Estimating the Production Function when Firms Are Constrained

Ajay Shenoy*

October 28, 2017
First Version: 30 July 2016

Abstract

I derive a test for the key assumption behind a broad set of methods for estimating production functions: that the firm's choice of intermediate inputs depends only on its observed choices of other inputs and on unobserved productivity. This assumption fails when firms are constrained or face other market distortions, as is common in developing countries or among small firms in developed countries. The test rejects undistorted choices in many countries and industries. I show that when firms are constrained a simple autoregressive estimator becomes viable. I propose a method to choose between choice-based and autoregressive estimators, which in simulations yields lower error than either approach alone. (JEL Codes: C52, D24, C51)

*University of California, Santa Cruz; email at azshenoy@ucsc.edu. Phone: (831) 359-3389. Postal Address: Rm. E2455, University of California, M/S Economics Department, 1156 High Street, Santa Cruz CA, 95064. I am grateful to Salvador Navarro for providing code and data. I also thank Dan Ackerberg for helpful comments and suggestions. I thank Liam Rose for excellent research assistance. This paper benefited greatly from the suggestions of Natalia Lazzati, Alan Spearot, and seminar participants at U.C. Santa Cruz.

1 Introduction

Central though it is to any model of the economy, the production function is one of the hardest primatives to estimate. Most models of production predict that a more productive firm will use more inputs. But the researcher does not observe and cannot control for productivity. Any attempt to estimate the effect on output of hiring more labor or installing more capital suffers from omitted variable bias.

Starting with the work of Olley and Pakes (1996), a host of new methods have addressed this problem by exploiting the information contained in the firm's choice of inputs. Proxy methods like that of Ackerberg et al. (2015) use the choices of the firm to infer its productivity. They assume that, conditional on its labor and capital, a more productive firm always uses more intermediate inputs. These inputs are a proxy for productivity, which is no longer an omitted variable. Meanwhile first-order methods such as Gandhi et al. (2013) assume the firm chooses its level of intermediates optimally. By combining the production function with the first-order condition for intermediates, Gandhi et al. (2013) derive an estimating equation that does not contain productivity. Though powerful, these choice-based methods rely on equally powerful assumptions.

I show that these assumptions need not hold when a firm's choice of intermediates is distorted by market imperfections—for example, an expenditure constraint. Proxy methods require a one-for-one relation between productivity and the choice of intermediates. If some firms lack credit or cannot find suppliers, two otherwise identical firms may use different levels of intermediates. First-order methods require that the choice of intermediates is optimal, meaning the marginal product equals the price. If a firm is constrained its choice need not be optimal.

Constraints and other market distortions are widespread in industries across the world. According to the 2006 World Bank Enterprise Survey, 32 percent of Central American firms and 37 percent of Chilean firms find getting electricity to be a serious or very serious obstacle. Nearly 60 percent of Zimbabwean firms have wasted production capacity because inputs were unavailable. Half suffered power failures, and 80 percent lack financing. According to the Economics Research Forum's survey of firms, 25 percent of Egyptian firms and 26 percent of Tunisian firms say that getting raw materials is a severe constraint. Roughly half in both countries say the cost of raw materials is a constraint. Given that most have little access to financial services, it is hard to imagine they are able to choose the optimal level of inputs. Even in the U.S. many firms are constrained. The 2007 Survey of Business Owners reports that 10 percent of all firms that shut down did so because they lack access to credit. Even if firms do not face a hard credit constraint, they may face other market distortions that cause their choices to depend on more than just their production technology. It is critical that a researcher be able to test for whether such distortions are present.

This paper's first contribution is to show that constraints and other distortions may create identification problems for choice-based methods. Since this first result makes it crucial to know whether such distortions are serious, the second contribution is a test for the assumptions behind choice-based methods. All such methods rely on what Ackerberg et al. (2015) call the Scalar Unobservable assumption, which states that the choice of intermediates is a function of only capital, labor, and productivity. When combined with the other identifying assumptions, the Scalar Unobservable assumption implies the cost of intermediates as a share of the firm's output is a function of only the choices of capital, labor, and intermediates used in production. After controlling for a nonparametric function of these inputs no other variable known to the firm in year $t-1$ or earlier should be informative. But if the firm is constrained or its choice is otherwise distorted, lags of these inputs may be informative about the constraint, which in turn is informative about the cost share of intermediates. Testing for whether these lags are informative is in effect a test of the Scalar Unobservable assumption. I apply the test to the sample of manufacturing firms in Chile and Colombia studied by Gandhi et al. (2013). The test rejects the Scalar Unobservable assumption in most industries, suggesting the Scalar Unobservable assumption cannot be taken for granted.

The paper's third contribution is to show that a simplified dynamic panel estimator may be used when firms are constrained. This estimator assumes that productivity is autoregressive (rather than an arbitrary Markov process), but does not assume a Scalar Unobservable. To be clear, the estimator itself is not new (Ackerberg et al., 2015, see, for example,); rather, I show that this estimator sidesteps the need for choice assumptions that may be undermined by market distortions. I also show that this estimator, though estimated by non-linear GMM, has a close connection to two-stage least squares. As I describe below, this connection is useful because it is possible to test whether the estimator is well-identified using standard weak instruments statistics.

Such statistics matter because when firms are relatively unconstrained, the autoregressive estimator—at least when used to estimate a gross rather than a value-added production function—is weakly identified. In the absence of constraints or other distortions, the firm's choice of intermediates has no exogenous variation independent of its choices of the other inputs. But market distortions induce variation in its choice that can then be used for identification. The same distortions that undermine choice-based methods are crucial for the consistency of the autoregressive estimator, implying the methods are complementary.

Showing how constraints and distortions induce this complementarity is the fourth and most important contribution of this paper. I show that the bias of Gandhi et al. (2013) is strictly increasing in the severity of the constraints. Meanwhile, the Cragg-Donald statistic for the identification of the autoregressive method is bounded above by an expression that captures how informative the instruments are about the distortion. Finally, I show that the problem of weak identification (and thus complementarity) arises only when estimating a gross production function. Assuming productivity is autoregressive, is no special reason to expect the autoregressive method will be weakly identified when used to estimate a value-added production function.

Monte Carlo simulations confirm the principle of complementarity. As the severity of constraints increases the performance of the autoregressive method improves even as that of the Gandhi-Navarro-Rivers deteriorates. Measures of strong identification improve with the severity of constraints, confirming that weak identification undermines the autoregressive method unless firms are constrained.

The principle of complementarity suggests that if it is possible to choose between Gandhi-Navarro-Rivers and the autoregressive estimator, it may be possible to construct estimates that are valid regardless of whether firms are constrained. The paper's final contribution is to propose such a selection criterion. The criterion is based on 1) the test for constraints, 2) a common-sense restriction that the chosen estimator should yield positive elasticities for all inputs, and 3) a weak identification statistic for two-stage least squares. Though such a statistic depends on the unknown autocorrelation coefficient in the stochastic process for productivity, I show that the autoregressive estimate of this coefficient is consistent even when the production function is weakly identified. This estimate can be used, for example, to construct the rK statistic proposed by Kleibergen and Paap (2006). The autoregressive method should only be used when this rK statistic is large.

2 The Problem of Identification under Constraints

2.1 Review of Choice-Based Methods

Consider a firm that produces output at time t by combining capital K_t, labor L_t, and real expenditure on intermediate inputs M_t using a gross production function F. I assume output depends on real expenditures of intermediates denoted in units of the output good.[1] Firms differ in their productivity, which most choice-based methods assume is Hicks neutral.[2]

Assumption 1 (Hicks Neutrality) *The productivity of the firm is Hicks neutral and has two parts: ω_t, which is persistent, and ε_t, which is unknown when inputs are chosen and is independent across time and firms. Gross output is*

$$Y_t = e^{\omega_t + \varepsilon_t} F(K_t, L_t, M_t) \tag{1}$$

Since all the methods discussed in this paper are nonparametric it is necessary to assume the production function is smooth. It is also convenient to assume concavity:

[1] Like capital, the term "intermediate inputs" is a catch-all for many different inputs (e.g. fuel, electricity, and raw materials), meaning there is no unambiguous definition for the "level" of intermediate inputs. If the researcher is willing to somehow define a price of intermediates, it would imply a level of intermediates. The researcher might further assume that output depends on this implied level of intermediates rather than expenditures. I show in Appendix C.2 that in this case the main results of the paper require little or no modification.

[2] For example, Olley and Pakes (1996); Levinsohn and Petrin (2003); Wooldridge (2009); Gandhi et al. (2013); Ackerberg et al. (2015).

Assumption 2 (Production) *The production function F is smooth and concave in each of its arguments.*

It is common to estimate the production function using a log-sieve approximation—that is, to approximate the log of the function with a polynomial in the logs of its arguments. That approach effectively assumes the production function has a power series representation of some order.

It is also standard to make some assumption about timing.[3] It is not necessary to make any assumptions about how capital and labor are chosen, except that they are not a function of future information. But it is necessary to assume there are valid instruments for both:

Assumption 3 (Dynamic Capital and Labor) *K_t and L_t are at least partly determined at $t-1$ or earlier.*

In their empirical application, Gandhi et al. (2013) assume that both capital and labor are completely determined at $t-1$. This assumption is unnecessary for their method as long as capital and labor are at least somewhat dynamic (e.g. there are some adjustment costs), which implies that their lags are informative about their current levels. But for consistency I make the same assumption as Gandhi et al. (2013) in my simulations; the assumption actually favors their method and is thus conservative.[4]

Having chosen its capital K_t and labor L_t, the firm now chooses its expenditure on intermediate inputs M_t. Though not crucial for the autoregressive method derived in Section 4, this timing is crucial for the choice-based methods:

Assumption 4 (Timing) *M_t is chosen after the other inputs, after ω_t is known, but before ε_t is known.*

[3]As Ackerberg et al. (2015) explain, the production function is identified only if labor and capital are chosen before intermediate inputs, if there are i.i.d. shocks to the price of labor or output (but not productivity) after inputs are chosen but before labor is chosen, or if there is i.i.d. optimization error in the choice of labor. I assume the first of these as it is easiest to model and seems plausible.

[4]Under this timing assumption the choice-based methods can use current capital and labor as instruments, but the autoregressive method of Section 4 cannot. Under the weaker assumption that capital and labor are chosen with some information about ω_t the choice-based methods would have to use the same set of instruments as the autoregressive method.

The method of Gandhi et al. (2013) assumes the firm's choice is optimal, meaning the firm solves

$$\max_{M_t} \mathbb{E}_{\varepsilon_t}[e^{\omega_t+\varepsilon_t} F(K_t, L_t, M_t) - M_t] \tag{2}$$

by setting the expected marginal product of intermediate expenditure equal to the price. Here I have assumed for simplicity that the M_t is defined as expenditures on intermediates, meaning the price is normalized to 1. Then the method assumes

Assumption 5 (Optimal Choices) *Firms choose M_t to satisfy*

$$1 = \mathbb{E}[e^{\varepsilon_t}] e^{\omega_t} F_M(K_t, L_t, M_t) \tag{3}$$

Gandhi et al. (2013) exploit the assumption of Hicks Neutrality, which implies that productivity enters both the production function and the righthand side of the first-order condition multiplicatively. Dividing by realized output and multiplying by M_t gives

$$\frac{M_t}{Y_t} = \frac{F_M(K_t, L_t, M_t) M_t}{F(K_t, L_t, M_t)} \mathbb{E}[e^{\varepsilon_t}] e^{-\varepsilon_t} \tag{4}$$

The lefthand side is simply the cost share of intermediates, which is observable.[5] Gandhi et al. nonparametrically estimate this "share regression" in logs, which recovers the elasticity of output with respect to intermediate inputs M. Let $\hat{\varepsilon}_t$ be the residual, which is a consistent estimate of the shock ε_t. Divide the predicted value from the share regression by M and the sample average of $e^{\hat{\varepsilon}_t}$ to isolate $\frac{F_M(K_t,L_t,M_t)}{F(K_t,L_t,M_t)}$. Integrate this ratio with respect to M to recover (the log of) the production function F up to a constant of integration $\mathscr{C}(K_t, L_t)$. Though the integral is now known, the production function cannot be extracted from it unless $\mathscr{C}(K_t, L_t)$ is known.

Let $\mathscr{I}_t$ denote the integral and define $\mathcal{Y}_t = y_t - \mathscr{I}_t - \varepsilon_t$, where $y_t = \log Y_t$.

[5]Gandhi et al. (2013) define the share as $\frac{P^M M_t}{Y_t}$, as they assume it is possible to calculate a separate real price and real level of intermediates, and that output depends on the real level rather than the real expenditure. By contrast, Ackerberg et al. (2015) in their exposition take the approach used here. I show in Appendix C.2 that under the assumptions made by Gandhi et al. (2013), the main results require no modification.

Since

$$y_t - \log F(K_t, L_t, M_t) = \omega_t + \varepsilon_t \tag{5}$$

the known part of productivity can be written as

$$\omega_t = \mathcal{Y}_t + \mathscr{C}(K_t, L_t) \tag{6}$$

Now Gandhi et al. (2013) make another assumption common in this literature:

Assumption 6 (Markov Productivity) *The known shock follows a first-order Markov process. For some function Ψ, productivity at t can be written as $\omega_t = \Psi(\omega_{t-1}) + \eta_t$.*

Then

$$\mathcal{Y}_t + \mathscr{C}(K_t, L_t) = \Psi(\mathcal{Y}_{t-1} + \mathscr{C}(K_{t-1}, L_{t-1})) + \eta_t \tag{7}$$

Since this is a Markov process the innovation in productivity η_t does not depend on capital K_{t-1} or labor L_{t-1}. Gandhi et al. (2013) assume capital and labor were chosen at $t-1$ before the innovation is known. Then capital and labor cannot depend on η_t, making them valid instruments that can be used to estimate $\mathscr{C}$ nonparametrically. Combined with the estimate of $\mathscr{I}_t$ this estimate of $\mathscr{C}$ gives the production function.

An alternative to the first-order approach is the proxy variable approach of Ackerberg et al. (2015), who build on the work of Olley and Pakes (1996), Levinsohn and Petrin (2003), and Wooldridge (2009). Ackerberg et al. estimate a value-added production function $\tilde{F}(K_t, L_t)$ rather than a gross production function. Unlike Gandhi et al. they need not assume the choice of intermediates is optimal, only that it is strictly increasing in productivity and depends only on productivity, labor, capital, and other observables. They drop Assumption 5 and instead assume:

Assumption 7 (Monotonicity) *All else equal, the choice of intermediates M_t is strictly increasing in ω_t.*

Assumption 8 (Scalar Unobservable) *The choice of intermediate inputs is $M_t = \bar{M}(K_t, L_t, \omega_t)$ for some smooth function $\bar{M}(\cdot)$.*

If we assume F is a neoclassical production function (concave and increasing in its arguments), Optimal Choices implies these two assumptions, but the converse need not hold. If either of these two assumptions fails, Optimal Choices must also fail. It is the second of these assumptions—the Scalar Unobservable assumption—that is the focus of this paper.

Under these assumptions productivity can be written as $\omega_t = \bar{M}^{-1}(K_t, L_t, M_t)$ for some function $\bar{m}^{-1}$. After controlling for capital and labor, intermediates are a proxy for productivity. Define $\tilde{f}(k_t, \ell_t) = \log \tilde{F}(e^{k_t}, e^{\ell_t})$ and $\bar{m}^{-1}(k_t, \ell_t, m_t) = \bar{M}^{-1}(e^{k_t}, e^{\ell_t}, e^{m_t})$. The log of value-added output can be written as

$$\begin{aligned} y_t &= \tilde{f}(k_t, \ell_t) + \bar{m}^{-1}(k_t, \ell_t, m_t) + \varepsilon_t \\ &= \Phi(k_t, \ell_t, m_t) + \varepsilon_t \end{aligned} \tag{8}$$

The term $\Phi(k_t, \ell_t, m_t)$ can be estimated nonparametrically, giving a consistent estimate of $\log \tilde{f}(k_t, \ell_t) + \omega_t$. By the Markov Productivity assumption,

$$y_t - \tilde{f}(k_t, \ell_t) - \Psi\big(\hat{\Phi}(k_{t-1}, \ell_{t-1}, m_{t-1}) - \tilde{f}(k_{t-1}, \ell_{t-1})\big) = \eta_t + \varepsilon_t \tag{9}$$

which is uncorrelated with $(k_t, \ell_t, m_{t-1}, k_{t-1}, \ell_{t-1}, \ldots)$. These variables are instruments that can be used to estimate the value-added production function by the generalized method of moments.

2.2 The Scalar Unobservable Assumption Fails When Firms Are Constrained

But suppose firms cannot choose their inputs freely. To be precise, suppose that each firm has a constraint Z_t on its choice of intermediates. The constraint may be a function of capital (which may be offered as collateral) and one or more other terms $S_t^1, S_t^2, \ldots, S_t^I$. These terms, some or all of which may be unobserved, might comprise retained earnings, the wealth of the entrepreneur, or her political connections to state-run banks.

Append to the production function (1) and the firm's optimization (2) the following conditions :

$$M_t \leq Z_t = \bar{Z}(K_t, S_t^1, \ldots, S_t^I) \quad (10)$$

$$S_t^i = \Gamma^i(S_{t-1}^i, S_{t-2}^i, \ldots) + v_t^i \quad \text{for all } i = 1, \ldots, I \quad (11)$$

where v_t^i is a serially independent shock. Equation 11 states that the other components of the constraint follow stochastic processes that need not have the Markov property.

Let λ be the Lagrange multiplier on (10). The new first-order condition is

$$1 + \lambda(K_t, S_t^1, \ldots, S_t^I, \ldots) = \mathbb{E}[e^{\varepsilon_t}]e^{\omega_t} F_M(K_t, L_t, M_t). \quad (12)$$

It is immediately clear that Assumption 5 of Optimal Choices fails whenever $\lambda > 0$—that is, whenever any firms are constrained. Rearrange this expression and invert F_M to show that the level of intermediate inputs is now

$$M_t = \dot{M}(K_t, L_t, \omega_t, \lambda_t) = \ddot{M}(K_t, L_t, \omega_t, S_t^1, \ldots, S_t^I) \quad (13)$$

which depends on more than one unobservable: productivity ω_t and one or more terms $\{S_t^j\}$. Assumption 8, the Scalar Unobservable assumption, also fails.

2.3 Generalization: Imperfections Other than Constraints

Throughout the paper I treat the suboptimal choice as having been caused by a credit constraint. The focus on constraints is largely for the sake of exposition. All of the important results hold for any unobserved feature of the firm or the economy that gives some firms easier access to intermediates. If some firms get an unobserved discount on their inputs because they are regular customers; or if some firms suffer periodic power cuts because their town elected a mayor of the opposition party; or if some firms are new to the industry and have not found enough suppliers to meet their needs; then the Scalar Unobservable assumption fails. The results of Sections 4 and 5 are no less valid.

To make this point more formally, let τ_t denote a market distortion that may vary across firms. The distortion drives a wedge between the marginal products of firms that appear similar. Then (12) is modified to

$$1 + \tau_t = \mathbb{E}[e^{\varepsilon_t}]e^{\omega_t} F_M(K_t, L_t, M_t). \tag{14}$$

As long as τ is a function of unobserved terms that cause the actual choice of m_t to deviate from the undistorted optimum, all of the important results that follow will hold.[6]

3 A Specification Test: Are Firms Able to Make Optimal Choices?

3.1 Approach

Using arguments similar to those used to derive Equation 8, it is easy to see that Assumptions 4, 7, and 8 imply that there exists a function $\xi(k_t, \ell_t, m_t)$ such that

$$y_t = \xi(k_t, \ell_t, m_t) + \varepsilon_t \tag{15}$$

where y_t is gross output (if the researcher is estimating a gross production function F) or value-added output (if the researcher is estimating a value-added production function $\tilde{F}$). Let $s_t^M = \log(M_t/Y_t)$ denote the log of the share of intermediates in output. Multiply both sides of (15) by -1 and add m_t to both sides:

$$s_t^M = \bar{\xi}(k_t, \ell_t, m_t) - \varepsilon_t \tag{16}$$

where $\bar{\xi}(k_t, \ell_t, m_t) = m_t - \xi(k_t, \ell_t, m_t)$.

Equation 16 implies the systematic variation in the share of intermediates is a function of only k_t, ℓ_t, and m_t. To be precise, after controlling for these inputs the residual variation in the share is simply ε_t, which is uncorrelated with any variable known before time t. Let $\boldsymbol{r}_{t-1}$ be a vector of instruments dated $t-1$ or earlier. Then if Assumptions 4 and 7 hold, one simple test of the Scalar Unob-

[6] The sole exception is the inequality in Equation 26 of Proposition 2. Though the expression for the bias of the estimate is valid regardless of the source of the market imperfection, the sign of the bias may or may not be negative.

servable assumption is to estimate the semiparametric regression

$$s_t^M = \bar{\xi}(k_t, \ell_t, m_t) + \boldsymbol{r}_{t-1}\boldsymbol{\varrho} + e_t \tag{17}$$

and test the hypothesis $\boldsymbol{\varrho} = \mathbf{0}$. (In practice it may be useful to control for variables beyond just the nonparametric term $\bar{\xi}(\cdot)$ such as year dummies.)

A rejection of this hypothesis suggests the Scalar Unobservable assumption fails. The consequences of its failure are especially clear for the method of Gandhi et al. (2013). Equation 17 is similar to the log share regression used to estimate the elasticity of intermediates, and as I show in Section 5.2 the bias of this estimate can be written directly as a function of moments and conditional moments of Λ_t. But the consequences for the method of Ackerberg et al. (2015) and other proxy methods are also clear. Equation 17 was derived from Equation 8, which is the first stage of a proxy estimator. As I show in Appendix C.1 the constraint can be thought of as inducing non-classical measurement error in the choice of intermediates, in that the actual m_t deviates from the optimal m_t^*.

3.2 Evidence of Imperfections

I run the test on the Chilean and Colombian census of manufacturers. I run the test separately for each of the five industries considered in Gandhi et al. (2013).

Chilean Manufacturing: I use exactly the same dataset as Gandhi et al. (2013), which is the Chilean manufacturing census used in Ackerberg et al. (2006) and expanded by Greenstreet (2007).

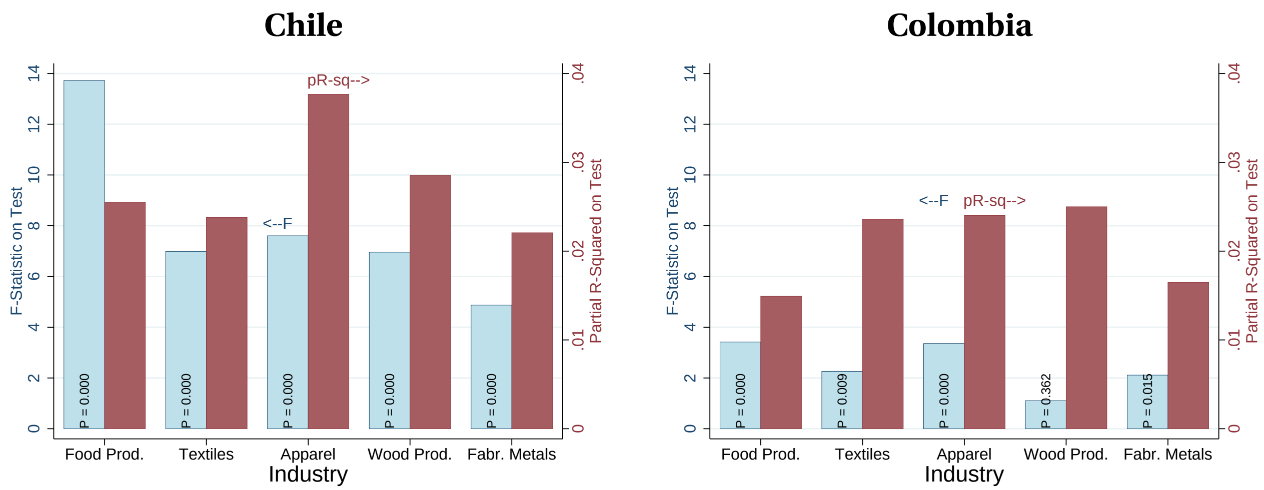
Colombian Manufacturing: I use exactly the same dataset as Gandhi et al. (2013), which is the Colombian manufacturing census.[7]

Figure 1 shows the F-statistic, p-value, and partial R-squared of the test in the two datasets of manufacturing firms.[8] The test rejects at the 5 percent le-

[7] I am grateful to Salvador Navarro for giving me the files and code needed to reproduce the data for both Chile and Colombia.

[8] As Gandhi et al. (2013) measure physical output and rescale the level of intermediates by a price index, I define $s^M = P_t^M M_t / (P_t^Y Y_t)$ as they do. In addition to the terms in Equation 17, I also control for year dummies. This is a simple way to deal with potential measurement error in the price of intermediates, which would otherwise be negatively correlated with the choice of intermediates. Running the test without year dummies makes little difference.

Figure 1
Firms are Constrained in Many Chilean and Colombian Industries



Note: All tests are based on standard errors clustered by plant.

vel in all five of Chile's industries and all but one of Colombia's industries. In many cases the p-value is close to zero. The F-statistics are on average smaller in the Colombian sample, likely because it is smaller. All five industries have at least 500 firms the Chilean sample, but this is true of only two in the Colombian sample. By contrast, the partial R-squared is of similar magnitude in both samples. As I show in Section 6, this invariance to sample size makes the partial R-squared a better gauge of the level of constraints.

As the test rejects in most of these industries, the Scalar Unobservable assumption should not be taken for granted. But in some industries the test either does not reject (Wood Products in Colombia) or the partial R-squared is relatively small (food products in Colombia). Taken together these two results suggest it is unwise to assume firms are or are not constrained. The condition should be tested.

4 The Autoregressive Method

If the Scalar Unobserable assumption fails, choice-based methods may be biased. A different class of methods drops that assumption and instead imposes linearity on the Markov process that governs productivity:

Assumption 9 (Autoregressive Productivity) *The known shock follows a first-order autoregressive process. For some parameters $(\bar{\omega}, \rho)$, productivity at $t+1$ can be written as $\omega_t = \bar{\omega} + \rho\omega_{t-1} + \eta_t$.*

As shown by Ackerberg et al. (2015), this assumption implies an estimator similar to the linear dynamic panel estimator (for example, Arellano and Bond, 1991; Blundell and Bond, 1998), though that estimator also controls for a firm-level fixed-effect. The procedure outlined here assumes no fixed-effect because most choice-based methods do not.[9] However, it is easy to allow for them by directly applying a dynamic panel estimator. Finally, in the main text I focus on estimating a gross production function, as that is the more challenging case. I discuss estimating a value-added production function in Appendix B.2.

4.1 Procedure

Let $f(k_t, \ell_t, m_t) = \log F(e^{k_t}, e^{\ell_t}, e^{m_t})$ denote the log of the gross production function. Under Autoregressive Productivity, gross output may be rewritten as

$$\begin{aligned} y_t &= f(k_t, \ell_t, m_t) + \omega_t + \varepsilon_t \\ &= f(k_t, \ell_t, m_t) + \bar{\omega} + \rho\omega_{t-1} + \eta_t + \varepsilon_t \\ &= f(k_t, \ell_t, m_t) + \bar{\omega} + \rho(y_{t-1} - f(k_{t-1}, \ell_{t-1}, m_{t-1})) + \underbrace{\eta_t - \rho\varepsilon_{t-1} + \varepsilon_t}_{\nu_t} \end{aligned} \tag{18}$$

Equation (18) can be estimated using generalized method of moments. If ε_t is only measurement error, under the timing assumptions of Section 2.1 any function of $(k_t, \ell_t, k_{t-1}, \ell_{t-1}, m_{t-1}, \ldots)$ is uncorrelated with the combined error

[9] Gandhi et al. (2013) do show how to extend their method to include an additive firm-level fixed-effect.

term ν_t. If ε_t is not measurement error but a true shock to revenue it might affect investment and hiring. Then $\rho\varepsilon_{t-1}$ may be correlated with k_t and ℓ_t, ruling these out as instruments. I make this more conservative assumption in the simulations that follow. Since m_t may be correlated with η_t, and y_{t-1} is correlated with $\rho\varepsilon_{t-1}$, neither is an instrument.[10] In the simulations below I follow Gandhi et al. (2013) in approximating $f(k_t, \ell_t, m_t)$ with a second-order translog polynomial. I instrument with the (de-meaned) lags of each term of the polynomial, and the second lags of m, k, and their interaction.[11]

4.2 Link to Two-Stage Least Squares

Though the autoregressive estimator is technically estimated using nonlinear GMM, it behaves very much like a version of two-stage least squares. To see why, consider again the example of Cobb-Douglas production. Suppose for a moment that ρ is known. Then (18) can be rewritten as

$$\begin{aligned} y_t &= \pi_k k_t + \pi_\ell \ell_t + \pi_m m_t + \bar{\omega} + \rho(y_{t-1} - \pi_k k_{t-1} - \pi_\ell \ell_{t-1} - \pi_m m_{t-1}) + \nu_t \\ y_t - \rho y_{t-1} &= \bar{\omega} + \pi_k(k_t - \rho k_{t-1}) + \pi_\ell(\ell_t - \rho \ell_{t-1}) + \pi_m(m_t - \rho m_{t-1}) + \nu_t \\ &= \bar{\omega} + \pi_k \Delta_\rho(k) + \pi_\ell \Delta_\rho(\ell) + \pi_m \Delta_\rho(m) + \nu_t \end{aligned} \tag{19}$$

where $\Delta_\rho(x) = x_t - \rho x_{t-1}$. These $\Delta_\rho(\cdot)$ terms are observed and thus be treated as outcomes or regressors. One could estimate π_k, π_ℓ, π_m by running two-stage least squares, instrumenting for $\Delta_\rho(k), \Delta_\rho(\ell), \Delta_\rho(m)$ with the vector of $\boldsymbol{r}_{t-1}$.

As it turns out, the GMM estimator does something similar to this. Suppose f is estimated using a log sieve polynomial. Define $\boldsymbol{X}_t$ as a matrix where each column is one term of this polynomial $(k_t, \ell_t, m_t, k_t^2, k_t\ell_t, \cdots)$ and each row is an observation. Let $\boldsymbol{\pi}$ be the coefficients of the sieve polynomial. Let $\boldsymbol{R}_{t-1}$ be a

[10] If labor and capital are chosen with some information about ω_t the instruments chosen when ε_t is not measurement error—all lags of capital, labor, and intermediates—are still valid for this method. However, Gandhi-Navarro-Rivers would have to drop k_t and ℓ_t from its list of instruments, leaving the two methods with the same set of instruments. This is why Assumption 4, the Timing Assumption, favors Gandhi-Navarro-Rivers over the autoregressive method.

[11] That is, let the vector of instruments be

$$\boldsymbol{r}_{t-1} = \{k_{t-1}, \ell_{t-1}, m_{t-1}, k_{t-1}^2, k_{t-1}\ell_{t-1}, k_{t-1}m_{t-1}, \ell_{t-1}^2, \ell_{t-1}m_{t-1}, m_{t-1}^2, k_{t-2}, m_{t-2}, k_{t-2}m_{t-2}\}$$

matrix of instruments (where each row gives the vector of instruments $\boldsymbol{r}_{t-1}$ for one observation). Let $\boldsymbol{y}_t$ be a vector of outcomes. The GMM estimator solves

$$\min_{\boldsymbol{\pi},\rho} \boldsymbol{R}_{t-1}^T(\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\pi} - \rho(\boldsymbol{y}_{t-1} - \boldsymbol{X}_{t-1}\boldsymbol{\pi}))W(\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\pi} - \rho(\boldsymbol{y}_{t-1} - \boldsymbol{X}_{t-1}\boldsymbol{\pi}))^T\boldsymbol{R}_{t-1} \quad (20)$$

Solving the first-order condition for $\boldsymbol{\pi}$ gives

$$\hat{\boldsymbol{\pi}} = \left[\Delta_{\hat{\rho}}(\boldsymbol{X})^T\boldsymbol{R}_{t-1}W\boldsymbol{R}_{t-1}^T\Delta_{\hat{\rho}}(\boldsymbol{X})\right]^{-1}\Delta_{\hat{\rho}}(\boldsymbol{X})^T\boldsymbol{R}_{t-1}W\boldsymbol{R}_{t-1}^T\Delta_{\hat{\rho}}(\boldsymbol{y}) \quad (21)$$

$$(22)$$

while $\hat{\rho}$ is estimated by solving the nonlinear equation

$$0 = -\boldsymbol{R}_{t-1}^T(\rho(\boldsymbol{y}_{t-1} - \boldsymbol{X}_{t-1}\hat{\boldsymbol{\pi}}))W(\boldsymbol{y}_t - \boldsymbol{X}_t\hat{\boldsymbol{\pi}} - \rho(\boldsymbol{y}_{t-1} - \boldsymbol{X}_{t-1}\hat{\boldsymbol{\pi}}))^T\boldsymbol{R}_{t-1} \quad (23)$$

after substituting the expression for $\hat{\boldsymbol{\pi}}$.

If ρ were known and $W = (\boldsymbol{R}_{t-1}^T\boldsymbol{R}_{t-1})^{-1}$ then the expression for $\hat{\boldsymbol{\pi}}$ would simply be the two-stage least squares estimator. What the GMM estimator effectively does is solve the nonlinear equation (23) for ρ and plug this estimate into (21). If the GMM estimate of ρ is consistent then (21) would converge to the two-stage least squares estimate. Stock and Wright (2000) lay out the conditions under which certain parameters of a GMM estimator are consistent even under weak identification asymptotics (to be precise, when they follow a drifting process that assumes identification is local to zero). The following proposition, which is proven in Appendix A.1, shows that these conditions hold:

Proposition 1 *Let N denote the sample size. Define the matrix*

$$\Xi(\boldsymbol{z}) = \sqrt{N}\left(\frac{\sum_{n=1}^N \boldsymbol{r}_{t-1}^{(n)}\boldsymbol{z}_n^T}{N} - \mathbb{E}[\boldsymbol{r}_{t-1}^{(n)}\boldsymbol{z}_n^T]\right)$$

and let

$$\boldsymbol{\Xi}_T = \begin{bmatrix}\Xi(y_t) & \Xi(\boldsymbol{x}_t) & \Xi(y_{t-1}) & \Xi(\boldsymbol{x}_{t-1})\end{bmatrix}$$

be a block matrix.

Assume

1. $\Xi_N \to_D \bar{\Xi} \sim N(\mathbf{0}, \Sigma^{\Xi})$

2. $\mathbb{E}[\boldsymbol{r}_{t-1}\omega_{t-1}] = \Upsilon \neq \mathbf{0}$

Then under the weak instruments asymptotics of Stock and Wright (2000), the autoregressive estimate for ρ is consistent (regardless of whether the other parameters are weakly identified).

The first assumption holds as long as the appropriate central limit theorem holds for output and for the inputs of production. The second assumption holds as long as the choices of inputs depends in some way on current or past productivity (as ω_{t-1} depends on ω_{t-2}, and so on). The

The consistency of ρ implies there is a close connection between the autoregressive and two-stage least squares estimates of $\boldsymbol{\pi}$.

This connection is useful because the effect of weak instruments on two-stage least squares has been longer studied and is better understood than the effect of weak identification on nonlinear GMM. For example Stock and Yogo (2005) show that under some assumptions, weak instruments bias two-stage least squares towards ordinary least squares. There are also diagnostic statistics (e.g. Stock and Yogo, 2005; Kleibergen and Paap, 2006) common in the applied literature that can be used, as I show below, to assess the severity of weak instruments.

5 Complementarity in Estimating the Gross Production

Though it may seem clear that the success or failure of choice-based estimators hinges on on how badly firms are constrained, this section argues that—at least when estimating a gross production function—the same is to some extent true of autoregressive estimators. While choice-based methods grow more biased as constraints tighten, an autoregressive estimator for the gross production function becomes less biased. This complementarity arises because although constraints undermine the identification assumptions of choice-based methods, they strengthen the instruments used for identification in the autoregressive estimator.

5.1 Intuition: Cobb-Douglas Case

Though the exclusion restriction of the autoregressive estimator holds regardless of whether the Scalar Unobservable assumption fails, the rank condition essentially requires it to fail. This ironic feature is the key to its complementarity with Gandhi-Navarro-Rivers. It is easiest to see this rank failure when production is Cobb-Douglas and choices are optimal. Let $F(K_t, L_t, M_t) = K_t^{\pi_k} L_t^{\pi_\ell} M_t^{\pi_m}$. After de-meaning, the optimal choice of intermediates is

$$m_t = \frac{1}{1-\pi_m}\left[\omega_t + \pi_k k_t + \pi_\ell \ell_t\right] \tag{24}$$

Let $\boldsymbol{r}_{t-1} = [r_{t-1}^1, \ldots, r_{t-1}^q]$ be a vector of valid instruments. The parameter π_M is identified only if a change in the value of π_m induces changes in the moment conditions $\mathbb{E}[\nu_t r_{t-1}^1] = \cdots = \mathbb{E}[\nu_t r_{t-1}^q] = 0$ that are linearly independent of those induced by changes in the other parameters π_k, π_ℓ. (In an ordinary least squares regression, this assumption is equivalent to saying the regressors are not perfectly collinear.) A change in π_m equals

$$\begin{aligned}
\frac{\partial \mathbb{E}[\nu_t r_{t-1}^n]}{\partial \pi_m} &= \mathbb{E}[r_{t-1}^n(\rho m_{t-1} - m_t)] \\
&= \mathbb{E}\left[r_{t-1}^n \frac{1}{1-\pi_m}\Big\{(\rho\omega_{t-1} - \omega_t) + \pi_k(\rho k_{t-1} - k_t) + \pi_\ell(\rho \ell_{t-1} - \ell_t)\Big\}\right] \\
&= \mathbb{E}\left[r_{t-1}^n \frac{1}{1-\pi_m}\Big\{\eta_t + \pi_k(\rho k_{t-1} - k_t) + \pi_\ell(\rho \ell_{t-1} - \ell_t)\Big\}\right] \\
&= \frac{1}{1-\pi_m}\mathbb{E}[r_{t-1}^n \eta_t] + \frac{\pi_k}{1-\pi_m}\frac{\partial \mathbb{E}[\nu_t r_{t-1}^n]}{\partial \pi_k} + \frac{\pi_\ell}{1-\pi_m}\frac{\partial \mathbb{E}[\nu_t r_{t-1}^n]}{\partial \pi_\ell}
\end{aligned}$$

where, as before, all variables are de-meaned. By the exclusion restriction r_{t-1}^n is uncorrelated with η_t, implying the first term is zero. But then any change in the moment condition induced by π_m is perfectly collinear with those induced by π_k and π_ℓ, implying π_m is not identified.

As shown in Section 5, this result generalizes to any non-parametrically estimated production function. The intuition is similar to the functional dependence critique raised in Ackerberg et al. (2015) and Gandhi et al. (2013). The optimal choice of intermediates is perfectly determined by the choices of other inputs and productivity, which depends on its own lag and an innovation. After controlling for this lag and the other inputs, the only remaining variation is the innovation. Since this variation cannot be used for identification without violating the exclusion restriction, there is no way to identify the effect of intermediates on output.

But if firms are constrained, the unobserved elements that determine the constraint $S_t^1, S_t^2, \ldots, S_t^I$ will induce additional systematic variation in the choice of intermediates. As I show in Section 5, there is a direct complementarity between Gandhi-Navarro-Rivers and the autoregressive estimator.

5.2 General Result

When firms are potentially constrained the Lagrange multiplier on the constraint—its mean, variance, and correlation with the instruments—governs the relative bias of Gandhi-Navarro-Rivers and the autoregressive estimator. The complementarity between them rests on the fact, embodied in two theorems below, that the Lagrange multiplier has opposite effects on their consistency.

First, formalize the notion of constraints by modifying Assumption 5 as follows:

Assumption 10 (Constrained Optimal Choices) *Define $\Lambda_t = \log(1 + \lambda_t)$, where λ_t is a Lagrange multiplier that gives the shadow cost to the firm of being unable to choose M_t optimally. Then the choice of the firm satisfies*

$$\Lambda_t = \omega_t + \log \mathbb{E}[e^{\varepsilon_t}] + \log F_M(K_t, L_t, M_t) \tag{25}$$

As noted earlier, nearly all of the results that follow hold even if the imperfection is not a constraint but some other distortion τ_t that drives a wedge between the marginal products of firms. Redefine $\Lambda_t = \log(1 + \tau_t)$ and all of the following results still hold.

5.2.1 The Bias of Gandhi-Navarro-Rivers Increases with Constraints

The following proposition, which is proven in Appendix A, shows that Gandhi-Navarro-Rivers grows more biased as constraints tighten:

Proposition 2 *Let $\ddot{\Lambda}_t = \Lambda_t - \mathbb{E}[\Lambda_t \mid k_t, \ell_t, m_t]$. The the average bias of the Gandhi-Navarro-Rivers estimate of the elasticity of intermediates is*

$$-\left(\mathbb{E}[\Lambda_t] + \log \mathbb{E}[e^{\ddot{\Lambda}_t}]\right) \tag{26}$$

In the special case where $\ddot{\Lambda}_t \sim N(0, \sigma^2_\Lambda)$ the bias is simply $-(\mathbb{E}[\Lambda_t] + \sigma^2_\Lambda/2)$.

Finally, if the market imperfection is caused by constraints like that shown in the constrained optimization (12), then the bias is less than or equal to zero (with equality holding if and only if firms are unconstrained).

The proposition implies that an increase in either the mean of Λ_t or the variability of $\ddot{\Lambda}_t$, the residual from a nonparametric regression of Λ_t on (k_t, ℓ_t, m_t), increases the absolute value of the bias. When there is a bias it is always downward—constraints make output appear less responsive to intermediates than in truth. Since the second stage of Gandhi-Navarro-Rivers relies on a consistent estimate of the elasticity of intermediates, the other parameters—in particular, the elasticities of capital and labor—will also be biased.

5.2.2 The Bias of the Autoregressive Estimator Decreases with Constraints

Like any GMM estimator, the autoregressive estimator is under-identified if the Jacobian of the moment conditions is singular at the true value of the parameters. Section 5.1 shows that the Jacobian is singular in the special case where firms are unconstrained and the production function is known to be Cobb-Douglas. In practice the production function is unknown and must be estimated non-parametrically, and the fraction of firms that are constrained will be neither 0 or 1, but lie somewhere in between. In that case the Jacobian of the moment conditions is not singular, but it may be near-singular. Though in principle the estimator is identified, in practice such weak identification will bias nonlinear GMM much as weak instruments bias two-stage least squares (Stock and Wright, 2000).

This problem arises for the autoregressive estimator when constraints are relatively slack. One measure of near-singularity, the smallest singular value of the Jacobian of the moment conditions, is bounded above by a statistic that is increasing in the severity of constraints and equals zero when firms are unconstrained.[12] But the statement of this proposition is not intuitive and is thus, together with its proof, relegated to Appendix A.3.

It is more intuitive to exploit the link between the autoregressive estimator and two-stage least squares (see Section 4.2). The severity of constraints, as I show below, places an upper bound on the size of a common measure of weak instruments. Assume that $\log F(K_t, L_t, M_t)$ and $\log F_M(K_t, L_t, M_t)$ each has a polynomial sieve approximation in logs:

$$\log F(K_t, L_t, M_t) = \underbrace{\sum_{a_1^0, a_2^0, a_3^0} A^0_{a_1^0, a_2^0, a_3^0} k_t^{a_1^0} \ell_t^{a_2^0} m_t^{a_3^0}}_{f(k_t, \ell_t, m_t)}$$
$$\log F_M(K_t, L_t, M_t) = \sum_{a_1^1, a_2^1, a_3^1} A^1_{a_1^1, a_2^1, a_3^1} k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1} \tag{27}$$

Let $\boldsymbol{x}_t$ denote the terms of the sieve approximation. As in Section 4.2, assume for now that ρ is known and denote $\Delta_\rho(x) = x_t - \rho x_{t-1}$. Consider the two-stage least squares regression that instruments $\Delta_\rho(\boldsymbol{x})$ with $\boldsymbol{r}_{t-1}$. One measure of the joint strength of the instruments is the Cragg-Donald statistic, which is the equivalent of the first-stage F-statistic in a case with multiple endogenous regressors. The following proposition, proven in Appendix A.4, links the Cragg-Donald statistic to the log-Lagrange multiplier Λ_t:

Proposition 3 *Let $\boldsymbol{V}$ be the matrix of residuals from a regression of $\Delta_\rho(\boldsymbol{x})$ on the matrix of instruments, and let $\Sigma_{\boldsymbol{V}}$ be the variance matrix. Let $\Delta_\rho(\widehat{\dot{\boldsymbol{x}}})$ and $\Delta_\rho(\widehat{\Lambda})$ be the predicted values from a regression of $\Delta_\rho(\dot{\boldsymbol{x}})$ and $\Delta_\rho(\Lambda)$ on the instruments (as would be generated in the first stage of two-stage least squares). Let β be the vector of coefficients and $[SSR]^{\Delta_\rho(\widehat{\Lambda})}$ the sum of squared residuals from a regression of $\Delta_\rho(\widehat{\Lambda})$ on $\Delta_\rho(\widehat{\dot{\boldsymbol{x}}})$. Let $q_0 = [A^1_{0,0,1} \vdots (\dot{\boldsymbol{A}}^{\mathbf{1}} - \beta)^T]^T$. Then the Cragg-Donald Statistic is asymptotically bounded above by*

$$\frac{[SSR]^{\Delta_\rho(\widehat{\Lambda})}}{q_0^T \Sigma_{\boldsymbol{V}} q_0} \tag{28}$$

[12]This approach follows in the spirit of Wright (2003), who derives a test for under-identification based on the distance between the Jacobian and the closest matrix that is not of full column rank; and of Kleibergen and Paap (2006), who propose a widely used test for weak instruments based on the singular value decomposition.

The numerator of (28) measures the extent to which the instruments induce variation in the Lagrange multiplier that is independent of that induced in the (differenced) terms of the sieve approximation other than $\Delta_\rho(m)$. The denominator is a rough measure of the fraction of the unexplained variation in $\Delta_\rho(m)$ that is independent of the unexplained variation in the other terms. In short, (28) is almost like an F-statistic.

If firms are unconstrained then $\Lambda_t = \Lambda_{t-1} = 0$, which implies $\Delta_\rho(\Lambda) = 0$ and thus (28) is zero. Then Proposition 3 implies the estimator is under-identified, generalizing the result from Section 5.1 to a nonparametric estimator. If firms are only slightly constrained, meaning constraints explain only a small part of their choice of intermediates, (28) will be small and the estimator is at best only weakly identified.

Though ρ is in fact unknown and the two-stage least squares estimator described in the main text is infeasible, as noted in Section 4.2 the autoregressive estimator is a close cousin. As proven in Appendix A.3, the local identification condition for nonlinear GMM, much like the Cragg-Donald statistic, hinges whether the instruments induce enough independent variation in the Lagrange multiplier.

5.2.3 Implications

The point of these propositions is not to show how to estimate the bias of either estimator (which is impossible because the Lagrange multiplier is unobserved). Propositions 2 and 3 imply that as the severity of constraints increases, the bias of Gandhi-Navarro-Rivers increases and the bias of the autoregressive estimator decreases. Both propositions imply the change is smooth and monotonic (with the caveat that Proposition 3 derives a bound rather than an equality). Then there is a range where although firms are constrained, Gandhi-Navarro-Rivers is still the less biased estimator. The challenge, addressed in Section 7, is to find a point at which to switch from one estimator to the other. That fact that switching between estimators can reduce overall bias is the key implication of complementarity.

One key caveat, however, is that the biases of the estimators depend on similar but not identical statistics. Gandhi-Navarro-Rivers depends on the average constraint as well as the variation unexplained by the other inputs (see Proposition 2). But the identification of the autoregressive estimator improves (and thus the bias decreases) *only* in the independent variation induced by the instruments. It is possible that firms are heavily constrained but the instruments are uninformative about the constraint. By this logic the two estimators are only imperfect complements. But if the constraint follows a stochastic process like that shown in Equation 11, then $m_{t-1}, m_{t-2}, \ldots$ should be informative about the constraint (because they are informative about its lags).

5.3 Estimating a Value-Added Production Function: There is No Complementarity

The weak identification problem of Sections 5.1 and 5.2 do not arise when estimating a value-added production function. Recall that the autoregressive estimator is weakly identified when firms are unconstrained because there is little or no independent variation in the choice of intermediates. The GMM estimator is unable to separate the effect of intermediates on output from the effects of the other inputs.

But a value-added production function assumes this effect is zero. Since intermediates are known not to enter the production function, GMM does not need independent variation in intermediates. As long as lagged capital and lagged labor induce independent variation in current capital and labor—which is guaranteed by Assumption 3—the value-added production function is well-identified. In other words, though constraints will bias the Ackerberg-Caves-Frazer estimator (see Appendix C.1), the autoregressive estimator does not require them to be biased. There is no trade-off and thus no complementarity. That said, it may still make sense to test for constraints before choosing the autoregressive estimator because Ackerberg-Caves-Frazer does not require Assumption 9 (Autoregressive Productivity).

6 Simulations

Section 2 shows that constraints may cause the Scalar Unobservable assumption to fail, in which case choice-based methods may be inconsistent while autoregressive methods become well-identified. But as the severity of constraints increases, how rapidly does the performance of choice-based methods deteriorate? How rapidly does the performance of autoregressive methods improve? Can the test for constraints help choose the method with lower bias?

I answer these questions by running Monte Carlo simulations. Throughout I focus on estimating a gross production function, as it is easier to form a data generating process grounded in economic theory when intermediates appear in the production function.[13] I compare the performance of the autoregressive method to the method of Gandhi et al. (2013). I then assess whether the test for constraints successfully detects their presence, and whether it is informative about the bias of the two methods.

6.1 Setup

I assume the production function is

$$Y_t = e^{\omega_t + \varepsilon_t} X(K_t, L_t) M_t^{\theta^M}$$

which gives a closed form solution for the (constrained) optimal choice of M_t even if $X(K_t, L_t)$ is not Cobb-Douglas. In the baseline case, known productivity ω_t evolves according to

$$\omega_{t+1} = \rho\omega_t + \eta_{t+1} \tag{29}$$

The initial distributions of log capital k and log labor ℓ are normal and evolve according to

$$\begin{aligned} k_{t+1} &= \alpha_0^k + \alpha_1^k k_t + \alpha_2^k \eta_t + \alpha_3^k \eta_{t-1} \\ \ell_{t+1} &= \alpha_0^\ell + \alpha_1^\ell k_t + \alpha_2^\ell \eta_t + \alpha_3^\ell \eta_{t-1} \end{aligned}$$

which allows one-year and two-year adjustment lags.

The firm chooses its intermediate inputs subject to a credit constraint. The constraint depends on the firm's capital and its wealth W_t, which has two com-

[13] The results from estimating a value-added production function using Ackerberg et al. (2015) are reported in Appendix B.2.

ponents S_t^1, S_t^2. Both parts evolve according to

$$S_t^i = \alpha_0^S + \alpha_1^S S_{t-1}^i + \alpha_2^S S_{t-2}^i + v_t^i \quad \text{for } i = 1, 2$$

where $\{v_t^1, v_t^2\}$ are independent and normally distributed white noise.

The firm's problem is

$$\max_{M_t} \mathbb{E}[e^{\omega_t + \varepsilon_t} X(K_t, L_t) M_t^{\theta^M}] - M_t$$

subject to a credit constraint. Let $\bar{K}, \bar{W}$ be the average level of capital and wealth. The constraint is

$$M_t \leq \zeta \left(\frac{K_t}{\bar{K}}\right)^{\frac{1}{2}} \left(\frac{W_t}{\bar{W}}\right)^{\frac{1}{2}}$$
$$W_t = \exp\left[\sqrt{\varphi} S_t^1 + (1 - \sqrt{\varphi}) S_t^2\right]$$

with $\varphi = .5$.[14] The multiplier ζ gives the credit limit of a firm with the average level of capital and wealth. In a country with a strong financial market ζ is large, letting even a relatively poor firm borrow as much as it needs. In the simulations I vary this parameter to change the fraction of firms that are constrained.

The constrained optimal choice of intermediates is

$$M_t^* = \min\left\{ \left[\theta^M e^{\omega_t} \mathbb{E}[e^{\varepsilon_t}] X(K_t, L_t)\right]^{\frac{1}{1-\theta^M}} \quad , \quad \zeta K_t^{\frac{1}{2}} W_t^{\frac{1}{2}} \right\} \tag{30}$$

In the baseline case I assume

$$X(K_t, L_t) = \left[\theta^K K_t^{\frac{\epsilon-1}{\epsilon}} + \theta^L L_t^{\frac{\epsilon-1}{\epsilon}}\right]^{\sigma \frac{\epsilon}{\epsilon-1}}$$

with $\epsilon = 5$. As described in Appendix B.1, all of the other parameters and the moments of each distribution are calibrated to match industry 311 from Gandhi et al. (2013). The only exception is that I center log productivity around zero. This assumption, which only affects the average level of intermediates chosen, is not important because the constraint ζ is chosen to ensure the desi-

[14] The standard deviation of v_t^1 and v_t^2 are both calibrated to match the log of short-term assets in Chile's Industry 311. As a result, $SDev(S_t) = SDev(S_t^2)$ in expectation. This implies that W_t will have the same mean and variance regardless of the value chosen for φ.

red fraction of firms is constrained.[15] I set the number of firms to 2613, the number of unique firms in industry 311, and the length of the panel to 7, roughly the average number of years per firm.

6.2 Estimation

For each specification I choose levels of ζ that on average make the constraint bind for 1, 10, 20, ..., 80 percent of choices. (Each firm makes one choice of intermediates each year). For each level of constraint I simulate 360 datasets. In each I estimate the production function using Gandhi-Navarro-Rivers (GNR) and the autoregressive method (AR), as well as computing several testing statistics (including the test proposed in Section 3).

I estimate Gandhi-Navarro-Rivers using code written by the authors.[16] I make only one change: in any simulation where the true Markov process for productivity is autoregressive, I impose this assumption. By doing so I avoid favoring the autoregressive method, which by construction imposes the (true) autoregressive assumption. In one set of simulations I let productivity be a third-order polynomial, as assumed by Gandhi et al. (2013). In these simulations I estimate Gandhi-Navarro-Rivers assuming productivity takes this form, again to avoid penalizing Gandhi-Navarro-Rivers.

I apply the autoregressive estimator to a second-order polynomial in logs. I first demean every variable to remove the constant in the autoregressive process. (In the true process the constant is zero, but in a real application the researcher would not know that.) I then estimate the equation using as instruments the one-period lags of every term in the second-order polynomial, and also the two-period lags of intermediates, capital, and their product. By using only lags of capital and labor as instruments I avoid the unrealistic assumption that the unanticipated shock ε is only measurement error.

[15]The constraint effectively scales up or down with average productivity. However, centering productivity around zero reduces the grid space over which I must search to find the ζ that constrains the desired number of firms.

[16]Special thanks to Salvador Navarro for the code.

6.3 How do the Estimators and Testing Statistics Behave as Constraints Tighten?

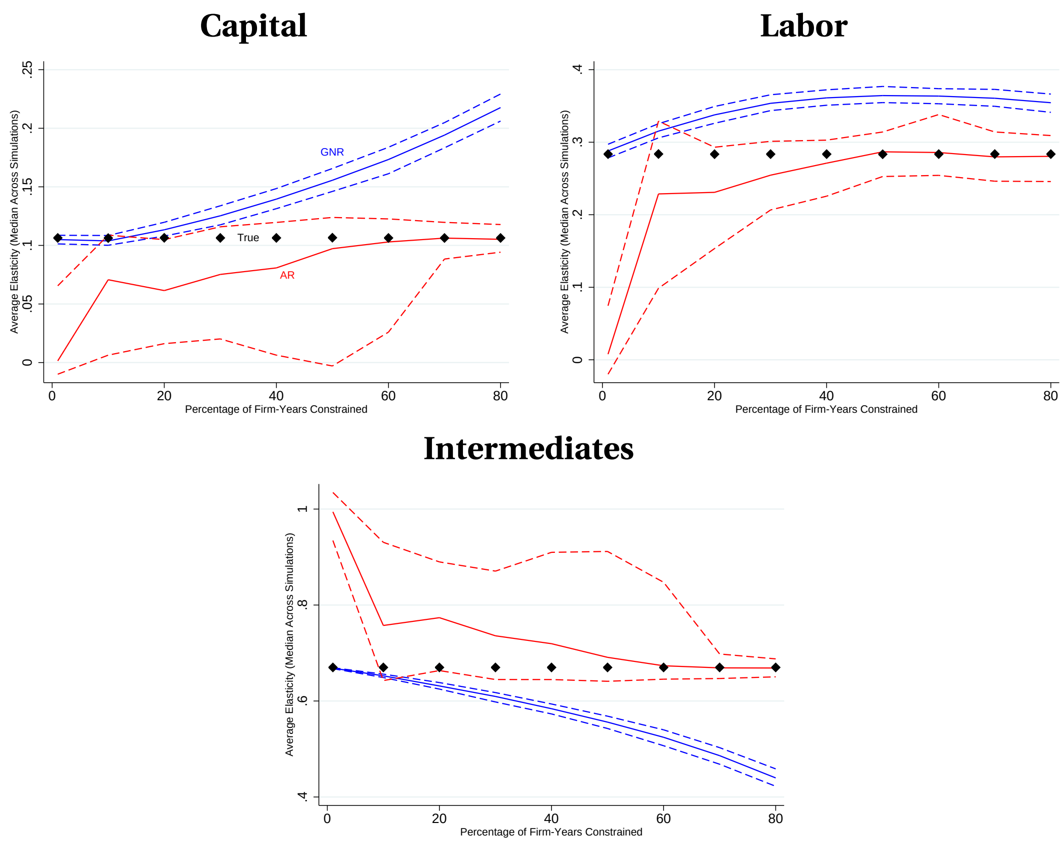
For each level of constraint I estimate the production function using both Gandhi-Navarro-Rivers and the autoregressive method. I estimate the average elasticity of output with respect to capital, labor, and intermediates by computing the elasticity for each firm in each year and taking the average. Figure 2 plots the 10th, 50th, and 90th percentiles of the average elasticity for Specification 1 at each level of constraint. The other specifications show a similar pattern to the baseline.

The estimates of Gandhi-Navarro-Rivers are far less noisy than those of the autoregressive method. For all three elasticities the difference between the 90th and 10th percentiles of the estimates is small. When firms are unconstrained, Gandhi-Navarro-Rivers gives precise and accurate estimates. But when firms are constrained the estimates, though still precise, are severely biased. As implied by Proposition 2, the estimate of the elasticity of intermediates is biased downward. Though the proposition does not make predictions about the elasticities of labor and capital, the simulations suggest both are biased upward in the presence of constraints.

The pattern of the bias is ironic, as ordinary least squares, fixed effects, and other linear methods are often criticized for doing the opposite. For example, Ackerberg et al. (2015) write "one common finding [about the fixed-effects estimator] is unreasonably low estimates of" the elasticity of output with respect to capital (p. 3). In other words, the fact that linear methods give higher estimates for intermediates and lower estimates for capital is taken as a sign that they are biased. But Proposition 2 and Figure 2 suggest it could just as likely be a sign that choice-based methods are biased.

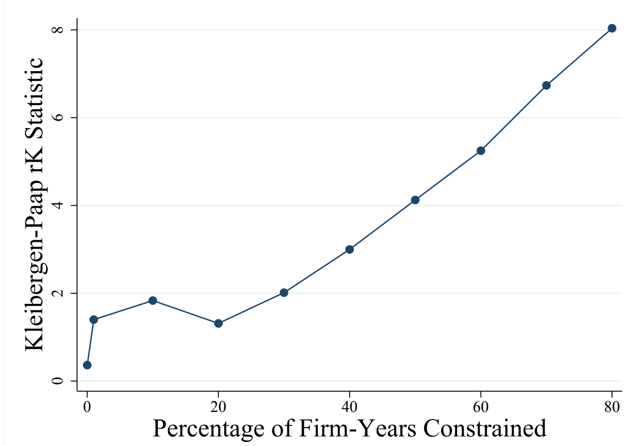
By contrast, the autoregressive estimator is precise and accurate only when firms are heavily constrained. As firms become unconstrained the estimates grow noisy and biased. As noted in Section 5, this bias arises because the effect of intermediates on output is only weakly identified. Figure 3 shows how the rK statistic of Kleibergen and Paap (2006) changes as constraints vary. This figure computes the statistic using the true value of ρ, making it infeasible (I outline how to compute a feasible rk statistic below). Nevertheless it is infor-

Figure 2
Estimated Elasticities: Baseline Specification



Note: I estimate the elasticity of output with respect to each factor of production using both Gandhi-Navarro-Rivers (GNR) and the autoregressive method (AR). I compute the average elasticity across all firms in each simulation. I then compute the median (solid line) and 90th and 10th percentiles (dashed lines) of the average elasticity across all 360 simulations.

Figure 3
The (Infeasible) rK-Statistic Rises with Constraints

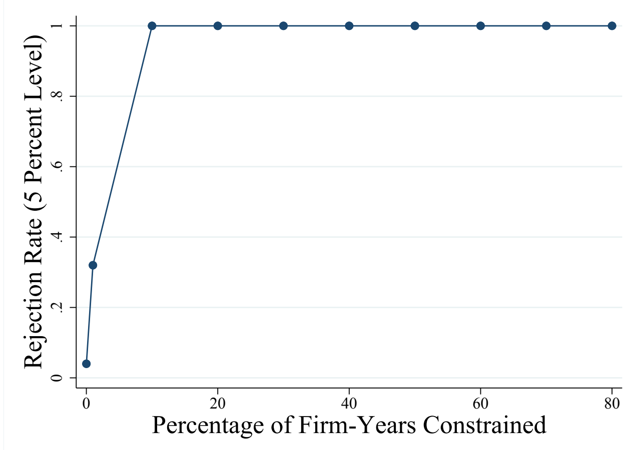


Note: This figure shows the rK statistic of Kleibergen and Paap (2006), which generalizes the Cragg-Donald statistic for weak instruments. The statistic is computed in each of 100 simulations at each level of constraints using the true value of ρ (which is why it is infeasible).

mative about the strength of identification. The figure shows that the average rK statistic is close to zero when a negligible fraction of firms is constrained. It remains low until 40 percent or more of firm-years are constrained. Figures 2 and 3 confirm that the two methods are complementary.

Figure 4 shows the rejection rate of the test for constraints. When firms are essentially unconstrained (just 0.01 percent of firm-years constrained), the rejection rate is equals the size of the test. When even 1 percent of firms are constrained the rejection rate jumps to over 30 percent, and when at least 10 percent of firm-years are constrained the test always rejects. In practice the power of the test will depend not only on the level of constraints but the variance of the unanticipated shock and the covariances of the other terms. But this figure suggests it has good power for simulations calibrated to match an actual manufacturing industry.

Figure 4
The Test of Section 3 Rejects in the Presence of Constraints



Note: This figure shows the rejection rate of the specification test. The test uses a degree-2 control function and tests for the significance of $k_{t-1}, \ell_{t-1}, m_{t-1}, k_{t-2}, m_{t-2}, k_{t-2}m_{t-2}$. Inference is clustered within firm.

7 Can the Test be Used to Choose Between Methods? A Simple Demonstration

This section gives a simple demonstration of how the test might be used to choose between choice-based and autoregressive estimators. Given that the two estimators are complementary, in an ideal world one would choose whichever has lower error. Which of the two performs better would depend on the context. For some statistic of interest—say, ϕ^X, the elasticity of output with respect to input X—consider an (infeasible) estimator that sets

$$\hat{\phi}^X_{Inf} = \begin{cases} \hat{\phi}^X_{GNR} & \text{if } |\hat{\phi}^X_{GNR} - \phi^X| < |\hat{\phi}^X_{AR} - \phi^X| \\ \hat{\phi}^X_{AR} & \text{otherwise} \end{cases}$$

This estimator is infeasible because it requires knowing the true elasticity ϕ^X (or at least the true error). But measuring the performance of this infesible estimator puts an upper bound on the potential gains from choosing between choice-based and autoregressive methods rather than relying on choice-based methods alone.

Figure 5
A Combined Approach Gives More Accurate Estimates

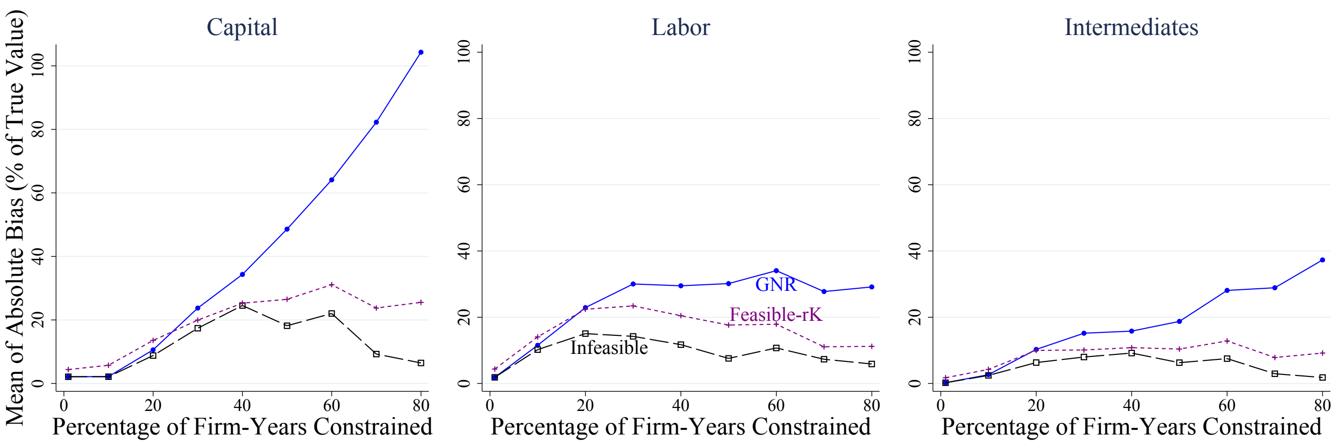


Figure 5 shows the mean absolute error of the infeasible estimator alongside that of Gandhi-Navarro-Rivers in the baseline specification. The error of the two estimators is nearly identical when firms are unconstrained. In these conditions the infeasible estimator uses Gandhi-Navarro-Rivers. But as constraints tighten the estimators diverge. The mean absolute error of Gandhi-Navarro-Rivers rises with the severity of the constraints, whereas that of the infeasible estimator stays at 20 percent or below. The divergence is especially stark for the elasticity of capital, but all three elasticities are estimated more accurately by the infeasible estimator. By switching to the autoregressive estimator when firms are increasingly constrained, the infeasible estimator achieves uniformly lower error.

Is there any feasible approach that can mimic the infeasible estimator by choosing the estimator that is likely to have lower bias? Complementarity implies that it is sufficient to find a statistic shows whether the autoregressive estimator is well-identified (as such implies Gandhi-Navarro-Rivers is poorly identified). Given the link between the autoregressive estimator and the two-stage least squares estimator (see Section 4.2), one approach is to compute a standard weak instruments statistic. Though Proposition 3 is proven in terms of the Cragg-Donald statistic, the Kleibergen-Paap statistic allows for a more general error structure.

Making this estimator feasible requires a consistent estimate of ρ. As noted in Section 4.2, the autoregressive estimate $\hat{\rho}$ is consistent even when its estimate of the production function is weakly identified. In practice, the estimate can be noisy when the estimator is completely non-identified—that is, when firms are unconstrained. The specification test of Section 3 can confirm whether there is evidence of constraints. Finally, I impose one last test: that all of the elasticities must be positive. This key piece of information sets production function estimation apart from most problems of estimation, where testing whether a parameter is positive is often the goal. By contrast, it is generally safe to assume that an increase in any input of production will increase output.

Let $\bar{H}$ be the threshold for the feasible rK statistic, which I set to 4. Then define the feasible estimator as

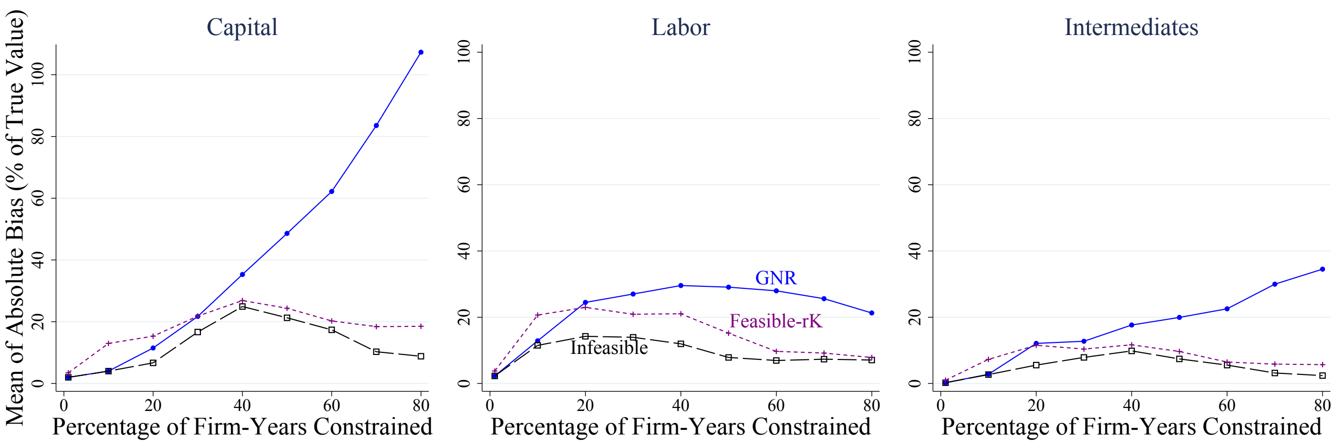
$$\hat{\phi}^X_{Feas} = \begin{cases} \hat{\phi}^X_{AR} & \text{if} \quad \begin{array}{l} \text{1. Feasible rK statistic } > \bar{H}, \\ \text{2. Test rejects at 0.01 level, and} \\ \text{3. } \hat{\phi}^X_{AR} > 0 \end{array} \\ \\ \hat{\phi}^X_{GNR} & \quad \text{otherwise} \end{cases}$$

As Figure 5 shows, this feasible estimator, though crude, achieves much of the improvement made by the infeasible estimator over Gandhi-Navarro-Rivers alone. This suggests the test is a useful guide to the error of each estimator.

What happens when the key assumption of the autoregressive estimator—Autoregressive Productivity—fails? Specification 6 assumes current productivity is a third-order polynomial in lagged productivity, which is the assumption made in Gandhi et al. (2013). I calibrate the parameters of this polynomial to match Industry 311 in their Chilean data (see Appendix B.1). I leave the autoregressive estimation as before, but I estimate Gandhi-Navarro-Rivers assuming productivity is a third-order polynomial (that is, imposing the true functional form). As before, the feasible and infeasible estimators are built from these two estimators.

Figure 6 shows that the performance of the feasible and infeasible estimators is little changed. Both estimators still have a lower error than Gandhi-Navarro-Rivers alone. This is not to say that either the autoregressive method or any estimator that relies on it is robust to all deviations from Autoregressive

Figure 6
Calibrated, Non-Autoregressive Productivity



Productivity. If the Markov process is very nonlinear the autoregressive method will likely do poorly. This simulation shows only that deviations from autoregressive productivity comparable to those found in the data do not derail this approach.

8 Summary

I show that a key identifying assumption behind choice-based methods for estimating production functions is undermined by the presence of market distortions or constraints in the firm's choice of intermediates. I derive a test for the assumption and show that it rejects many manufacturing industries. I then derive the properties of a simple autoregressive estimator. I show that this autoregressive estimator is complementary to the method of Gandhi et al. (2013) when estimating a gross production function. Gandhi-Navarro-Rivers is biased when firms are heavily constrained, while the autoregressive estimator is heavily biased when firms are generally unconstrained. I propose a method for choosing between estimators to get estimates that are valid regardless of whether firms are constrained. Monte Carlo simulations suggest this selection method yields lower error than either Gandhi-Navarro-Rivers or the autoregressive estimator alone.

References

ACKERBERG, D. A., K. CAVES, AND G. FRAZER (2006): "Structural Identification of Production Functions," .

——— (2015): "Identification Properties of Recent Production Function Estimators," .

ARELLANO, M. AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies*, 58, 277–297.

BLUNDELL, R. AND S. BOND (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of econometrics*, 87, 115–143.

GANDHI, A., S. NAVARRO, AND D. RIVERS (2013): "On the Identification of Production Functions: How Heterogeneous Is Productivity?" .

GREENSTREET, D. (2007): "Exploiting Sequential Learning to Estimate Establishment-Level Productivity Dynamics and Decision Rules," *Economics Series Working Papers 345, University of Oxford, Department of Economics*.

KLEIBERGEN, F. AND R. PAAP (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of econometrics*, 133, 97–126.

LEVINSOHN, J. AND A. PETRIN (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of economic studies*, 70, 317–341.

OLLEY, G. AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

SHENOY, A. (2014): "Market Failures and Misallocation," .

STOCK, J. H. AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

STOCK, J. H. AND M. YOGO (2005): "Testing for Weak Instruments in Linear IV Regression," *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80.

WOOLDRIDGE, J. M. (2009): "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters*, 104, 112–114.

WRIGHT, J. H. (2003): "Detecting Lack of Identification in GMM," *Econometric theory*, 19, 322–330.

A Theoretical Appendix

A.1 Proof of Proposition 1

Aside from the assumption that weak identification in the model follows a drifting process, Theorem 1 of Stock and Wright (2000) makes two further assumptions that ensure $\hat{\rho}$ is consistent. The first is that

$$\sqrt{N}\left(\frac{\sum_{n=1}^{N}\left\{\boldsymbol{r}_{t-1}^{(n)}\nu_n(\hat{\boldsymbol{\pi}},\hat{\rho})-\mathbb{E}[\boldsymbol{r}_{t-1}^{(n)}\nu_n(\hat{\boldsymbol{\pi}},\hat{\rho})]\right\}}{N}\right) \tag{31}$$

converges to a Gaussian stochastic process in $(\hat{\boldsymbol{\pi}},\hat{\rho})$.

Let

$$v(\hat{\boldsymbol{\pi}}, \hat{\rho}) = \begin{bmatrix} 1 \\ -\hat{\boldsymbol{\pi}} \\ -\hat{\rho} \\ \hat{\rho}\hat{\boldsymbol{\pi}} \end{bmatrix}$$

Recall that

$$\nu_n(\hat{\boldsymbol{\pi}}, \hat{\rho}) = y_{nt} - \boldsymbol{x}_{nt}\hat{\boldsymbol{\pi}} - \hat{\rho}(y_{n,t-1} - \boldsymbol{x}_{n,t-1}\hat{\boldsymbol{\pi}})$$

which implies (31) can be written as

$$\boldsymbol{\Xi}_N v(\hat{\boldsymbol{\pi}}, \hat{\rho}) \to_D \bar{\boldsymbol{\Xi}} v(\hat{\boldsymbol{\pi}}, \hat{\rho}) \sim N(0, v(\hat{\boldsymbol{\pi}}, \hat{\rho})^T \Sigma^{\Xi} v(\hat{\boldsymbol{\pi}}, \hat{\rho})) \tag{32}$$

where convergence holds by the Cramér-Wold Theorem. This expression is a Gaussian stochastic process in $(\hat{\boldsymbol{\pi}}, \hat{\rho})$.

The second condition is phrased in terms of the decomposition

$$\begin{aligned}
\frac{\sum_{n=1}^N \boldsymbol{r}_{t-1}^{(n)} \nu_n(\hat{\boldsymbol{\pi}}, \hat{\rho})}{N} = {} & \frac{\sum_{n=1}^N \boldsymbol{r}_{t-1}^{(n)} \nu_n(\boldsymbol{\pi}, \rho)}{N} \\
& + \underbrace{\left(\frac{\sum_{n=1}^N \boldsymbol{r}_{t-1}^{(n)} \nu_n(\hat{\boldsymbol{\pi}}, \hat{\rho})}{N} - \frac{\sum_{n=1}^N \boldsymbol{r}_{t-1}^{(n)} \nu_n(\boldsymbol{\pi}, \hat{\rho})}{N} \right)}_{g_{\boldsymbol{\pi}}(\hat{\boldsymbol{\pi}}, \hat{\rho})} \\
& + \underbrace{\left(\frac{\sum_{n=1}^N \boldsymbol{r}_{t-1}^{(n)} \nu_n(\boldsymbol{\pi}, \hat{\rho})}{N} - \frac{\sum_{n=1}^N \boldsymbol{r}_{t-1}^{(n)} \nu_n(\boldsymbol{\pi}, \rho)}{N} \right)}_{g_{\rho}(\hat{\rho})}
\end{aligned}$$

The condition is that $g(\hat{\rho}) \to_p \mathbf{0}$ for $\hat{\rho} = \rho$ and $g(\hat{\rho}) \to_p \bar{g} \neq \mathbf{0}$ otherwise. Note that

$$\begin{aligned}
g(\hat{\rho}) &= \frac{1}{N} \sum_{n=1}^N \boldsymbol{r}_{t-1}^{(n)} \Big(\nu_n(\boldsymbol{\pi}, \hat{\rho}) - \nu_n(\boldsymbol{\pi}, \rho) \Big) \\
&= (\hat{\rho} - \rho) \cdot \frac{1}{N} \sum_{n=1}^N \boldsymbol{r}_{t-1}^{(n)} \Big(y_{n,t-1} - \boldsymbol{x}_{n,t-1} \boldsymbol{\pi} \Big) \\
&= (\hat{\rho} - \rho) \cdot \frac{1}{N} \sum_{n=1}^N \boldsymbol{r}_{t-1}^{(n)} \Big(\omega_{t-1} + \varepsilon_{t-1} \Big) \\
&\to_p (\hat{\rho} - \rho) \Upsilon
\end{aligned}$$

which is zero if and only if $\hat{\rho} = \rho$ because $\Upsilon \neq \mathbf{0}$ by assumption.

Finally, note that in this context the drifting assumption (which is the basis for weak identification asymptotics) is that $g_{\boldsymbol{\pi}}(\hat{\boldsymbol{\pi}}, \hat{\rho})$ can be approximated by

$$\frac{\bar{g}_N^{\boldsymbol{\pi}}(\hat{\boldsymbol{\pi}}, \hat{\rho})}{\sqrt{N}}$$

where $\bar{g}_N^{\boldsymbol{\pi}}(\hat{\boldsymbol{\pi}}, \hat{\rho})$ converges uniformly to a continuous and bounded function $\bar{g}^{\boldsymbol{\pi}}(\hat{\boldsymbol{\pi}}, \hat{\rho})$.

A.2 Proof of Proposition 2

Add $s_t^M = m_t - y_t$ to both sides of Equation 25, let $\bar{\varepsilon} = \log \mathbb{E}[e^{\varepsilon_t}]$, and let ϕ^M denote the elasticity of F with respect to intermediates. The first stage nonparametrically estimates

$$s_t^M = \bar{\varepsilon} + \log \phi^M(k_t, \ell_t, m_t) - \varepsilon_t - \Lambda_t$$

which is consistent for

$$E[s_t^M \mid k_t, \ell_t, m_t] = \bar{\varepsilon} + \log \phi^M(k_t, \ell_t, m_t) - \mathbb{E}[\Lambda_t \mid k_t, \ell_t, m_t]$$

Let $\ddot{\Lambda}_t = \Lambda_t - \mathbb{E}[\Lambda_t \mid k_t, \ell_t, m_t]$. As per the method, the point-wise log elasticity of intermediates is estimated as

$$\widehat{\log \phi^M}(k_t, \ell_t, m_t) = \hat{E}[s_t^M \mid k_t, \ell_t, m_t] - \log\left[\frac{1}{N}\sum \exp(\hat{E}[s_t^M \mid k_t, \ell_t, m_t] - s_t^M)\right]$$

where the summations are over the sample and N is the sample size. This expression is consistent for

$$\begin{aligned} &E[s_t^M \mid k_t, \ell_t, m_t] - \log \mathbb{E}[\exp(E[s_t^M \mid k_t, \ell_t, m_t] - s_t^M)] \\ &= \bar{\varepsilon} + \log \phi^M(k_t, \ell_t, m_t) - \mathbb{E}[\Lambda_t \mid k_t, \ell_t, m_t] - \log \mathbb{E}[e^{\varepsilon_t + \ddot{\Lambda}_t}] \\ &= \log \phi^M(k_t, \ell_t, m_t) - (\mathbb{E}[\Lambda_t \mid k_t, \ell_t, m_t] + \log \mathbb{E}[e^{\ddot{\Lambda}_t}]) \end{aligned}$$

where the last equality follows because ε_t is independent of $\Lambda_t, k_t, \ell_t, m_t$, implying $\mathbb{E}[e^{\varepsilon_t + \ddot{\Lambda}_t}] = \mathbb{E}[e^{\varepsilon_t}]\mathbb{E}[e^{\ddot{\Lambda}_t}]$.

Recall that $\Lambda_t = \log(1 + \lambda_t) > 0$ because λ_t equals the firm's willingness to pay to relax the constraint, which is positive if the firm is constrained and 0 otherwise. Then $\mathbb{E}[\Lambda_t \mid k_t, \ell_t, m_t] > 0$. And by Jensen's Inequality, $\log \mathbb{E}[e^{\ddot{\Lambda}_t}] > \mathbb{E}[\log e^{\ddot{\Lambda}_t}] = 0$ because $\ddot{\Lambda}_t$ has a mean of 0 by construction. Together these imply that

$$\widehat{\log \phi^M}(k_t, \ell_t, m_t) \leq \log \phi^M(k_t, \ell_t, m_t) \qquad \text{for all} \;\; k_t, \ell_t, m_t$$

with equality holding if and only if $\Lambda_t = 0$.

The estimator for the average elasticity is simply

$$\widehat{\log \phi^M} = \frac{1}{N} \sum \widehat{\log \phi^M}(k_t, \ell_t, m_t)$$

implying the bias converges to

$$\widehat{\log \phi^M} - \log \phi^M = -(\mathbb{E}[\Lambda_t] + \log \mathbb{E}[e^{\ddot{\Lambda}_t}])$$

In the special case where $\ddot{\Lambda}_t \sim N(0, \sigma^2_\Lambda)$ the function $e^{\ddot{\Lambda}_t}$ has a log-normal distribution, implying the bias is simply $-(\mathbb{E}[\Lambda_t] + \sigma^2_\Lambda/2)$.

A.3 Relating the Jacobian of the Moment Conditions to the Severity of Constraints

This appendix derives bounds for the smallest singular value of the Jacobian of the moment conditions, and for the smallest eigenvalue of the quadratic form of the Jacobian and the weighting matrix.

Proposition 4 *Assume the sieve approximation for $f(k_t, \ell_t, m_t)$ has p terms, and thus p parameters $\{A^0_{a^0_1,a^0_2,a^0_3}\}$ to estimate. Assume there is a vector of valid instruments $\boldsymbol{r}_{t-1}$ of length $q \geq p$. Assume all variables are demeaned, and that Assumptions 9 and 10 hold.*

Let $\dot{\boldsymbol{x}}_t$ be a $(p-1)$-length vector of the terms of the sieve approximation excluding m_t. Let β be the vector of coefficients from an ordinary least squares regression of the $q \times 1$ vector $\mathbb{E}[\boldsymbol{r}_{t-1}\Delta_\rho(\Lambda)]$ on the $q \times p - 1$ matrix $\mathbb{E}[\boldsymbol{r}_{t-1}\Delta_\rho(\dot{\boldsymbol{x}})^T]$, and let $[SSR]^\Lambda$ denote the sum of squared residuals. Then the smallest singular value of the Jacobian is bounded above by

$$\sqrt{\frac{[SSR]^{\Lambda}}{\beta^T \beta}}$$

Suppose W is the GMM weighting matrix, and J the Jacobian. Let β_W be the coefficients and $[SSR]^{\Lambda}_W$ be the sum of squared residuals from the generalized least squares regression using the weighting matrix W applied to the same regressors and regressand used to estimate β and $[SSR]^{\Lambda}$. Then the smallest eigenvalue of the quadratic form JWJ is bounded above by

$$\frac{[SSR]^{\Lambda}_W}{\beta_W^T \beta_W}$$

The following lemma is useful in proving the proposition:

Lemma 1 *Let $\widetilde{x}$ denote the demeaned transformation of a variable x. Then under the assumptions of Proposition 4,*

$$\begin{aligned}\widetilde{m_t} = \rho \widetilde{m_{t-1}} + \sum_{(a_1^1, a_2^1, a_3^1) \neq (0,0,1)} A'_{a_1^1, a_2^1, a_3^1} \big(\widetilde{k_t^{a_1^1} \ell_t^{a_2^1}} m_t^{a_3^1} - \rho k_{t-1}^{a_1^1} \widetilde{\ell_{t-1}^{a_2^1}} m_{t-1}^{a_3^1} \big) \\ - \frac{1}{-A_{0,0,1}} (\widetilde{\Lambda}_t - \rho \widetilde{\Lambda_{t-1}}) + \frac{1}{-A_{0,0,1}} \widetilde{\eta}_t \end{aligned} \tag{33}$$

Substitute the sieve approximation of $\log F_M(K_t, L_t, M_t)$ into Equation 25:

$$\begin{aligned} \Lambda_t &= \omega_t + \log \mathbb{E}[e^{\varepsilon_t}] + \sum_{a_1^1, a_2^1, a_3^1} A^1_{a_1^1, a_2^1, a_3^1} k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1} \\ &= \omega_t + \log \mathbb{E}[e^{\varepsilon_t}] + \sum_{(a_1^1, a_2^1, a_3^1) \neq (0,0,1)} A^1_{a_1^1, a_2^1, a_3^1} k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1} + A^1_{0,0,1} m_t \\ \Rightarrow m_t &= \frac{1}{-A^1_{0,0,1}} \omega_t + \frac{1}{-A^1_{0,0,1}} \log \mathbb{E}[e^{\varepsilon_t}] + \sum_{(a_1^1, a_2^1, a_3^1) \neq (0,0,1)} \underbrace{\frac{A^1_{a_1^1, a_2^1, a_3^1}}{-A^1_{0,0,1}}}_{A'_{a_1^1, a_2^1, a_3^1}} k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1} - \frac{1}{-A^1_{0,0,1}} \Lambda_t \end{aligned} \tag{34}$$

Apply the Autoregressive Productivity assumption to the definition of ω_t:

$$\tag{35}$$

$$m_t = \frac{1}{-A_{0,0,1}}(\bar{\omega} + \rho\omega_{t-1} + \eta_t) + \frac{1}{-A_{0,0,1}} \log \mathbb{E}[e^{\varepsilon_t}] + \sum_{(a_1^1,a_2^1,a_3^1)\neq(0,0,1)} A'_{a_1^1,a_2^1,a_3^1} k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1} - \frac{1}{-A_{0,0,1}}\Lambda_t \tag{36}$$

Take the lag of (34):

$$\omega_{t-1} = \Lambda_{t-1} - A_{0,0,1} m_{t-1} - \log \mathbb{E}[e^{\varepsilon_t}] - \sum_{(a_1^1,a_2^1,a_3^1)\neq(0,0,1)} A'_{a_1^1,a_2^1,a_3^1} k_{t-1}^{a_1^1} \ell_{t-1}^{a_2^1} m_{t-1}^{a_3^1}$$

Substitute this expression into (36):

$$\begin{aligned} m_t &= \frac{1}{-A_{0,0,1}}(\bar{\omega} + \rho[\Lambda_{t-1} - A_{0,0,1} m_{t-1} - \log \mathbb{E}[e^{\varepsilon_t}] - \sum_{(a_1^1,a_2^1,a_3^1)\neq(0,0,1)} A'_{a_1^1,a_2^1,a_3^1} k_{t-1}^{a_1^1} \ell_{t-1}^{a_2^1} m_{t-1}^{a_3^1}] + \eta_t) \\ &+ \frac{1}{-A_{0,0,1}} \log \mathbb{E}[e^{\varepsilon_t}] + \sum_{(a_1^1,a_2^1,a_3^1)\neq(0,0,1)} A'_{a_1^1,a_2^1,a_3^1} k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1} - \frac{1}{-A_{0,0,1}}\Lambda_t \\ m_t &= \rho m_{t-1} + \sum_{(a_1^1,a_2^1,a_3^1)\neq(0,0,1)} A'_{a_1^1,a_2^1,a_3^1} \big(k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1} - \rho k_{t-1}^{a_1^1} \ell_{t-1}^{a_2^1} m_{t-1}^{a_3^1}\big) \\ &+ \frac{1}{-A_{0,0,1}}\Big[(1-\rho)\log \mathbb{E}[e^{\varepsilon_t}] + \bar{\omega}\Big] - \frac{1}{-A_{0,0,1}}(\Lambda_t - \rho\Lambda_{t-1}) + \frac{1}{-A_{0,0,1}}\eta_t \end{aligned}$$

After demeaning, this expression becomes

$$\begin{aligned} \widetilde{m_t} = \rho\widetilde{m_{t-1}} &+ \sum_{(a_1^1,a_2^1,a_3^1)\neq(0,0,1)} A'_{a_1^1,a_2^1,a_3^1} \big(\widetilde{k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1}} - \rho \widetilde{k_{t-1}^{a_1^1} \ell_{t-1}^{a_2^1} m_{t-1}^{a_3^1}}\big) \\ &- \frac{1}{-A_{0,0,1}}(\widetilde{\Lambda}_t - \rho\widetilde{\Lambda_{t-1}}) + \frac{1}{-A_{0,0,1}}\widetilde{\eta}_t \end{aligned}$$

□

Proof of Proposition 4:

Let
$$\boldsymbol{x}_t^T = \begin{bmatrix} x_{1t} & x_{2t} & \cdots & x_{pt} \end{bmatrix} = \begin{bmatrix} m_t & k_t & \ell_t & m_t^2 & m_t k_t & \cdots \end{bmatrix}$$
denote a vector of the (demeaned) terms in the polynomial used to approximate the production function. Then the sieve approximations can be written as

$$\begin{aligned} \log F(K_t, L_t, M_t) &= \boldsymbol{x}_t^T \boldsymbol{A^0} \\ \log F_M(K_t, L_t, M_t) &= \boldsymbol{x}_t^T \boldsymbol{A^1} \end{aligned} \tag{37}$$

The residual of the estimating equation is

$$\nu_t = y_t - \boldsymbol{x}_t^T \boldsymbol{A^0} - \rho \left(y_{t-1} - \boldsymbol{x}_{t-1}^T \boldsymbol{A^0} \right)$$

and the Jacobian of the moment conditions is the sample analog of

$$J = \mathbb{E}[\boldsymbol{r}_{t-1} \cdot (\nabla_{[\boldsymbol{A^0} \quad \rho]} \nu_t)^T]$$

where
$$(\nabla_{[\boldsymbol{A^0} \quad \rho]} \nu_t)^T = \begin{bmatrix} -[\boldsymbol{x}_t - \rho \boldsymbol{x}_{t-1}]^T & -[y_{t-1} - \boldsymbol{x}_{t-1}^T \boldsymbol{A^0}] \end{bmatrix}$$

Without loss of generality assume $x_{1t} = m_t$ be Let $\dot{\boldsymbol{x}} = [x_{2t}, \cdots, x_{pt}]$ be a vector that contains all the other elements of $\boldsymbol{x}$. By Lemma 1 there is a vector of coefficients $\dot{\boldsymbol{A}}^{\boldsymbol{1}}$ such that

$$\begin{aligned} -A^1_{0,0,1}(m_t - \rho m_{t-1}) &= (\dot{\boldsymbol{x}}_t - \rho \dot{\boldsymbol{x}}_{t-1})^T \dot{\boldsymbol{A}}^{\boldsymbol{1}} - (\Lambda_t - \rho \Lambda_{t-1}) + \eta_t \\ -A^1_{0,0,1}\mathbb{E}[\boldsymbol{r}_{t-1}(m_t - \rho m_{t-1})] &= \mathbb{E}[\boldsymbol{r}_{t-1} \cdot (\dot{\boldsymbol{x}}_t - \rho \dot{\boldsymbol{x}}_{t-1})^T] \dot{\boldsymbol{A}}^{\boldsymbol{1}} - \mathbb{E}[\boldsymbol{r}_{t-1}(\Lambda_t - \rho \Lambda_{t-1})] \end{aligned} \tag{38}$$

where the second line follows because $\boldsymbol{r}_{t-1}$ is a vector of valid instruments and thus uncorrelated with η_t. Let $P_{\dot{\boldsymbol{x}}}$ be the orthogonal projector for the matrix $\mathbb{E}[\boldsymbol{r}_{t-1} \cdot (\dot{\boldsymbol{x}}_t - \rho \dot{\boldsymbol{x}}_{t-1})^T]$. Equation 38 implies that $(I - P_{\dot{\boldsymbol{x}}})(A^1_{0,0,1})\mathbb{E}[\boldsymbol{r}_{t-1}(m_t - \rho m_{t-1})] = (I - P_{\dot{\boldsymbol{x}}})\mathbb{E}[\boldsymbol{r}_{t-1}(\Lambda_t - \rho\Lambda_{t-1})]$. Let $\beta_{\dot{\boldsymbol{x}}} = [\beta_{x_2}, \cdots, \beta_{x_N}]^T$ be the coefficient vector associated with the ordinary least squares regression of $\mathbb{E}[\boldsymbol{r}_{t-1}(\Lambda_t - \rho\Lambda_{t-1})]$ on $\mathbb{E}[\boldsymbol{r}_{t-1} \cdot (\dot{\boldsymbol{x}}_t - \rho \dot{\boldsymbol{x}}_{t-1})^T]$, and define

$$v_0 = \begin{bmatrix} A^1_{0,0,1} & (\dot{\boldsymbol{A}}^{\boldsymbol{1}} - \beta_{\dot{\boldsymbol{x}}})^T & 0 \end{bmatrix} \tag{39}$$

where the final 0 is in the same position in the vector of coefficients $[\boldsymbol{A}^{\boldsymbol{0}} \quad \rho]$ being estimated as the autoregressive coefficient ρ.

The smallest singular value of the Jacobian is equal to

$$minSV(J) = \sqrt{\min\left\{mineig(J^TJ), mineig(JJ^T)\right\}} \leq \sqrt{mineig(J^TJ)}$$

where $mineig$ denotes the smallest eigenvalue. Since J^TJ is a quadratic form (and thus positive semidefinite),

$$\begin{aligned} mineig(J^TJ) &= \min_v \frac{v^TJ^TJv}{v^Tv} \\ &\leq \frac{v_0^TJ^TJv_0}{v_0^Tv_0} \\ &= \frac{\left\{(I - P_{\dot{\boldsymbol{x}}})\mathbb{E}[\boldsymbol{r}_{t-1}(\Lambda_t - \rho\Lambda_{t-1})]\right\}^T \left\{(I - P_{\dot{\boldsymbol{x}}})\mathbb{E}[\boldsymbol{r}_{t-1}(\Lambda_t - \rho\Lambda_{t-1})]\right\}}{(A^1_{0,0,1})^2 + \sum \beta^2_{x_a}} \end{aligned} \tag{40}$$

The numerator of (40) is simply the residual sum of squares from a regression of $\mathbb{E}[\boldsymbol{r}_{t-1}(\Lambda_t - \rho\Lambda_{t-1})]$ on the columns of $\mathbb{E}[\boldsymbol{r}_{t-1} \cdot (\dot{\boldsymbol{x}}_t - \rho \dot{\boldsymbol{x}}_{t-1})^T]$.

A similar argument shows that the smallest eigenvalue of the weighted Jacobian JWJ is bounded above by the rescaled residual of a generalized least squares regression with weighting matrix W^{-1}.

A.4 Proof of Proposition 3

Let $P_{\boldsymbol{R}_{t-1}}$ be the projection matrix for the instruments. By an argument similar to that used in Proposition 4

$$\begin{aligned}-A^1_{0,0,1}P_{\boldsymbol{R}_{t-1}}\Delta_\rho(m) &= P_{\boldsymbol{R}_{t-1}}\Delta_\rho(\dot{\boldsymbol{x}})^T\dot{\boldsymbol{A}}^{\mathbf{1}} - P_{\boldsymbol{R}_{t-1}}\Delta_\rho(\Lambda) + P_{\boldsymbol{R}_{t-1}}\eta_t \\ &\to \Delta_\rho(\widehat{\dot{\boldsymbol{x}}})^T\dot{\boldsymbol{A}}^{\mathbf{1}} - \Delta_\rho(\widehat{\Lambda})\end{aligned}$$

where the last line follows because $P_{\boldsymbol{R}_{t-1}}\eta_t$ is asymptotically 0 by the validity of the instruments. Then

$$-A^1_{0,0,1}P_{\Delta_\rho(\widehat{\dot{\boldsymbol{x}}})}\Delta_\rho(\widehat{m}) = \Delta_\rho(\widehat{\dot{\boldsymbol{x}}})(\dot{\boldsymbol{A}}^{\mathbf{1}} - \beta) \tag{41}$$

$$-A^1_{0,0,1}(I - P_{\Delta_\rho(\widehat{\dot{\boldsymbol{x}}})})\Delta_\rho(\widehat{m}) = -(I - P_{\Delta_\rho(\widehat{\dot{\boldsymbol{x}}})})\Delta_\rho(\widehat{\Lambda}) \tag{42}$$

Since the Cragg-Donald statistic G is positive definite, its smallest eigenvalue can be written as

$$mineig(G) = \min_v \frac{v^T\Sigma_{\boldsymbol{V}}^{-1/2}\hat{\boldsymbol{X}}^T\hat{\boldsymbol{X}}\Sigma_{\boldsymbol{V}}^{-1/2}v}{v^Tv} \tag{43}$$

Since $\Sigma_{\boldsymbol{V}}$, and thus $\Sigma_{\boldsymbol{V}}^{-1/2}$, is positive definite, we can re-parameterize the minimization in terms of $q = \Sigma_{\boldsymbol{V}}^{-1/2}v$.

$$mineig(G) = \min_q \frac{q^T\hat{\boldsymbol{X}}^T\hat{\boldsymbol{X}}q}{q^T\Sigma_{\boldsymbol{V}}q} \tag{44}$$

Let $q_0 = [A^1_{0,0,1}, (\dot{\boldsymbol{A}}^{\mathbf{1}} - \beta)^T]^T$. By (41) and (42),

$$\begin{aligned}\min_q \frac{q^T\hat{\boldsymbol{X}}^T\hat{\boldsymbol{X}}q}{q^T\Sigma_{\boldsymbol{V}}q} &\le \frac{q_0^T\hat{\boldsymbol{X}}^T\hat{\boldsymbol{X}}q_0}{q_0^T\Sigma_{\boldsymbol{V}}q_0} \\ &= \frac{[(I - P_{\Delta_\rho(\widehat{\dot{\boldsymbol{x}}})})\Delta_\rho(\widehat{\Lambda})]^T[(I - P_{\Delta_\rho(\widehat{\dot{\boldsymbol{x}}})})\Delta_\rho(\widehat{\Lambda})]}{q_0^T\Sigma_{\boldsymbol{V}}q_0}\end{aligned} \tag{45}$$

The numerator is the sum of squared residuals from the regression of $\Delta_\rho(\widehat{\Lambda})$ on

Table 1
Parameters: Stochastic Processes

Capital		Labor	
α_0^k	-0.01	α_0^ℓ	0.19
α_1^k	1.00	α_1^ℓ	0.95
α_2^k	0.04	α_2^ℓ	0.47
α_3^k	0.20	α_3^k	0.18

Productivity		Wealth	
ρ	0.77	α_0^S	1.94
ρ_0	86.31	α_1^S	0.49
ρ_1	-33.36	α_2^S	0.32
ρ_2	4.57		
ρ_3	-0.20		
$\bar{\omega}$	7.08		

$\Delta_\rho(\hat{\dot{\boldsymbol{x}}})$.

$\square$

B Simulation Appendix

B.1 Calibration

Using the same code as Gandhi et al. (2013) I reconstruct their estimates of productivity ω_t. I regress ω_t on a third-order polynomial in ω_{t-1}, which is the functional form assumed for Ψ in Gandhi et al. (2013). I take the residual as η_t. I then regress y_t on a third-order polynomial of k_t, ℓ_t, m_t as well as on ω_t, and take the residual from this equation as ε_t.

Table 1 shows the parameters of the stochastic processes. I regress ω_t on just ω_{t-1} to estimate ρ. I take the estimates from the earlier third-order polynomial as $\rho_0, \rho_1, \rho_2, \rho_3$. I take the mean of ω_{t-1} as $\bar{\omega}$. I then regress k_t on k_{t-1}, η_{t-1}, and η_{t-1} to estimate the parameters of the capital process. I do the same for the labor process. Finally, I regress the log of short-term assets on its first and second lag to estimate the parameters of the observed wealth process.

Table 2 shows the standard deviations of the three shocks in the simulation.

Table 2
Parameters: Shocks

$SD(\eta)$	0.09
$SD(\varepsilon)$	0.25
$SD(v^S) = SD(v^X)$	1.35

Table 3
Parameters: Production Function

	Spec. 1	Spec. 2	Spec. 3
θ^K	0.015	0.064	0.11
θ^L	0.95	0.33	0.28
θ^M	0.67	0.67	0.67
σ	0.39	0.39	1
ϵ	5	1.25	1

I set each to equal the standard deviation of the shock estimated above, where I take v^S as the residual from the regression of short-term assets on its two lags.

Table 3 shows the parameters of the three production functions (basline, almost Cobb-Douglas, and Cobb-Douglas). For each elasticity ϵ I chose the other parameters to produce average elasticities that match the elasticities Gandhi et al. (2013) estimate for Industry 311 in Chile.

The additional specifications generated for Figure **??** vary the production function as follows: let $u \sim U[-\frac{1}{2}, \frac{1}{2}]$ be a random variable drawn independently across simulations. Then

$$\sigma = 0.39 + 0.4u$$
$$\theta^K = 0.015 - 0.04(u - .4)$$
$$\theta^L = 0.95 + u + .1$$

The "Basic" scenario varies only the production function. The other specifications also vary one other parameter, as summarized in Table 4:

Table 4
Parameters Varied to Produce Figure **??**

	[+]	[-]
SD(Innovation)	$SD(\eta_t) = .185$	$SD(\eta_t) = .046$
SD(Shock)	$SD(\varepsilon_t) = .51$	$SD(\varepsilon_t) = .13$
AR Coef. on Prod	$\rho = .92$	$\rho = .62$
AR(1) Coef. on S	$\alpha_1^S = .69$	$\alpha_1^S = .29$
AR(2) Coef. on S	$\alpha_2^S = .42$	$\alpha_2^S = .25$

B.2 Estimating the Value-Added Production Function when Firms are Constrained

Suppose the researcher believes the true production function takes a value-added form, meaning intermediate inputs are used in production but do not appear in the production function. Then

$$Y = e^{\omega+\varepsilon}\tilde{F}(K, L)$$

Estimating the autoregressive method in this case is easy, as intermediates are excluded from the estimation entirely. Simply demean the log of output, capital, and labor, form the residual

$$\begin{aligned} &y_{it} - \tau_k k_{it} - \tau_\ell \ell_{it} - \tau_{kk} k_{it}^2 - \tau_{\ell\ell}\ell_{it}^2 - \tau_{k\ell} k_{it}\ell_{it} \\ &\quad - \rho(y_{i,t-1} - \tau_k k_{i,t-1} - \tau_\ell \ell_{i,t-1} - \tau_{kk} k_{i,t-1}^2 - \tau_{\ell\ell}\ell_{i,t-1}^2 - \tau_{k\ell} k_{i,t-1}\ell_{i,t-1}) \end{aligned} \tag{46}$$

and estimate the coefficients by generalized method of moments using a constant and lags of $k, \ell, k^2, \ell^2, k\ell$ as instruments. (I find that also using the second lag of capital as an instrument increases the precision.)

Since the flexible input m does not appear in the residual (46), the autoregressive method sidesteps the problem of under-identification. As I show below it works regardless of whether firms are constrained. This is a major benefit of the value-added approach. Whether the value-added approach is valid, however, depends on the production environment.[17]

[17]Ackerberg et al. (2015) assume there is a "structural value-added production function" in

For the simulations that follow I assume the firm sets the log of m equal to a second-order polynomial in the logs of capital, labor, and known productivity. I calibrate the coefficients of the polynomial to match industry 311 in the Chilean data (see Appendix B for details). I adjust the paramters of the production function to give elasticities of capital and labor to match the value-added estimates of Gandhi et al. (2013). All other parameters are left as in the main text.

Figure 7 compares the error of Ackerberg-Caves-Frazer to that of the autoregressive method. I show only the autoregressive method rather than the feasible and infeasible estimators because the main weakness of the autoregressive method—the problem of under-identification—is absent when estimating a value-added production function. This is clear in the figure. The error of the autoregressive method remains low regardless of how many firms are constrained.

By contrast, Ackerberg-Caves-Frazer only has low error when firms are unconstrained. When firms are constrained the estimates are inaccurate. The problem is compounded when productivity is nonlinear. As in the main text, I assume that the researcher using Acerberg-Caves-Frazer knows the true functional form of the Markov process for productivity. But though there is no specification error in the nonlinearity, its presence aggravate the failure of the Scalar Unobservable assumption. The autoregressive estimator does not have this problem. Though it imposes the incorrect assumption of a linear Markov process, it does not impose Scalar Unobservability, which is the more misleading assumption when over half of firms are constrained.

---

which output is

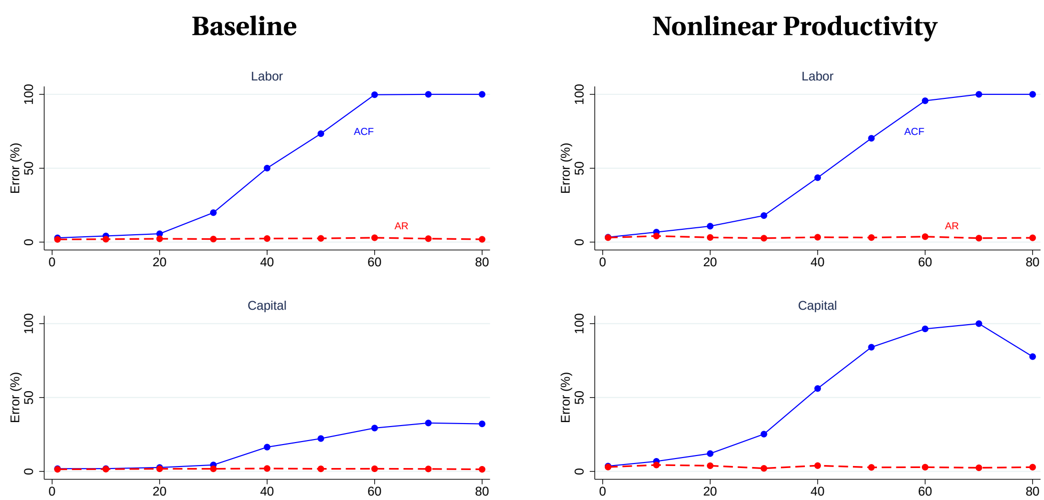
$$Y = e^{\varepsilon} \min\{e^{\omega}\tilde{F}(K,L), \theta^M M\}. \tag{47}$$

Then the unconstrained firm sets $\theta^M M = e^{\omega}\tilde{F}(K,L)$. The Scalar Unobservable assumption is satisfied, and realized output is not a function of M.

Though theoretically appealing the structural value-added production function has one problem: it is impossible to estimate if firms are constrained. A constrained firm sets $M = Z < e^{\omega}\tilde{F}(K,L)$, meaning $Y = e^{\varepsilon}\theta^M Z$. The firm has spare capacity; two firms with different levels of labor and capital have identical output. Since output is no longer a function of capital and labor it is uninformative about the value-added part of production.

The test for the Scalar Unobservable assumption will work nevertheless, letting the researcher know if firms might be constrained.

Figure 7
The Autoregressive Method and Ackerberg-Caves-Frazer



C Miscellaneous Results

C.1 Ackerberg-Caves-Frazer is Biased in the Presence of Constraints

As in the case of Gandhi-Navarro-Rivers, to say something concrete about the bias requires assuming something about how intermediates are chosen in the presence of constraints. To be precise, assume

Assumption 11 (Constrained Unobservable) *The choice of intermediate inputs is* $M_t = \bar{M}(K_t, L_t, \omega_t, \Lambda_t)$ *for some smooth function* $\bar{M}(\cdot)$*, with* $M_t^* = \bar{M}(K_t, L_t, \omega_t, 0)$.

Assumption 12 (Constrained Monotonicity) *The derivatives satisfy* $\bar{M}_\omega > 0$ *and* $\bar{M}_\Lambda < 0$

These functions can be written in logs $m_t = \bar{m}(k_t, \ell_t, \omega_t, \Lambda_t)$.

The easiest way to characterize the bias of Ackerberg-Caves-Frazer (and proxy methods more generally) is to think of constraints as inducing non-classical "measurement error" of sorts in the choice of intermediates. The econometrician would like to observe m_t^* because this optimal or desired choice is a function of only a scalar unobservable. But m_t, the constrained optimal choice, is all that is available. The difference between them is

$$\begin{aligned} m_t^* - m_t &= \bar{m}(k_t, \ell_t, \omega_t, 0) - m_t \\ &= \bar{m}(k_t, \ell_t, \bar{m}^{-1}(k_t, \ell_t, m_t, \Lambda_t), 0) - m_t \\ &= \mu(k_t, \ell_t, m_t, \Lambda_t) \end{aligned}$$

where $\mu(k_t, \ell_t, m_t, 0) = 0$ and $\mu_4 > 0$.

From Equation 8 the first-stage procedure is to get a nonparametric estimate of

$$\begin{aligned} y_t &= \tilde{f}(k_t, \ell_t) + \bar{m}^{-1}(k_t, \ell_t, m_t^*) + \varepsilon_t \\ &= \Phi(k_t, \ell_t, m_t^*) + \varepsilon_t \\ &= \Phi(k_t, \ell_t, m_t) + \underbrace{\Big[\Phi(k_t, \ell_t, m_t + \mu(k_t, \ell_t, m_t, \Lambda_t)) - \Phi(k_t, \ell_t, m_t)\Big]}_{\Phi_t^\Delta = \Phi^\Delta(k_t, \ell_t, m_t, \Lambda_t)} + \varepsilon_t \end{aligned} \tag{48}$$

which is consistent for $\mathbb{E}[y_t \mid k_t, \ell_t, m_t] = \Phi(k_t, \ell_t, m_t) + \mathbb{E}[\Phi_t^\Delta \mid k_t, \ell_t, m_t]$.

Define

$$\begin{aligned} \Psi_t^\Delta &= \Psi^\Delta(k_t, \ell_t, m_t, \Lambda_t) \\ &= \Psi\big(\hat{\mathbb{E}}[y_t \mid k_t, \ell_t, m_t + \mu(k_t, \ell_t, m_t, \Lambda_t)] - \tilde{f}(k_t, \ell_t)\big) - \Psi\big(\hat{\mathbb{E}}[y_t \mid k_t, \ell_t, m_t] - \tilde{f}(k_t, \ell_t)\big) \end{aligned} \tag{49}$$

The second stage then estimates

$$\begin{aligned} y_t - \tilde{f}(k_t, \ell_t) &= \Psi\big(\hat{\mathbb{E}}[y_{t-1} \mid k_{t-1}, \ell_{t-1}, m_{t-1}^*] - \tilde{f}(k_{t-1}, \ell_{t-1})\big) + \eta_t + \varepsilon_t \\ &= \Psi\big(\hat{\mathbb{E}}[y_{t-1} \mid k_{t-1}, \ell_{t-1}, m_{t-1}] - \tilde{f}(k_{t-1}, \ell_{t-1})\big) + \Psi_{t-1}^\Delta + \eta_t + \varepsilon_t \end{aligned} \tag{50}$$

At the true $\tilde{f}$ the moment condition equals

$$\mathbb{E}[(\Psi_{t-1}^\Delta + \eta_t + \varepsilon_t)\boldsymbol{r}_{t-1}] = \mathbb{E}[\Psi_{t-1}^\Delta \boldsymbol{r}_{t-1}]$$

which in general does not equal 0. In the neighborhood of $\Lambda_{t-1} = 0$ is approximately equal to

$$\begin{aligned}
&\mathbb{E}[\Psi'[\bar{m}^{-1}(k_{t-1}, \ell_{t-1}, m_{t-1}, 0)]\Phi_3(k_{t-1}, \ell_{t-1}, m_{t-1})\mu_4(k_{t-1}, \ell_{t-1}, m_{t-1}, 0)\Lambda_{t-1}\boldsymbol{r}_{t-1}] \\
&= \mathbb{E}\Big[\Psi'[\omega_{t-1}]\bar{m}_3^{-1}(k_{t-1}, \ell_{t-1}, m_{t-1}, 0)\mu_4(k_{t-1}, \ell_{t-1}, m_{t-1}, 0)\boldsymbol{r}_{t-1}\mathbb{E}[\Lambda_{t-1} \mid \boldsymbol{r}_{t-1}]\Big]
\end{aligned}$$

which is nonzero as long as $\mathbb{E}[\Lambda_{t-1} \mid \boldsymbol{r}_{t-1}] \neq 0$.

C.2 Allowing for the Price of Intermediates

In the main text I assume the production function is a function of "real expenditures" of intermediates. If the production function is instead a function of "levels" of intermediates, meaning expenditures divided by a price P_t^M, it is straightforward to adapt the test for constraints.

Under the assumption of Constrained Optimal Choices no adjustment is necessary, though the Optimal Choice assumption must be modified to

Assumption 13 (Optimal Choices, Modified) *Firms choose M_t to satisfy*

$$P_t^M = \mathbb{E}[e^{\varepsilon_t}]e^{\omega_t}F_M(K_t, L_t, M_t) \tag{51}$$

which implies the Constrained Optimal Choice assumption becomes

Assumption 14 (Constrained Optimal Choices, Modified) *Define $\Lambda_t = \log(1 + \lambda_t/P_t^M) = \log(1 + \tilde{\lambda}_t)$, where $\tilde{\lambda}_t$ is a Lagrange multiplier that gives the shadow cost to the firm (in units of the intermediate good) of being unable to choose M_t optimally. Let p_t^M be the log of the price of intermediates. Then the choice of the firm satisfies*

$$p_t^M + \Lambda_t = \omega_t + \log \mathbb{E}[e^{\varepsilon_t}] + \log F_M(K_t, L_t, M_t) \tag{52}$$

Adding $m_t - y_t$ to both sides and defining the cost share as $s_t^M = \log(P_t^M M_t/Y_t)$ yields an expression identical to that proposed in the main text.

Suppose there is reason to doubt that the firm makes optimal choices but it is still plausible that the Scalar Unobservable assumption holds. This assumption must now be modified, as it is hard to imagine the price of intermediates would not affect the firm's choice of intermediates.[18]

[18] Otherwise the test given in the text can be run without modification.

Assumption 15 (Scalar Unobservable, Modified) *The choice of intermediate inputs is* $M_t = \bar{M}(K_t, L_t, \omega_t, P_t^M)$ *for some smooth function* $\bar{M}(\cdot)$.

This modified assumption implies the new testing equation would be

$$s_t^M = \bar{\xi}(k_t, \ell_t, m_t, p_t^M) + \boldsymbol{r}_{t-1}\boldsymbol{\varrho} + e_t \tag{53}$$

This modification is only necessary if the researcher is unwilling to make Assumption 14 but is willing to make Assumption 15. Since optimal choices are necessary for Gandhi-Navarro-Rivers, I effectively take Assumption 14 as given in the empirical application.

Finally, Property **??** of Proposition **??** will still hold as long as there is no systematic variation in the price of intermediates across firms. If such variation exists then the autoregressive estimator may be identified even if firms are unconstrained as long as the instruments are informative about the price of intermediates. In this case, the price of intermediates might directly serve as an instrument for the choice of intermediates. But the literature often deems it implausible that there is exogenous variation in the price of intermediates across firms, as any aggregate variation would likely be correlated with productivity. Idiosyncratic variation, if it exists, is rarely observed and immediately implies that the Scalar Unobservable assumption fails.