also need to be time varying (i.e., $E[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}] = \phi_t(k_{jt}, l_{jt}, m_{jt}) + d_t$).

model for output, y_{jt}, there is a model for the proxy variable—in this case, intermediate inputs, m_{jt}. In appendix O3, we extend our result to the full model involving f, h, and $\mathbb{M}$, using the full distribution of the data.

B. Monte Carlo Evidence on the Use of Time-Series Variation

The result in theorem 1 shows that under the model described above, there are not enough sources of cross-sectional variation that can be used to identify the gross output production function. In particular, the problem is associated with flexible intermediate inputs. While aggregate time-series variation provides a potential source of identification, relying on it runs a risk of weak identification in practice.

To evaluate the performance of using time-series variation as a source of identification, we conduct several Monte Carlo experiments. As we show in equation (5), the firm's optimal choice of intermediate inputs depends on the relative price of intermediate inputs to output, as opposed to the levels. In our simulations, we fix the price of output to be one and let the price of intermediate inputs vary. Specifically, the (log) price of intermediate inputs is assumed to follow an AR(1) (first-order autoregressive) process. We refer to the variance of the innovation in this process as the level of time-series variation.

The parameters of the data-generating process are chosen to roughly match the properties of our data, as well as the variances of our productivity estimates, described in section VII. A full description of the setup is provided in appendix O4 (Monte Carlo 1). The key features are as follows. For simplicity, we abstract away from labor and specify a Cobb-Douglas production function in capital and intermediate inputs, with elasticities of 0.25 and 0.65, respectively. Firms maximize the expected stream of future discounted profits. Productivity is assumed to evolve according to an AR(1) process with a persistence parameter of 0.8. The law of motion for capital is given by $K_{jt} = (1 - \kappa_j)K_{jt-1} + I_{jt-1}$, where investment I is chosen a period ahead in $t - 1$ and the depreciation rate $\kappa_j \in [0.05, 0.15]$ varies across firms. Intermediate inputs are chosen flexibly in period t as a function of capital, productivity, and the relative price of intermediates to output. The price of investment is assumed to be fixed. For the time-series process for the price of intermediate inputs, we set the AR(1) coefficient to 0.6 and the variance of the innovation at a baseline value of 0.0001.[22] In addition to the baseline value of time-series variation, we also create versions with half, twice, and 10 times this baseline variation (0.00005, 0.0002, and 0.001, respectively).

[22] This corresponds to the values obtained from a regression of the log relative price of intermediate inputs on its lag for the largest industry in Chile: food products (International Standard Industrial Classification [ISIC] code 311). The level of time-series variation in Colombia is considerably smaller.

Industrial Organization (Backus)
PHDBA 297T-LEC-002

Please note: all of these questions can be answered in just a few sentences or lines of algebra. If you get carried away writing long answers, you will risk running out of time.

Demand:

In class, we spent a lot of time talking about logit demand systems. In these demand systems, utility for consumer i and product j is given by

$$u_{ij} = X_j\beta + \xi_j - \alpha p_j + \epsilon_{ij}, \tag{1}$$

and if ϵ_{ij} is an iid logit error, then the choice probabilities are given by

$$P(u_{ij} = max_k\{u_{ik}\}) = \frac{e^{X_j\beta+\xi_j-\alpha p_j}}{1+\sum_k e^{X_k\beta+\xi_k-\alpha p_k}}. \tag{2}$$

1) Using equation (2), derive the 2SLS estimator for the parameters of this model when price is endogenous and you possess a set of instruments Z.

2) Describe one of the (several) commonly used set of instruments used to estimate this problem, and and discuss limitations your choice involves.

3) Suppose now that you have estimated such a demand model. You use this demand system to compute the consumer surplus *losses* that would be involved in removing a product from the market. How would you do so?

4) Stepping out of your model for a moment, would you trust that estimate? How would think about the limitations of such an exercise?

[NB: We are working with a simple logit here to make your computations easier. Please do not reply to 4) "you should have estimated BLP"]

Testing:

Stepping back into your model, you have estimated demand and you have a set of instruments Z. You are interested in testing two competing models of seller conduct: that they are perfectly competitive, and that they set a fixed 30% markup.

5) Could you use the testing framework in Backus, Conlon, and Sinkinson (2021) to do so? If so, offer some intuition for the test (do not write out the entire procedure). If not, why not?

Supply:

Now let's shift gears and think about estimation of production functions,

$$y_{jt} = X\beta + \omega_{jt} + \varepsilon_{jt}. \tag{3}$$

Recall that y is log output, X consists of logged factors of production, and may include static or dynamic inputs, ω_{jt} is a a productivity shock that follows an exogenous Markov process, and ε_{jt} is an iid transitory shock.

In the first stage of Olley and Pakes (1996), the econometrician runs a regression

$$y_{jt} = \beta_\ell \ell_{jt} + g(k_{jt}, i_{jt}) + \varepsilon_{jt} \tag{4}$$

Or, in Levinsohn and Petrin (2003),

$$y_{jt} = \beta_\ell \ell_{jt} + g(k_{jt}, m_{jt}) + \varepsilon_{jt}. \tag{5}$$

In these equations, ℓ stands for log labor, k stands for log capital, m stands for materials, and i stands for investment. The function $g(\cdot)$ stands in for a flexible semiparametric form (e.g., a polynomial series expansion).

6) What is the motivation for this regression? That is, how do you get from equation (3) to equations (4) or (5), and what is the point of doing so?

7) This first-stage regression has turned out to be problematic, as we saw in our discussion of Ackerberg, Caves, and Frazer (2015). Why? What additional assumption(s) would be helpful to fix the problem?