Estimating the Production Function when Firms Are Constrained

Ajay Shenoy*

April 18, 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, as is common in developing countries or among small firms in developed countries. The test rejects unconstrained choices in many countries and industries. I show that when firms are constrained a simple autoregressive estimator becomes viable. Simulations suggest using the test to choose between choice-based and autoregressive estimators 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 constrained. 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.

Such constraints 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. It is critical that a researcher be able to test for whether such constraints are present.

This paper's first contribution is to derive 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, 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 three different contexts: manufacturing firms in Chile and Colombia, and farmers in Thailand. The test rejects the Scalar Unobservable assumption in many manufacturing industries as well as among Thai farmers. The result suggests the Scalar Unobservable assumption cannot be taken for granted; its validity must be tested.

This paper's second contribution is to show that a simplified dynamic panel estimator may be used when firms are constrained, and that this estimator performs well *only* when firms are constrained. This estimator assumes that productivity is autoregressive (rather than an arbitrary Markov process), but does not assume a Scalar Unobservable. I show that the method is underidentified when firms are unconstrained. The result follows because unconstrained choices of intermediates have no systematic independent variation beyond that contained in capital and labor, making it impossible to identify the contribution of intermediates to output. Constraints, however, will induce

such variation, ensuring the estimator is identified.

I run simulations to show that as the severity of constraints increases the performance of the autoregressive method improves even as that of the choice-based methods deteriorates. This complementarity suggests the researcher should use autoregressive methods when firms are constrained and choice-based methods when they are not. The challenge, however, is to identify whether in a given setting the constraints are too severe for choice-based methods.

The paper's third contribution is to show that the test for constraints can help the researcher decide which method is less biased. I show that under the assumption that each firm's choice is constrained-optimal, the coefficients estimated by the test give the partial correlation of a set of instruments (say, those used for the autoregressive estimator) with the Lagrange multiplier on the constraint. I show that when this multiplier is non-zero the choice-based methods may be inconsistent, whereas when it is zero the autoregressive estimators are under-identified. This complementarity in the two methods is what makes the test useful. By linking inconsistency (or consistency) of choice-based methods to the strong (or under-) identification of autoregressive methods, it helps the researcher adopt whichever approach is most suited to any situation.

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]

[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 D.1 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 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*

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

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 2 (Instruments for Capital and Labor) *There are valid instruments that are informative about K_t and L_t but chosen at $t-1$ or earlier.*

In their empirical application, Gandhi et al. (2013) assume that both capital and labor are pre-determined at $t-1$, meaning they are instruments for themselves. This assumption is unnecessary for their method as long as capital and labor are dynamic, 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 4.1, this timing is crucial for the choice-based methods:

Assumption 3 (Timing) *M_t is chosen at the beginning of period t 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 4.1 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 4 (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 by M and the sample average of $\hat{\varepsilon}_t$ 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 $y_t = \log Y_t$. Since

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

[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 D.1 that under the assumptions made by Gandhi et al. (2013), the main results require no modification.

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 5 (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 4 and instead assume:

Assumption 6 (Monotonicity) *All else equal, the choice of intermediates M_t is strictly increasing in ω_t.*

Assumption 7 (Scalar Unobservable) *The choice of intermediate inputs is $M_t = \bar{M}(K_t, L_t, \omega_t)$ for some smooth function $\bar{M}(\cdot)$.*

Assuming Optimal Choices (and the earlier assumptions) 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 $\omega_t = \bar{M}^{-1}(K_t, L_t, M_t)$. After controlling for capital and labor, intermediates are a proxy for productivity. Value-added output can be written

$$\begin{aligned} y_t &= \log \tilde{F}(K_t, L_t) + \bar{M}^{-1}(K_t, L_t, M_t) + \varepsilon_t \\ &= \Phi(K_t, L_t, M_t) + \varepsilon_t \end{aligned} \tag{8}$$

The term $\Phi(K_t, L_t, M_t)$ can be estimated nonparametrically, giving a consistent estimate of $\log \tilde{F}(K_t, L_t) + \omega_t$. By the Markov Productivity assumption,

$$\hat{\Phi}(K_t, L_t, M_t) - \log \tilde{F}(K_t, L_t) - \Psi\left(\hat{\Phi}(K_{t-1}, L_{t-1}, M_{t-1}) - \log \tilde{F}(K_{t-1}, L_{t-1})\right) = \eta_t \tag{9}$$

which is uncorrelated with $(K_t, L_t, M_{t-1}, K_{t-1}, L_{t-1}, \ldots)$. 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 :

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

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

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

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

It is immediately clear that Assumption 4 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

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

which depends on more than one unobservable: productivity ω_t and one or more terms $\{S_t^j\}$. Assumption 7, the Scalar Unobservable assumption, also fails.

Though throughout the paper I assume the suboptimal choice arises from a credit constraint, the argument made here holds for any unobserved feature of the firm or the economy that gives some firms easier access to intermediates than others. 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.

3 A Test for Constraints

3.1 Approach

Using arguments similar to those used to derive Equation 8, it is easy to see that Assumptions 3, 6, and 7 imply that there exists a function $\xi(k_t, \ell_t, m_t)$ such that

$$y_t = \xi(k_t, \ell_t, m_t) + \varepsilon_t \tag{14}$$

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 (14) by -1 and add m_t to both sides:

$$s_t^M = \bar{\xi}(k_t, \ell_t, m_t) - \varepsilon_t \tag{15}$$

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

Equation 15 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 su_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 3 and 6 hold, one simple test of the Scalar Unobservable assumption is to estimate the semiparametric regression

$$s_t^M = \bar{\xi}(k_t, \ell_t, m_t) + \boldsymbol{r}_{t-1}\boldsymbol{\varrho} + e_t \tag{16}$$

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 Unobservatble assumption fails. The consequences of its failure are especially clear for the method of Gandhi et al. (2013), as this equation is similar to the log share regression it uses to estimate the elasticity of intermediates. But the consequences for the method of Ackerberg et al. (2015) and other proxy methods are also clear, as Equation 16 was derived from Equation 8, which is the first stage of a proxy estimator. In all cases a key assumption—and thus a key step in the estimation—fails.

3.2 Evidence of Constraints

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). I also run the test on rice farmers in Thailand, whose "firms" may be more likely to be constrained than the formal sector manufacturers.

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.[6]

Thai Rice Farming: I use the Townsend Thai annual panel of households from 1997 to 2009. I construct the factors of production—land, labor, and capital—exactly as done in Shenoy (2014) except that I keep expenditure on intermediate inputs (seeds, fertilizer, etc.) separate from capital. Capital is defined as what remains: the value of structures, machinery, and vehicles used in production.

Figure 1 shows the F-statistic, p-value, and partial R-squared of the test in the two datasets of manufacturing firms.[7] The test rejects at the 5 percent level 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 5, this invariance to sample size makes the partial R-squared a better gauge of the level of constraints.

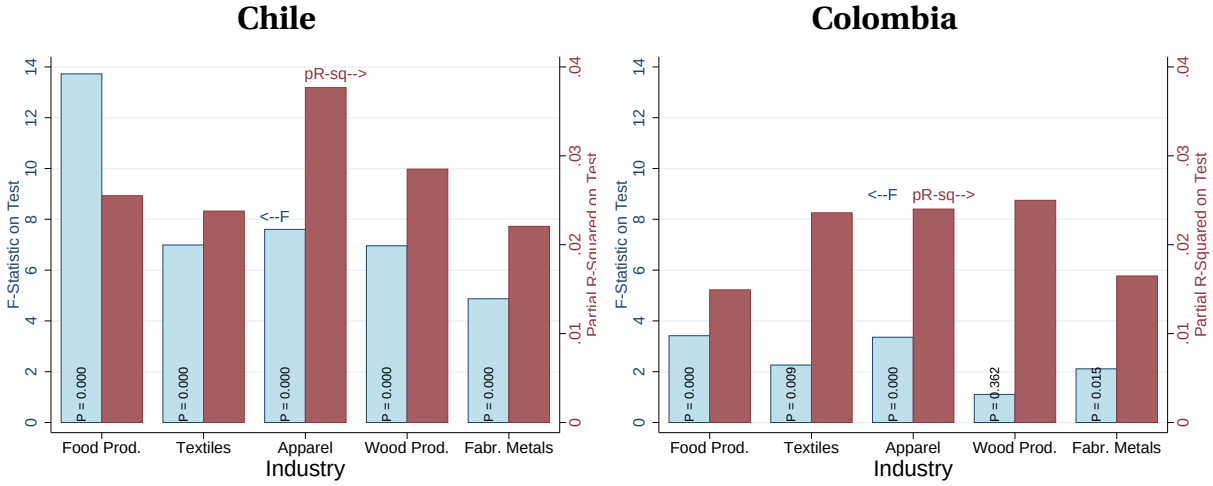
As the test rejects in most of these industries, the Scalar Unobservable assumption should not be taken for granted. But in some industries the test either does not reject (Wood Products in Colombia) or the partial R-squared is relatively small (food products in Colombia). Taken together these two results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

Figure 1 suggests the outcome of the test depends heavily on the economic environment. In an environment in which more firms are constrained, the partial R-squared should be larger. I check whether this pattern holds by running the test in the sample of Thai rice farmers. Many of these farmers are poor; their

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

[7] 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 16, 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. The Indian sample is restricted to large plants, which are the only plants included in the panel for three years in a row.

operations are typically small, with an average annual revenue of roughly $1200 (see Shenoy, 2014). It is a context in which a lack of credit and other market imperfections are likely to constrain choices.

I assume real rice revenue is a function of land, labor, capital, and intermediates. I approximate $\bar{\xi}$ using a log sieve approximation of these inputs. As it is not entirely clear in this context what the set of instruments $\boldsymbol{r}_{t-1}$ should include, I run the test using three different sets of instruments. All three sets include the first-order lags of the terms used in the log sieve approximation and the second-order lag of intermediates. They differ only in which other second-order terms they use. The first set uses only land and its interaction with intermediates; the second only capital and its interaction with intermediates; the third uses all of these terms and the interaction of land and capital.

Table 1 reports the results. The F-statistic is largest for the set instruments using second-order lags of land, likely because land is more likely than capital to be offered as collateral for loans. This F-statistic is over 8, larger than any industry in the manufacturing censuses except for Chilean food production (where the sample is nearly three times as large). The partial R-squared in

Table 1
Applying the Test to Thai Rice Farmers

	F-Statistic	Partial R-Squared	Observations	Households
Second-Order Land Instruments	8.34	0.041	4076	666
Second-Order Capital Instruments	6.99	0.032	4125	670
All Instruments	7.01	0.040	3989	648

Note: The test is applied assuming land, labor, capital, and intermediates are the factors of production. The specification includes year dummies. Inference is clustered by household.

this specification is over 0.04, larger than in any of the manufacturing industries. The result is intuitive. Poor farmers are likely to face heavier constraints than formal manufacturing firms.

4 Estimating the Production Function when Firms are Constrained

4.1 An 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 8 (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.[8] 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 A.2.

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

4.1.1 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

$$\begin{aligned} y_t &= f(k_t, \ell_t, m_t) + \omega_t + \varepsilon_t \\ &= f(k_t, \ell_t, m_t) + \bar{\omega} + \rho\omega_{t-1} + \eta_t + \varepsilon_t \\ &= f(k_t, \ell_t, m_t) + \bar{\omega} + \rho(y_{t-1} - f(k_{t-1}, \ell_{t-1}, m_{t-1})) + \underbrace{\eta_t - \rho\varepsilon_{t-1} + \varepsilon_t}_{\nu_t} \end{aligned} \tag{17}$$

Equation (17) 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 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.[9] 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.[10]

4.1.2 Identification

The autoregressive estimator comes with one ironic caveat: though the exclusion restriction holds regardless of whether the Scalar Unobservable assumption fails, the rank condition essentially requires it to fail. It is easiest to see

[9] 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 3, the Timing Assumption, favors Gandhi-Navarro-Rivers over the autoregressive method.

[10] That is, let

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

the problem 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

$$m_t = \frac{1}{1-\pi_m}\left[\omega_t + \pi_k k_t + \pi_\ell \ell_t\right] \tag{18}$$

Let $\boldsymbol{r}_{t-1} = [r_{t-1}^1, \ldots, r_{t-1}^R]$ 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 $\mathbb{E}[\nu_t r_{t-1}^1] = \cdots = \mathbb{E}[\nu_t r_{t-1}^R] = 0$ 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.

A variation of this argument applies to any non-parametrically estimated production function. The intuition is similar to the functionl dependence critique raised in Ackerberg et al. (2015). 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. That is the insight behind the reduced-form test of constraints derived in Section 3. In the next section I exploit this logic to derive a criterion that can be used to select between choice-based and autoregressive methods.

4.2 Complementarity: A Structural Interpretation of the Test

The success or failure of choice-based and autoregressive estimators hinges on how badly firms are constrained. The severity of constraints in part governs which of the two methods is less biased. A rejection by the test of Section 3 suggests firms are constrained. But the F-statistic and the partial R-squared do not necessarily have any structural interpretation, and do not necessarily capture the complementarity between the two estimators, without additional assumptions.

Assume that productivity is autoregressive (Assumption 8) and that Equation 12 holds. This latter assumption is formalized in

Assumption 9 (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*

$$\Lambda_t = \omega_t + \log \mathbb{E}[e^{\varepsilon_t}] + \log F_M(K_t, L_t, M_t) \tag{19}$$

Also 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:

$$\begin{aligned} \log F(K_t, L_t, M_t) &= \sum_{a_1^0, a_2^0, a_3^0} A^0_{a_1^0, a_2^0, a_3^0} k_t^{a_1^0} \ell_t^{a_2^0} m_t^{a_3^0} \\ \log F_M(K_t, L_t, M_t) &= \sum_{a_1^1, a_2^1, a_3^1} A^1_{a_1^1, a_2^1, a_3^1} k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1} \end{aligned} \tag{20}$$

Let $\boldsymbol{r}_{t-1}$ be the vector of instruments used for the autoregressive estimator, and recall that s_t^M is the cost share of intermediates in output. After de-meaning all variables, estimate

$$s_t^M = \sum_{(b_1, b_2, b_3)} B_{b_1, b_2, b_3} k_t^{b_1} \ell_t^{b_2} m_t^{b_3} + \boldsymbol{r}_{t-1} \boldsymbol{\varrho} + e_t \tag{21}$$

and test the hypothesis $\boldsymbol{\varrho} = 0$. Aside from the log sieve approximation this estimating equation is almost identical to Equation 16. But as I show in Appendix C, under the additional assumptions the test satisfies the following properties:

Proposition 1 *Under Assumption 9, if there is an approximation (20) and $\boldsymbol{r}_{t-1}$ is a vector of valid instruments for the autoregressive method, the following properties hold:*

1. *If $\boldsymbol{\varrho} \neq 0$ then for some firms $\Lambda_t \neq 0$, implying neither the Optimal Choice assumption nor the Scalar Unobservable assumption are satisfied. Choice-based estimators may be inconsistent.*

2. *If $\boldsymbol{\varrho} \neq 0$ then $\boldsymbol{\varrho}$ gives the rescaled coefficient from running a regression on the residual variation in Λ_t on the residual variation in $\boldsymbol{r}_{t-1}$ after controlling for $\{k_t^{b_1} \ell_t^{b_2} m_t^{b_3}\}_{(b_1,b_2,b_3)}$, making it informative about the extent of constraints.*

3. *Under Assumption 8, if for all firms $\Lambda_t = 0$ then the autoregressive estimator that estimates $f(\cdot)$ using the approximation (20) is under-identified.*

Properties 1 and 3 show the complementarity between choice-based and autoregressive estimators. The *necessary* conditions for the autoregressive estimator to be identified are *sufficient* conditions for the choice-based estimators to be inconsistent, and vice-versa. When firms are unconstrained the assumptions behind the choice-based estimators are met, but the autoregressive estimator is not identified. When firms are constrained the autoregressive estimator may be well-identified while the assumptions behind the choice-based methods fail. The level of constraints is encapsulated in the Lagrange multiplier Λ_t. By Property 2 the test is directly informative about the size of the Lagrange multiplier, making it informative about the level of constraints. Taken together these properties show why the test can help in choosing between the two estimators. When the testing statistics are large a key assumption behind choice-based methods is suspect. When they are small a key assumption behind autoregressive methods is suspect.

5 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.[11] 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.

5.1 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

$$\omega_{t+1} = \rho\omega_t + \eta_{t+1} \tag{22}$$

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

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

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

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

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 components 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 $\varphi = .5$.[12] 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

$$M_t^* = \min\left\{\left[\theta^M e^{\omega_t} \mathbb{E}[e^{\varepsilon_t}] X(K_t, L_t)\right]^{\frac{1}{1-\theta^M}} \quad , \quad \zeta K_t^{\frac{1}{2}} W_t^{\frac{1}{2}}\right\} \tag{23}$$

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 A.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.

[12] 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.

Table 2
Specifications

	Title	Description
1	Baseline	Baseline case described in the text
2	Near Cobb-Douglas	The elasticity of substitution is set to $\epsilon = 1.25$
3	Cobb-Douglas	Assumes $X(K_t, L_t) = K_t^{\theta^K} L_t^{\theta^L}$
4	Small N	Lowers number of firms to 1000
5	Small T	Shortens panel to 4 years
6	Non-AR Productivity	Makes Ψ a degree-3 polynomial

Note: All parameters are as in the baseline case unless noted otherwise. The polynomial in the case of nonlinear productivity is calibrated to match industry 311 from Gandhi et al. (2013).

This assumption, which only affects the average level of intermediates chosen, is not important because the constraint ζ is chosen to ensure the desired fraction of firms is constrained.[13] 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.

Aside from the baseline I run 5 other variations that illustrate the strengths and weaknesses of each method. These are summarized in Table 2.

5.2 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 both the F-statistic and the partial R-squared of the test.

I estimate Gandhi-Navarro-Rivers using code written by the authors.[14] I make only one change: I impose the assumption that productivity is autoregressive in all but the last specification (nonlinear productivity). By doing so I avoid favoring the autoregressive method, which by construction imposes the (true) autoregressive assumption. In the last specification I let productivity be

[13] 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.

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

a third-order polynomial exactly as Gandhi et al. (2013) does. Again, this is imposing the truth to avoid penalizing Gandhi-Navarro-Rivers.

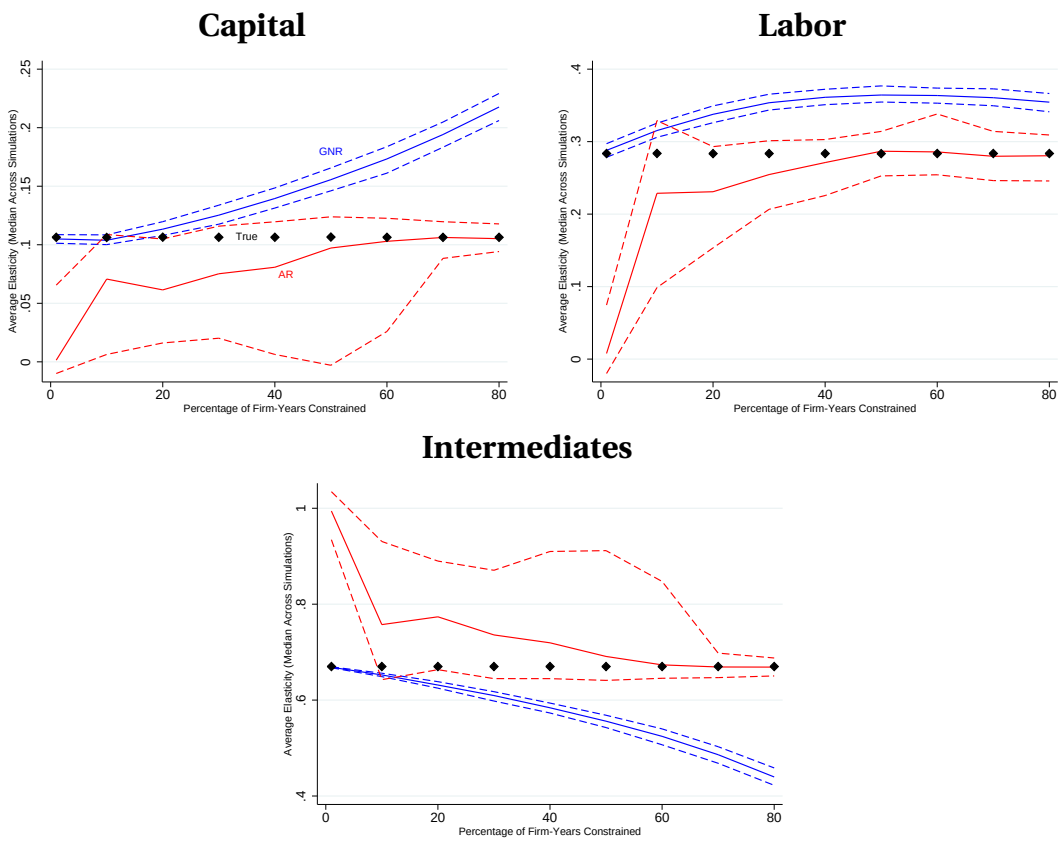
I apply the autoregressive estimator to a second-order polynomial in logs. I first demean every variable to remove the constant in the autoregressive process. (In the true process the constant is zero, but in a real application the researcher would not know that.) I then estimate the equation using as 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.

5.3 Comparing Choice-Based and Autoregressive Methods

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. The estimated elasticity of intermediates is too low while the estimated elasticity of capital is too high. The pattern of the bias is ironic, as ordinary least squares, fixed effects, and other linear methods are often critized 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 Figure 2 suggests it could just as likely be a sign that choice-based methods are biased.

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.

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 4.1.2, this bias is caused by under-identification.

The key conclusion from Figure 2 is that the two methods are complementary. As implied by Proposition 1, one does well precisely when the other does poorly.

5.4 The Test for Constraints

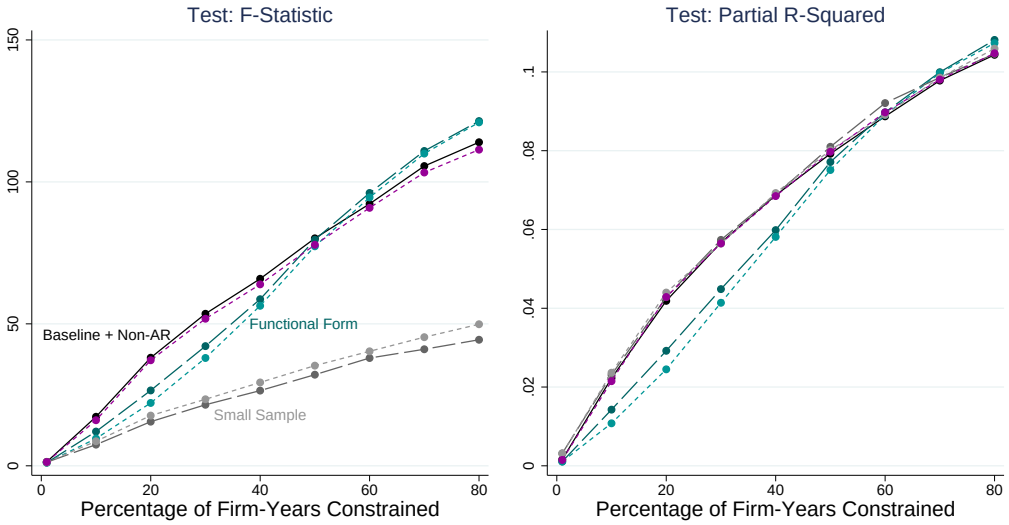
Figure 3 shows the F-statistic and partial R-squared of the test. For a given level of constraints each point shows the median of each statistic across all simulations. Each line is drawn using simulations from one of the six specifications in Table 2. Both statistics rise with the level of constraints. But the F-statistic does not rise as quickly or as high in the two specifications with smaller samples. The partial R-squared by contrast rises at a similar rate across all three specifications. Its invariance to sample size may make it a more reliable measure of the level of constraints.

Figure 4 confirms that the partial R-squared is informative about the root mean-squared error of each method. For each simulated dataset I compute the partial R-squared of the test and estimate the elasticity of each input using Gandhi-Navarro-Rivers and the autoregressive method. I discretize the partial R-squared into bins of width 0.015. Within each bin I compute the mean of the root mean-squared error of each elasticity .[15]

Figure 4 shows that the error of Gandhi-Navarro-Rivers is close to zero for low values of the partial R-squared. The error rises with the partial R-squared, reaching as high as 100 percent of the true value when the partial R-squared is 0.1. The autoregressive estimator is exactly the reverse. Its bias is high when the partial R-squared is low, but falls close to zero when it is 0.1. The crossing point at which the average error of the autoregressive estimator falls below that of Gandhi-Navarro-Rivers depends on the elasticity being estimated. However, by the time the partial R-squared hits 0.07 the autoregressive estimator is almost

[15]For each simulated dataset I compute the absolute difference between the estimate of the average elasticity and the truth. I rescale the difference to be a percentage of the true elasticity. I then take the mean across all simulated datasets that fall within a bin.

Figure 3
The F-Statistic and Partial R-Squared of
the Test Rise as Constraints Tighten



Note: The lines trace out the median of each statistic as ζ is adjusted to change the percentage of firm-year choices that are constrained. Each line gives a different specification (see Table 2). The lines marked "Small Sample" represent Specifications 4 and 5. Those marked "Functional Form" give Specifications 2 and 3. Those marked "Baseline + Non-AR" give Specifications 1 and 6.

Figure 4
The Test Identifies the Estimator with Lower Error

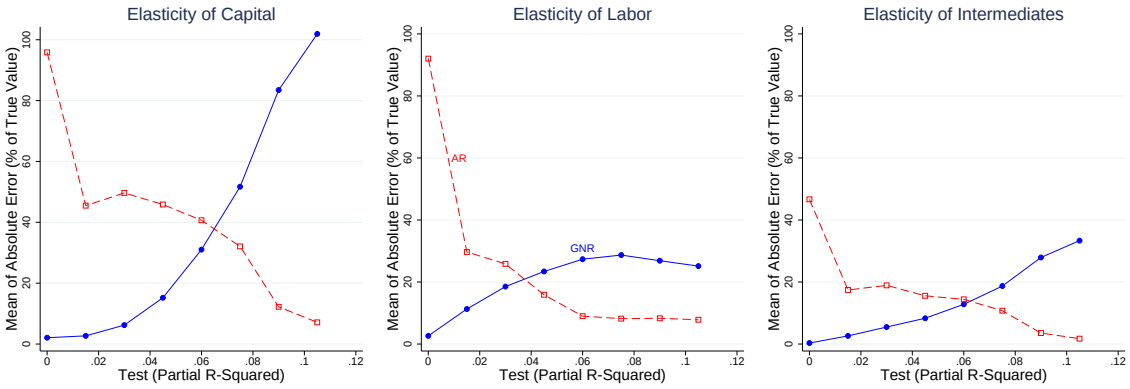
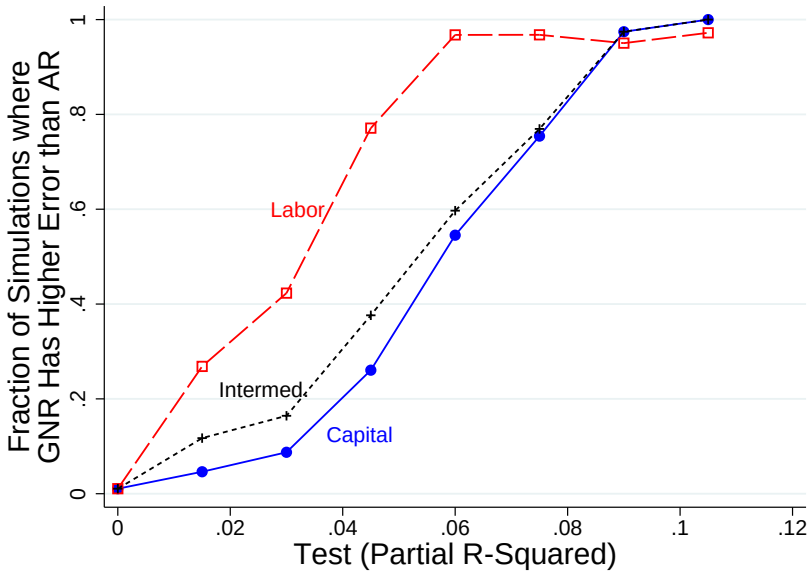


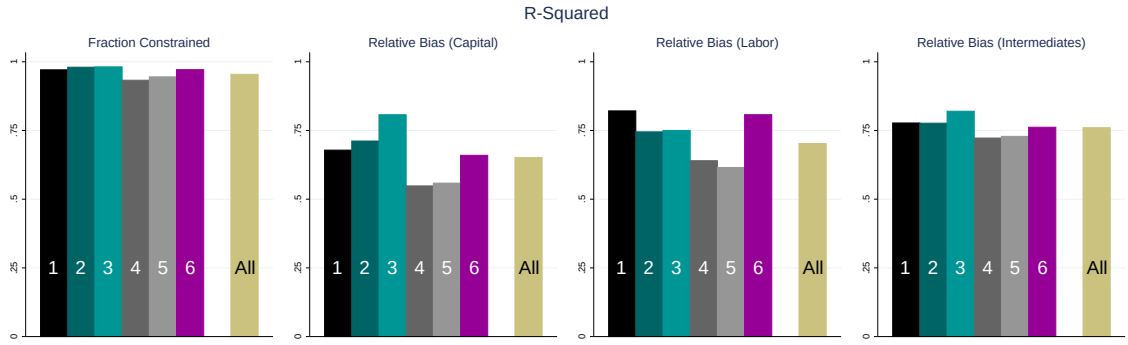
Figure 5
Relative Error of Choice-Based Method
Rises with the Structural Criterion



always the less-biased estimator. Figure 5 which shows the fraction of simulations in which the error of Gandhi-Navarro-Rivers exceeds that of the autoregressive estimator. Here the switching point comes slightly sooner (likely because the error of the autoregressive estimator is right-skewed). But the pattern is similar to Figure 4—the autoregressive estimator becomes more attractive at higher levels of the partial R-squared.

Figure 6 directly confirms that the criterion is informative about both constraints and the relative bias of the two estimators. Taking the set of simulation results as a sample of observations I regress first the fraction of firms constrained and next the relative bias on a high-order polynomial of the partial R-squared of the test. The explanatory power of this regression shows how much is known to a researcher about constraints and bias if she observes only the outcome of the test. Figure 6 suggests she would know quite a bit about the level of constraints. The outcome of the test explains over 90 percent of the variation in the fraction of firms constrained in all 6 specifications. I also report the explanatory power of a regression that pools all specifications, which effectively

Figure 6
The Test is Informative about the Extent
of Constraints and the Relative Bias of the Estimators



assumes the researcher knows neither the level of constraints nor what specification generated the data. Even in this case the test explains nearly all of the variation in the fraction of firms constrained.

The test is less informative about relative bias. It explains roughly 75 percent of the variation in the relative bias of the elasticity of intermediate inputs, and in most cases slightly less of the bias of the other two elasticities. The test is less informative about bias because the performance of each estimator depends on more than just the level of constraints. For example, a small sample size makes the performance each estimator (especially the autoregressive estimator) more variable and thus harder to predict. As expected, Figure 6 suggests the test is less informative about relative bias in the small sample Specifications 4 and 5. Nevertheless, in all cases the outcome of the test explains at least half the variation in relative bias and in many cases close to 75 percent.

6 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. My aim is not to propose a precise scheme for estimating the production function, but merely to provide suggestive evidence of how the statistics proposed in this paper can

guide researchers in choosing the method better suited to their study.

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

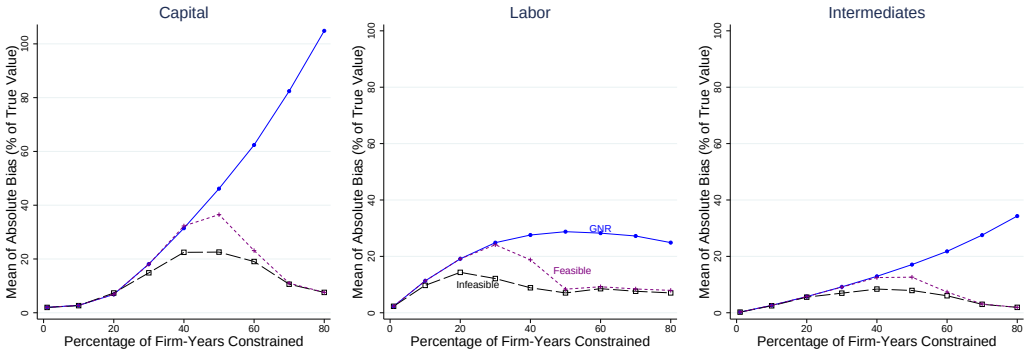
$$\hat{\phi}^X_{Inf} = \begin{cases} \hat{\phi}^X_{GNR} & \text{if } |\hat{\phi}^X_{GNR} - \phi^X| < |\hat{\phi}^X_{AR} - \phi^X| \\ \hat{\phi}^X_{AR} & \text{otherwise} \end{cases}$$

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 7 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? Recall from Figure 4 that the average error of the autoregressive estimator falls below that of Gandhi-Navarro-Rivers for high levels of the partial R-squared of the test. A simple, rule-of-thumb approach is to choose a threshold and use the Gandhi-Navarro-Rivers estimator if the partial R-squared is below this level, and to use the autoregressive estimator otherwise. As an added precaution I first check whether the F-statistic is large enough to reject at the 1 percent level. If not I choose Gandhi-Navarro-Rivers regardless of the partial R-squared.

Figure 7
A Combined Approach Gives More Accurate Estimates



Let $R_p^2\left(\boldsymbol{r}_{t-1}\boldsymbol{\varrho}\right)$ denote the structural criterion, and let $\mathcal{P}\left(\boldsymbol{\varrho}\right)$ be the p-value of the F-statistic. The choice of threshold is necessarily arbitrary (as I describe below), but for this exercise I choose a fairly conservative threshold of 0.07. Then define the feasible estimator as

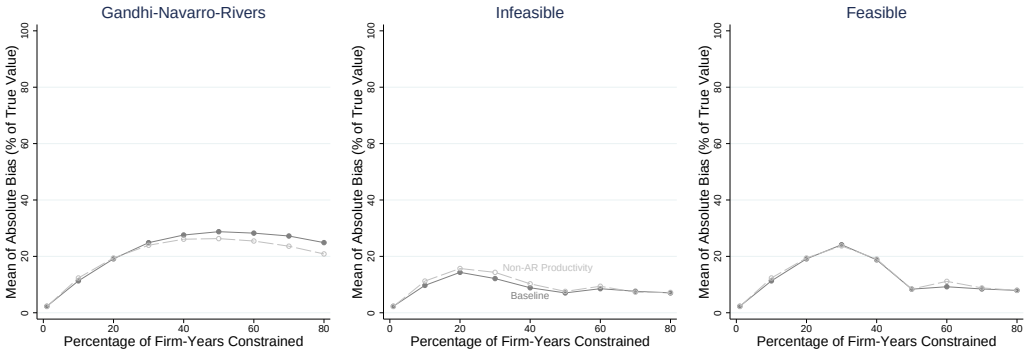
$$\hat{\phi}^X_{Feas} = \begin{cases} \hat{\phi}^X_{GNR} & \text{if} \quad R_p^2\left(\boldsymbol{r}_{t-1}\boldsymbol{\varrho}\right) < 0.07 \quad \text{or} \quad \mathcal{P}\left(\boldsymbol{\varrho}\right) > \mathcal{T}_{0.01} \\ \hat{\phi}^X_{AR} & \text{otherwise} \end{cases}$$

As Figure 7 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 A.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 8 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

Figure 8
Calibrated, Non-Autoregressive Productivity



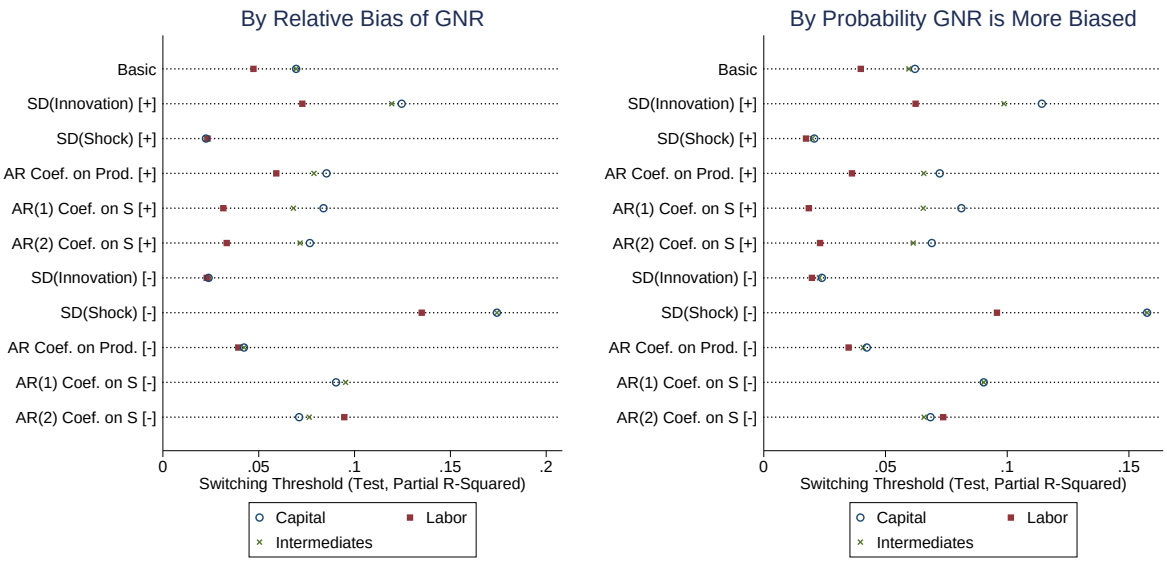
Note: All figures show the mean absolute error in estimating the labor elasticity.

or any estimator that relies on it is robust to all deviations from Autoregressive 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.

As noted at the beginning of this section my aim is only to demonstrate how the test and the criterion might be used as diagnostic tools, not to propose a precise method for estimating the production function. These simulations suggest switching from choice-based to autoregressive methods for large values of the criterion may reduce the error of the estimates. But the threshold, chosen here to be conservative, may not be universal.

Figure 9 confirms this intuition by generating additional simulations. In all cases the parameters of the production function are drawn at random. In additional, all of the cases except the "basic" case vary one of the key parameters previously calibrated to Chile's Industry 311. For example, the specification labeled "SD(Innovation) [+]" increases the standard deviation of η_t while "SD(Innovation) [-]" decreases it. (The details are in Appendix A.) For each specification I calculate the threshold at which the expected bias of Gandhi-Navarro-Rivers exceeds that of the autoregressive estimator (lefthand panel). I also compute the threshold at which Gandhi-Navarro-Rivers is more likely than not to have a larger bias.

Figure 9
The Precise Optimal Switching Threshold May Vary with the Parameters but Typically Lies between 0.025 and 0.1



The critical threshold varies across specifications. A wise researcher would avoid taking any single number as the "true threshold" and, as suggested above, use the test for guidance. But it does seem that, even among these 11 different simulations, Gandhi-Navarro-Rivers is always the less biased estimator when the partial R-squared of the test is below 0.025. For values of the partial R-squared over 0.1 the autoregressive estimator is almost certain to be the less biased estimator. For values in between there may be cause for concern about Gandhi-Navarro-Rivers, especially with estimates of the labor elasticity, but at the lower end of this range Gandhi-Navarro-Rivers probably still yields less biased estimates of the elasticities of capital and intermediates. Looking back to the empirical results of Section 3.2, the Apparel industry of Chile and the rice sector in Thailand may be better estimated using an autoregressive method (at least for the elasticity of labor). Food and wood products in Colombia and fabricated metals in both countries are likely best estimated using Gandhi-Navarro-Rivers. Several of the other industries are more marginal, though it is likely that Gandhi-Navarro-Rivers is the better estimator.

7 Summary

I derive a test for one of the key identifying assumptions of choice-based methods for estimating production functions. I show that the test rejects the assumptions in a wide range of industries and countries. I then show that the test can be used to assess the relative bias of choice-based methods and a simple autoregressive estimator. In simulations, using the test to choose between choice-based and autoregressive methods has lower error than either method alone. This result suggests that by identifying when and where firms are constrained, the test can guide researchers in choosing the best suited method for estimating the production function.

References

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

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

ALLCOTT, H., A. COLLARD-WEXLER, AND S. D. O'CONNELL (2014): "How Do Electricity Shortages Affect Productivity? Evidence from India," *NBER Working Paper*.

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*.

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," .

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

A Simulation Appendix

A.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

Table 3
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		

Table 4
Parameters: Shocks

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

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 3 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 4 shows the standard deviations of the three shocks in the simulation. 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 5 shows the parameters of the three production functions (basline, almost Cobb-Douglas, and Cobb-Douglas). For each elasticity ϵ I chose the other

Table 5
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

Table 6
Parameters Varied to Produce Figure 9

	[+]	[-]
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	$\alpha_2^S = .42$	$\alpha_2^S = .25$

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 9 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

$$\begin{aligned}\sigma &= 0.39 + 0.4u \\ \theta^K &= 0.015 - 0.04(u - .4) \\ \theta^L &= 0.95 + u + .1\end{aligned}$$

The "Basic" scenario varies only the production function. The other specifications also vary one other parameter, as summarized in Table 6:

A.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

$$\begin{aligned} &y_{it} - \tau_k k_{it} - \tau_\ell \ell_{it} - \tau_{kk} k_{it}^2 - \tau_{\ell\ell}\ell_{it}^2 - \tau_{k\ell} k_{it}\ell_{it} \\ &\quad - \rho(y_{i,t-1} - \tau_k k_{i,t-1} - \tau_\ell \ell_{i,t-1} - \tau_{kk} k_{i,t-1}^2 - \tau_{\ell\ell}\ell_{i,t-1}^2 - \tau_{k\ell} k_{i,t-1}\ell_{i,t-1}) \end{aligned} \tag{24}$$

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 (24), 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.[16]

For the simulations that follow I assume the firm sets the log of m equal to a

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

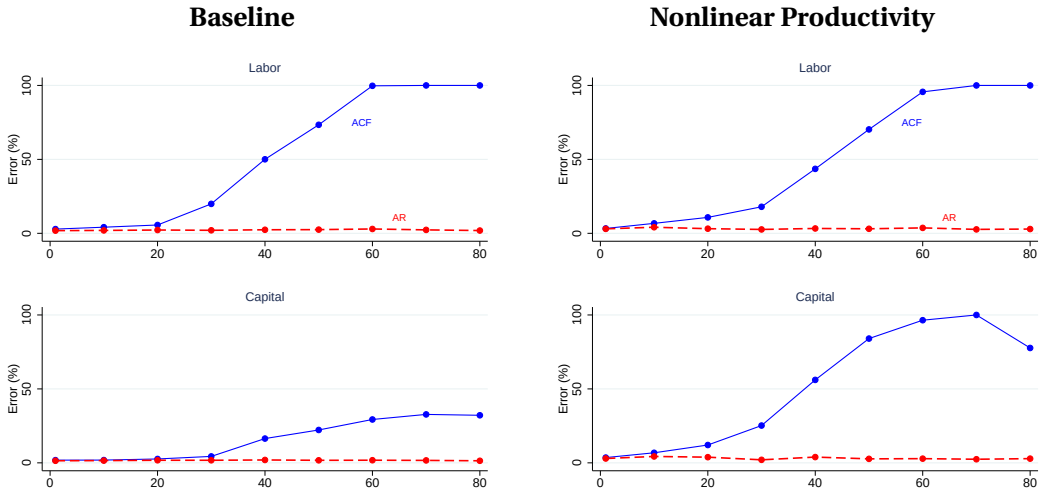
$$Y = e^{\varepsilon}\min\{e^{\omega}\tilde{F}(K,L), \theta^M M\}. \tag{25}$$

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 10
The Autoregressive Method and Ackerberg-Caves-Frazer



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 A 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 10 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.

B Data Appendix

B.1 Indian Farming

I construct output and the factors of production as follows:

- **Output —** The survey records both the quantity and revenue of each crop produced on each plot of land in each season. The quantities are recorded in several different units (quintals, bags, etc.), which I convert to kilograms.[17] I discard observations recorded as "bunches." In cases where multiple crops were planted on the same plot (less than 4 percent of cases) I use the first crop, which is the crop that used most of the land.

- **Land —** The survey records the number of acres planted for each crop on each plot. I take the number of hectares as my measure of land.

- **Labor —** The survey records the number of days of family and hired labor spent on each operation in production. I define labor to be the total labor of all family and workers excluding time spent on supervision, which is more like management than labor.

- **Capital —** I first compute the value of owned capital. I define the following pieces of machinery as capital used in farming: Tractor, Trailer, Thresher, Tractor with trailer, Tractor with thresher, Tractor with oil engine, Tractor with gauge wheel, Tractor with plough, Plough disc/mould board, Seed drill, Power tiller, Power sprayer, Chaff cutter, Combine harvester, Harrow, Leveler, Hoe / Manual earth remover, Rice pounder. After computing the value of these machines for each household I compute a daily user cost assuming machinery depreciates at 10 percent per year

[17] I assume each bag weighs 100 kilograms, which is a common selling size.

and an interest rate of 4 percent. To this I add a user cost of owned bullocks, which I take as the number of bullocks owned times the daily price at which the household rents bullocks.[18]

- **Intermediates —** I define intermediates as the value of all seeds, seedlings, manure, chemical fertilizer, pesticides, canal use fees, and transportation costs. Some seeds, seedlings, and manure is home-produced. I compute the value of these by calculating the average price of each within the village and multiply the price by the quantity of the home-produced inputs. I define nominal intermediate inputs as the total of expenditure on all seven inputs. For the real value I take the average price of inputs across the entire sample (for each crop) and take the product of these average prices with the quantity of each input (either home produced or purchased). The one exception is transport costs, for which no quantity is provided—for this I must simply use the total expenditure. Since transport costs are not a big part of crop production, this is unlikely to create too much noise. I define the sum of these real expenditures to be real intermediate inputs.

C Complementarity

This section proves Proposition 1.

C.1 Property 1

Substitute the sieve approximation of $\log F_M(K_t, L_t, M_t)$ into Equation 37 and add s_t^M to both sides:

[18] If the household rents no bullocks I take the village average; if no households in the village rent bullocks I use the average within the smallest level of aggregation—subdistrict, district, state—for which I observe at least one household renting bullocks. In some villages bullocks are rented by the acre instead of by the day. To rescale these prices I look for villages that quote prices in both days and acres and use these two prices to compute the average acres plowed per day, then take the median across the sample. I use the acres per day to rescale by-acre prices to daily prices. I use a similar method for prices quoted in hours.

$$
\begin{aligned}
s_t^M + \Lambda_t &= \omega_t + \log \mathbb{E}[e^{\varepsilon_t}] + \sum_{a_1^1, a_2^1, a_3^1} A^1_{a_1^1, a_2^1, a_3^1} k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1} + m_t - y_t \\
&= \log \mathbb{E}[e^{\varepsilon_t}] + \sum_{a_1^1, a_2^1, a_3^1} A^1_{a_1^1, a_2^1, a_3^1} k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1} + m_t - \sum_{a_1^0, a_2^0, a_3^0} A^0_{a_1^0, a_2^0, a_3^0} k_t^{a_1^0} \ell_t^{a_2^0} m_t^{a_3^0} - \varepsilon_t \\
&= \sum_{(b_1, b_2, b_3)} B_{b_1, b_2, b_3} k_t^{b_1} \ell_t^{b_2} m_t^{b_3} - \varepsilon_t
\end{aligned}
$$

for some set of coefficients $\{B_{b_1,b_2,b_3}\}$. Define $e_t = -\Lambda_t - \varepsilon_t$. Then

$$
s_t^M = \sum_{(b_1, b_2, b_3)} B_{b_1, b_2, b_3} k_t^{b_1} \ell_t^{b_2} m_t^{b_3} + e_t \tag{26}
$$

Suppose $\Lambda_t = 0$ for all firms. Then $e_t = -\varepsilon_t$ which by assumption is unanticipated and thus uncorrelated with anything known to the firm at t or earlier—in particular, $\boldsymbol{r}_{t-1}$. Then $\boldsymbol{r}_{t-1}$ is uninformative about s_t^M after controlling for the polynomial in $\{k_t, \ell_t, m_t\}$, implying $\boldsymbol{\varrho} \xrightarrow{a.s.} 0$ in an ordinary least squares estimate of Equation 21. □

C.2 Property 2

Let $\ddot{\boldsymbol{r}}_{t-1}$ be the vector of residuals from a regression of $\boldsymbol{r}_{t-1}$ on $\{k_t^{b_1} \ell_t^{b_2} m_t^{b_3}\}_{(b_1,b_2,b_3)}$. Let $\hat{e}_t$ be the residual from estimating (26). By the Partitioned Regression Theorem, the estimated vector of coefficients $\hat{\boldsymbol{\varrho}}$ is equivalent to the coefficients estimated in the regression

$$
\hat{e}_t = \ddot{\boldsymbol{r}}_{t-1} \boldsymbol{\varrho} + o_t
$$

Suppose $\Lambda_t \neq 0$. Let P be the orthogonal projection matrix implies by the ordinary least squares estimator of (26), and let I be the identity matrix. Then

$$
\begin{aligned}
\hat{e}_t &= (I - P)(-\Lambda_t - \varepsilon_t) \\
&= -\ddot{\Lambda}_t - \ddot{\varepsilon}_t
\end{aligned}
$$

implying the residual $\hat{e}_t$ is equal to the sum of two sub-residuals: the residual from a regression of $-\Lambda_t$ on $\{k_t^{b_1} \ell_t^{b_2} m_t^{b_3}\}_{(b_1,b_2,b_3)}$, and the residual from a regres-

sion of $-\varepsilon_t$ on the same set of regressors. Then by the linearity of the ordinary least squares estimator

$$\hat{\boldsymbol{\varrho}} = \hat{\boldsymbol{\varrho}}^{\Lambda} + \hat{\boldsymbol{\varrho}}^{\varepsilon}$$

where $\hat{\boldsymbol{\varrho}}^{\Lambda}$ and $\hat{\boldsymbol{\varrho}}^{\varepsilon}$ are the coefficients estimated by regressing the first and second sub-residual on $\ddot{\boldsymbol{r}}_{t-1}$. By the argument above, $\hat{\boldsymbol{\varrho}}^{\varepsilon} = \mathbf{0}$, implying $\hat{\boldsymbol{\varrho}} = \hat{\boldsymbol{\varrho}}^{\Lambda}$. □

C.3 Property 3

The following lemma is useful in proving this property:

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

$$\begin{aligned}\widetilde{m_t} = \rho\widetilde{m_{t-1}} + \sum_{(a_1^1,a_2^1,a_3^1)\neq(0,0,1)} A'_{a_1^1,a_2^1,a_3^1}\big(\widetilde{k_t^{a_1^1}\ell_t^{a_2^1}m_t^{a_3^1}} - \rho\widetilde{k_{t-1}^{a_1^1}\ell_{t-1}^{a_2^1}m_{t-1}^{a_3^1}}\big) \\ - \frac{1}{-A_{0,0,1}}(\widetilde{\Lambda}_t - \rho\widetilde{\Lambda_{t-1}}) + \frac{1}{-A_{0,0,1}}\widetilde{\eta}_t \end{aligned} \tag{27}$$

Substitute the sieve approximation of $\log F_M(K_t, L_t, M_t)$ into Equation 37:

$$\begin{aligned}\Lambda_t &= \omega_t + \log\mathbb{E}[e^{\varepsilon_t}] + \sum_{a_1^1,a_2^1,a_3^1} A^1_{a_1^1,a_2^1,a_3^1} k_t^{a_1^1}\ell_t^{a_2^1}m_t^{a_3^1} \\ &= \omega_t + \log\mathbb{E}[e^{\varepsilon_t}] + \sum_{(a_1^1,a_2^1,a_3^1)\neq(0,0,1)} A^1_{a_1^1,a_2^1,a_3^1} k_t^{a_1^1}\ell_t^{a_2^1}m_t^{a_3^1} + A^1_{0,0,1}m_t \\ \Rightarrow m_t &= \frac{1}{-A^1_{0,0,1}}\omega_t + \frac{1}{-A^1_{0,0,1}}\log\mathbb{E}[e^{\varepsilon_t}] + \sum_{(a_1^1,a_2^1,a_3^1)\neq(0,0,1)} \underbrace{\frac{A^1_{a_1^1,a_2^1,a_3^1}}{-A^1_{0,0,1}}}_{A'_{a_1^1,a_2^1,a_3^1}} k_t^{a_1^1}\ell_t^{a_2^1}m_t^{a_3^1} - \frac{1}{-A^1_{0,0,1}}\Lambda_t\end{aligned} \tag{28}$$

Apply the Autoregressive Productivity assumption to the definition of ω_t:

(29)

$$m_t = \frac{1}{-A_{0,0,1}}(\bar{\omega} + \rho\omega_{t-1} + \eta_t) + \frac{1}{-A_{0,0,1}} \log \mathbb{E}[e^{\varepsilon_t}] + \sum_{(a_1^1,a_2^1,a_3^1)\neq(0,0,1)} A'_{a_1^1,a_2^1,a_3^1} k_t^{a_1^1} \ell_t^{a_2^1} m_t^{a_3^1} - \frac{1}{-A_{0,0,1}} \Lambda_t \tag{30}$$

Take the lag of (28):

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

Substitute this expression into (30):

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

After demeaning, this expression becomes

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

□

Proof of Property: To keep the notation simple assume all variables in what follows are demeaned and suppress the $\widetilde{\ }$.

Suppose as in the text that the production function is approximated with a polynomial of arbitrary order in the logs of each input. The residual is

$$\nu_t = y_t - \sum_{(b_1,b_2,b_3)} \pi_{b_1,b_2,b_3} k_t^{b_1} \ell_t^{b_2} m_t^{b_3} - \rho\Big(y_{t-1} - \sum_{(b_1,b_2,b_3)} \pi_{b_1,b_2,b_3} k_{t-1}^{b_1} \ell_{t-1}^{b_2} m_{t-1}^{b_3}\Big) \tag{31}$$

Let $\boldsymbol{\pi} = [\pi_{1,0,0}, \pi_{0,1,0}, \pi_{0,0,1}, \ldots, \rho]^T$ be the vector of parameters. If the polynomial approximation has B terms this vector has $B+1$ elements. As before, let $\boldsymbol{r}_{t-1} = [r_{t-1}^1, r_{t-1}^2, \ldots, r_{t-1}^R]^T$ be the vector of instruments uncorrelated with η_t. By assumption, this vector contains the lags of all the terms of the polynomial approximation and several second-order lags $\boldsymbol{r}_{t-1}^{[2]}$.

The rank condition is

$$rank\left(\mathbb{E}\left[\frac{\partial(\nu_t \boldsymbol{r}_{t-1})}{\partial \boldsymbol{\pi}}\right]\right) = B+1 \tag{32}$$

Define $J_x = \frac{\partial(\nu_t)}{\partial x}$. The Jacobian of the moment condition equals

$$\frac{\partial(\nu_t \boldsymbol{r}_{t-1})}{\partial \boldsymbol{\pi}} = \begin{pmatrix} r_{t-1}^1 J_{\pi_{1,0,0}} & r_{t-1}^1 J_{\pi_{0,1,0}} & r_{t-1}^1 J_{\pi_{0,0,1}} & \cdots & r_{t-1}^1 J_\rho \\ r_{t-1}^2 J_{\pi_{1,0,0}} & \ddots & & & \vdots \\ \vdots & & & & \\ r_{t-1}^R J_{\pi_{1,0,0}} & r_{t-1}^R J_{\pi_{0,1,0}} & r_{t-1}^R J_{\pi_{0,0,1}} & \cdots & r_{t-1}^R J_\rho \end{pmatrix} \tag{33}$$

which is an $R \times (B+1)$ matrix.

The rank condition fails if there exist coefficients $\{C_{c_1,c_2,c_3}\}$ such that

$$\mathbb{E}[r_{t-1}^n J_{\pi_{0,0,1}}] = \sum_{(c_1,c_2,c_3)\neq(0,0,1)} C_{c_1,c_2,c_3} \mathbb{E}[r_{t-1}^n J_{\pi_{c_1,c_2,c_3}}] \tag{34}$$

for more than $R-(B+1)$ instruments. This rank condition is equivalent to

$$\begin{aligned} 0 &= \mathbb{E}[r_{t-1}^n J_{\pi_{0,0,1}}] - \sum_{(c_1,c_2,c_3)\neq(0,0,1)} C_{c_1,c_2,c_3} \mathbb{E}[r_{t-1}^n J_{\pi_{c_1,c_2,c_3}}] \\ &= \mathbb{E}[r_{t-1}^n \big(J_{\pi_{0,0,1}} - \sum_{(c_1,c_2,c_3)\neq(0,0,1)} C_{c_1,c_2,c_3} J_{\pi_{c_1,c_2,c_3}}\big)] \end{aligned} \tag{35}$$

The partial derivatives of (31) imply

$$J_{\pi_{c_1,c_2,c_3}} = \pi_{c_1,c_2,c_3}(k_t^{c_1}\ell_t^{c_2}m_t^{c_3} - \rho k_{t-1}^{c_1}\ell_{t-1}^{c_2}m_{t-1}^{c_3}) \qquad \forall c_1, c_2, c_3$$

Let $C_{c_1,c_2,c_3} = A'_{c_1,c_2,c_3}/\pi_{c_1,c_2,c_3}$. If $\Lambda_t = 0$, by Equation 27 the term in parentheses equals $-\eta_t/A^1_{0,0,1}$ at the true value of ρ. But by the assumption that $\boldsymbol{r}_{t-1}$ is a vector of valid instruments that satisfy the exclusion restriction,

$$\mathbb{E}[-r_{t-1}^n \eta_t / A^1_{0,0,1}] = 0$$

for all n, proving that the rank condition fails.

D Miscellaneous Results

D.1 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 10 (Optimal Choices, Modified) *Firms choose M_t to satisfy*

$$P_t^M = \mathbb{E}[e^{\varepsilon_t}]e^{\omega_t}F_M(K_t, L_t, M_t) \tag{36}$$

which implies the Constrained Optimal Choice assumption becomes

Assumption 11 (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*

$$p_t^M + \Lambda_t = \omega_t + \log \mathbb{E}[e^{\varepsilon_t}] + \log F_M(K_t, L_t, M_t) \tag{37}$$

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.[19]

Assumption 12 (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

$$s_t^M = \bar{\xi}(k_t, \ell_t, m_t, p_t^M) + \boldsymbol{r}_{t-1}\boldsymbol{\varrho} + e_t \tag{38}$$

This modification is only necessary if the researcher is unwilling to make Assumption 11 but is willing to make Assumption 12. Since optimal choices are necessary for Gandhi-Navarro-Rivers, I effectively take Assumption 11 as given in the empirical application.

Finally, Property 3 of Proposition 1 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.

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