# Production functions without control functions: Estimating input elasticities when monotonicity does not hold.

Mark Ponder

University of Minnesota

October 24, 2021

Click here for the most current version

**Abstract**

It is becoming increasingly common to use cost-minimization first-order conditions to recover firm-level markups. In principal, researchers do not need to specify a competitive environment in order to estimate price-cost margin which adds considerable appeal of this approach. However, researchers need to be careful that their method of estimating input elasticities, a crucial ingredient to this approach, is internally consistent with the estimates they are attempting to recover. The validity of control function methods, a common technique used to estimate input elasticities, rely on several assumptions that can be incompatible with a wide range of competitive environments. This paper proposes a new method of estimating production functions that relaxes some of the main assumptions of the control function approach, namely scalar unobservables and monotonicity. It does so by extending the results of Hausman et al. (1991) to a dynamic panel setting. Because these assumptions place the most restrictions on the economic environment, this has the potential of broadening the applicability of using cost-minimization moments to recover firm markups.

# 1 Introduction

Using cost-minimization first-order conditions has become a popular approach to estimating firm and time-specific markups. In theory, the method of De Loecker and Warzynski (2012) allows researchers to recover markups without specifying a competitive environment. Standard approaches in the industrial organization literature require researchers to specify the nature of competition, estimate demand parameters and then back out implied marginal costs. By maintaining cost-minimization, researchers can instead estimate input elasticities and use them to scale the sales-to-operating expense ratio to get an estimate of markups. If the researcher also observes physical units of output, they can further recover an estimate of marginal costs. One of the key arguments for using a cost-minimization approach is that the researcher does not need to specify the form of competition, e.g. Nash-Bertrand or Cournot pricing, as estimated markups can be sensitive to the proposed environment. However, in practice, this strength can often be illusory.

The principal issue is that estimation routines used to recover input elasticities often implicitly or explicitly constrain the economic environments that they are compatible with. Consider the control function approach of Olley and Pakes (1996). The assumption that there is a single unobservable, productivity, precludes the leading case where a firm's product market decisions depend on the state variables of their competitors. Levinsohn and Petrin (2003) (henceforth LP) point out that we can continue to maintain the scalar unobservable assumption by including these state variables in the control function, but this requires the researcher to take a stand on which state variables are relevant and therefore requires assumptions on the competitive environment. Similarly, a rote application of the LP estimation routine is not always possible as the researcher needs to first establish that monotonicity holds even when markets are not perfectly competitive, which is the case the authors consider.

In this paper, I extend the work of Hausman et al. (1991) to estimate production functions

when productivity is subject to non-transitory shocks. Under the standard assumption that these shocks are i.i.d. the parameters of the production function can be consistently estimated using two-stage least squares. The key idea is that after projecting the relevant covariates onto the space spanned by a set of instruments, the transitory errors only contaminate the estimates of the productivity process. As long as we are primarily interested in the production function then there is nothing further to do. However, if the true Markov process governing productivity is polynomial, then all model parameters can be consistently estimated.

This paper makes several contributions. First, it provides a means of estimating production functions that does not depend on the assumptions of monotonicity and a scalar unobservable. As such, this estimation routine has the potential of being more generally applicable to different economic environments. Second, it shows how both the standard control function approach and the proposed approach can be estimated using two-stage least squares. When properly specified, the control function estimator is efficient, and so we can use a Durbin-Wu-Hausman test to test the validity of the control function assumptions during estimation. Finally, this paper explores the properties of proposed estimator, including how to use iterated least-squares to generate estimates with improved small-sample bias.

One strength of the proposed estimator is that the Markov process governing productivity can be arbitrarily specified, which stands in contrast to the Arellano-Bond estimator where productivity is assumed to be AR(1). The main drawback is that we can only estimate ex-post productivity and therefore, ex-post markups. Unlike the control function approach, unobserved productivity cannot be separately identified from the transitory shock. As such, the proposed estimator can be viewed as an extension of the Arellano-Bond estimator to accommodate more general productivity processes and as a complement to the standard OP, LP, and ACF estimators by providing an statistical test of their validity. When the assumptions of monotonicity and scalar unobserability hold, the control function approach

will allow a more complete description of the firms environment

This paper proceeds as follows. In Section (2) I detail common assumptions used when estimating production functions and how parameters estimates are recovered from these assumptions. In Section (3), I propose a new method of estimating production functions when productivity follows an arbitrary Markov Process. I further discuss how we can test the validity of the control function assumptions . Next, in Section (4), I present a Monte Carlo exercise and discuss the properties of the estimator. Finally, in Section (5) I discuss the empirical application in Ponder (2021) and how the proposed method impacts the results.

# 2 Production Function Estimation

Following Griliches and Mairesse (1995), the production function is assumed to be log-separable in productivity

**Assumption 1.** *The relationship between output and inputs takes the log-separable form*

$$Q_{it} = F(K_{it}, V_{it}; \beta) \exp^{\omega_t + \epsilon_t}$$

The term $\omega_{it}$ is the econometrician's problem. This represents a persistent productivity that is unobserved by the researcher but is known to the firm. The term $\epsilon_{it}$ is an i.i.d. random error, which can represent unanticipated shocks to output or measurement error. The timing is as follows. Firms choose their fixed input $k_t$ in period $t-1$. They then observe $\omega_{it}$ and choose the flexible input to maximize profits. This causes an endogeneity problem between $\omega_{it}$ and $v_{it}$.[1] Firms know $\omega_{it}$ when $v_{it}$ is chosen and so $v_{it}$ will be correlated with $\omega_{it}$. This simultaneity issue has long been recognized in the literature, and has been handled in several different ways. I focus on two popular methods: dynamic panels and control

[1]Following the literature, I write logged values in lowercase and levels in uppercase.

functions. In the dynamic panel approach $\omega_{it}$ is assumed to follow an $AR(1)$ process.[2]

**Assumption 2.** *The unobserved productivity $\omega_t$ follows an auto-regressive process*

$$\omega_t = \rho_0 + \rho_1 \omega_{t-1} + \nu_t$$

Differencing then allows the researcher to control for the structural part of the residual. The autoregressive assumption is key for a differencing approach, as it allows the researcher to construct the relevant moments without worrying about unobservables entering nonlinearly. However, Assumption 2 is strong and rules out several economically meaningful processes by which technology might evolve.[3] Additionally, the process of differencing can remove significant amounts of identifying variation in the data and can lead to imprecise estimates. Bond and Söderbom (2005) provide several Monte Carlos that explore these issues and show that when adjustment costs are high, parameter estimates become imprecise.

Olley and Pakes (1996) (OP) proposed a novel estimator that solves several of these problems. Their estimator relies on the existence of a control function that can account for variation in unobserved productivity. They weaken Assumption 2, and replace it with the more general specification

**Assumption 3.** *The unobserved productivity $\omega_t$ follows a Markov Process*

$$\omega_t = E[\omega_t | \omega_{t-1}] + \nu_t$$

*and $E[\omega_t|\omega_{t-1}]$ can be approximated by a high-order polynomial.*

To estimate the model, they include the following assumption

[2]This approach is actually more general than just an $AR(1)$ process and can include firm fixed effects and other determinants of idiosyncratic productivity.

[3]For instance, a process where firms have low productivity for awhile, but then a large shock pushes them to a higher productivity state is easily modeled with a high order polynomial but is ruled out in a linear model.

**Assumption 4** (Scalar unobservability)**.** *Investment demand can be written as*

$$i_t = I(k_t, \omega_t)$$

*and is strictly monotonic in* $\omega_t$*.*

Assumption (4) implies that we can invert the investment function for $\omega_t$ and substitute for $\omega_t$ in the production function

$$q_{it} = f(k_{it}, v_{it}; \beta) + h(k_{it}, i_{it}) + \epsilon_{it}$$

where $h$ is an unknown function. We can non-parametrically regress $y_{it}$ on $k_{it}$, $v_{it}$, and $i_{it}$ to recover an estimate of $\epsilon_{it}$. Then, for each guess of $\beta$, we can generate an approximate version of $\hat{\omega}_t = y_{it} - f(k_{it}, v_{it}; \beta) - \hat{\epsilon}_t$. Regressing $\hat{\omega}_t$ on a polynomial of its lags and then interacting the residuals with valid instruments generates a GMM quadratic form, which we minimize with a search over $\beta$. There are several strengths to this approach. First, it does not rely on differencing and therefore leaves more identifying variation in the data, allowing for more precise estimates. Second, the process governing the structural part of productivity is allowed to evolve in a more general way. Finally, this method is able to separate out the structural component of productivity from the transitory component, allowing the researcher to analyze firm-specific markups. However, these strengths are not without drawbacks.

Levinsohn and Petrin (2003) (LP) note that the inversion relies on strict monotonicity, rather than weak monotonicity. This is problematic because many firms have zero investment in at least one year. These data either need to be discarded or the estimation routine needs to be adjusted to account for these points. The authors proposed using intermediate input demand for the control function instead. Under the assumption of perfect competition, they show that demand for intermediate inputs is monotonically increasing in $\omega_{it}$. Therefore, one could replace $i_{it}$ with $m_{it}$ (intermediate input demand) and then estimate

the model in an analogous manner. This has the benefit that most firms always report demand for intermediates, even when investment is zero. Several subsequent papers explored the implications of using intermediate input demand for a control function. Ackerberg et al. (2015) (ACF) note that in several leading cases intermediate inputs are functionally dependent on $v_{it}$, $k_{it}$ and $\omega_{it}$. This results in there being no identifying variation between $m_{it}$ and $v_{it}$. If there was a component of input demand that would separately vary it from $v_{it}$ (such as input prices), then we would need to include this component in the control function to satisfy Assumption 4. ACF propose several timing assumptions that lead to consistent estimation with a value-added production function. Gandhi et al. (2020) describe a similar problem for gross-production functions. They show that the gross-production function is in general nonparametrically unidentified for a similar reason: lack of independent variation in flexible inputs. Their argument does offer an insight on how one might be identified using a panel approach. If input prices are constant across firms, but are time varying[4], then there is sufficient variation to identify the model parameters. LP note that we can estimate a separate control function for each subset of the data where monotonicity and scalar unobservability holds. We can therefore estimate a separate control function by year and be left with enough variation to identify the model parameters. However, this requires a specific data generating process and still depends crucially on perfect competition.

Unfortunately, the scalar unobservables assumption is particularly problematic in imperfect markets. As shown in Bond et al. (2020), the intermediate input demand is in general a function of the firm's markup. Typically, the firm's markup is unknown and must be estimated by other means[5] and so scalar unobservables is necessarily violated if markups are firm specific and time varying. We can follow Gandhi et al. (2020) and assume that output and input prices are common across firms in the industry, but this requires firms to face the same residual demand curve and restricts the type of competition compatible

[4]Or constant across time but vary by firm

[5]As evidenced by the popularity of using financial data to estimate markups a la De Loecker and Warzynski (2012).

with the estimation routine. Additionally, even if the assumption a scalar unobservable holds, weak monotonicity might not hold. If a competitor has high levels of capital, a firm might employ less labor given a higher draw of productivity because it doesn't want to increase output further. In practice we only need the investment function to be monotonic in an region around the data, but this assumption is impossible to empirically verify and depends on the assumed form of competition. Even observing output prices, as I do, we still need to make assumptions on markups and input prices to get scalar unobservables and monotonicity. It is commonly assumed that input prices are constant across firms (and often implicitly across time). For each subset of firms where these prices are constant, we can estimate a separate control function that satisfies Assumption 4. However, we either need to explicitly control for the determinants of markups, in the manner of [ZZ], or we need to assume that they are piece-wise constant for a partition of the data.

# 3 Projected Control Function

To circumvent these issues, I maintain Assumption 3 and replace Assumption 4 with the following

**Assumption 5** (Independent Errors)**.** *The transitory error component satisfies*

$$\epsilon_t \perp (k_t, v_t, \omega_t, Z_t)$$

*where $Z_t$ is a vector of instruments. All moments of $\epsilon_t$ exist and are finite.*

This assumption is stronger than necessary, but simplifies the discussion. A weaker assumption would be that

$$E[\epsilon_t^i | y_t^*, Z_t] = E[\epsilon_t^i | y_t^*]$$

where $y_t^* = f(k_t, v_t) + \omega_t$. That is, conditional on knowing $y_t^*$, the instruments provide

no information on the moments of $\epsilon_t$. The advantage of this approach is that it does not rely on scalar unobservables and it allows for a more general process for $\omega_t$. Because the measurement error enters the semi-parametric estimator, we need to assume that all of its moments exist and are finite. Sufficient conditions include $\epsilon_t$ being sub-Gaussian or drawn from a compact set. Note that this essentially puts $\epsilon_t$ on equal footing with the independent regressors. In Ai and Chen (2003), it is assumed that all random variables entering the non-parametric function are compactly supported. Therefore, this assumption ensures that both the observed and unobserved variables can enter the non-parametric estimator.

Before showing how Assumption 5 identifies the model, I consider the case where $\epsilon_{it} = 0$. This is the assumption that is made in Ackerberg and Hahn (2015) and allows us to estimate the production function without using a control function or restricting ourselves to an AR(1) process. As such, it is a leading case that is testable under Assumption 5 and helps elucidate the estimation routine. Given instruments $Z_{it}$ such that $E[\epsilon_{it}|Z_{it}] = 0$, the model can be simply estimated using nonlinear least squares (i.e. the second step of OP, LP, and ACF). The key idea is that the first-stage is only necessary to control for the transitory shocks $\epsilon_{it}$ and identification largely comes from the second stage.[6] If there is no $\epsilon_{it}$ to control for, then we can directly use the second stage and get consistent estimates. We can estimate this model less efficiently, but avoid the nonlinear least squares, by estimating the following semi-parameteric model

$$q_{it} \approx f(k_{it}, v_{it}; \beta) + \sum_{|i|=1}^{T} \alpha_i y_{it-1}^{i_1} k_{it-1}^{i_2} v_{it-1}^{i_3} + \eta_{it} \tag{1}$$

which is exact in the limit of $T$, as long as $f$ can be approximated arbitrarily well by a polynomial. This follows because $\omega_{it} = y_{it} - f(k_{it}, v_{it})$ and we have assumed the $g$ can be approximated by a polynomial. This specification can be estimated using regularized 2SLS, following Newey and Powell (2003) or Chen (2007), and identification follows from

[6]This point is made explicit in Wooldridge (2009).

the standard completeness assumption.

If $\epsilon_{it} \neq 0$, then the above estimator is inconsistent. Equation (1) is now written as

$$q_{it} \approx f(k_{it}, v_{it}; \beta) + \sum_{|i|=1}^{T} \alpha_i (y_{it-1} - \epsilon_{it-1})^{i_1} k_{it-1}^{i_2} v_{it-1}^{i_3} + \eta_{it} + \epsilon_{it}$$

This is effectively a nonlinear errors-in-variables model, where output is measured with error $\epsilon_{it}$. In the appendix, I show that under Assumption 5 the terms $\{E[y_{t-1}^{*j}|Z_t]\}_{j\leq J}$ asymptotically span the same space as $\{E[y_{t-1}^{j}|Z_t]\}_{j\leq J}$, where $y^* = f(k_{it}, v_{it}) + \omega_{it}$. If $k_{t-1}$ and $v_{t-1}$ are included in $Z_t$, then we can bring them outside the expectation and write basis function as

$$E[y_{t-1}^{*j} k_{t-1}^{n} v_{t-1}^{m}|Z] = (k_{t-1}^{n} v_{t-1}^{m}) E[y_{t-1}^{*j}|Z]$$

which will imply that $\{E[y_{t-1}^{*j} k_{t-1}^{n} v_{t-1}^{m}|Z]\}_{j\leq J}$ and $\{E[y_{t-1}^{j} k_{t-1}^{n} v_{t-1}^{m}|Z]\}_{j\leq J}$ also span the same space, for any $n$ and $m$.[7] Therefore, we can rely on Frisch-Waugh-Lovell to control for $y_{t-1}^*$ using the variation of $y_{t-1}$ projected onto the instruments. Taken together, we have the following proposition

**Proposition 1.** Assume $\epsilon_{it} \perp (\omega_{it}, Z_{it}, X_{it})$ and $g$ can be approximated arbitrarily well by a polynomial. Then $\beta$ can be consistently recovered using 2SLS. Further, if $g$ is a polynomial,

[7] A basic property of the Hadamard product is that $(A \circ B)C = A \circ (BC)$
which implies that

$$\begin{aligned}(k_{t-1}^{n} \circ v_{t-1}^{m} \circ Y_{t-1}) &= (k_{t-1}^{n} \circ v_{t-1}^{m}) \circ Y_{t-1} \\ &= (k_{t-1}^{n} \circ v_{t-1}^{m}) \circ Y_{t-1}^{*} C \\ &= (k_{t-1}^{n} \circ v_{t-1}^{m} \circ Y_{t-1}^{*}) C\end{aligned}$$

where $Y_{t-1} = E[y_{t-1}^{*j}|Z_t]$. $C$ is a nonsingular, upper-triangular matrix and so $\{E[y_{t-1}^{*j} k_{t-1}^{n} v_{t-1}^{m}|Z]\}_{j\leq J}$ spans the same space as $\{E[y_{t-1}^{j} k_{t-1}^{n} v_{t-1}^{m}|Z]\}_{j\leq J}$ for any $n$ and $m$.

$\gamma$ can consistently estimated as well.

**Proof** see appendix. This results applies more broadly to nonlinear measurement error models and can be used in other dynamic panel settings.

**Wooldridge Formulation** To clarify, we can compare this estimator with the GMM set-up proposed in Wooldridge (2009). In this paper, the author notes that the following equation is sufficient to identify the parameters of the production functions[8]

$$q_{it} = \beta_0 + k_{it}\beta_k + v_{it}\beta_l + \rho_1 (c_{i,t-1}\lambda) + ... + \rho_q (c_{i,t-1}\lambda)^G + u_{it}$$

where $c_{it-1}$ is a polynomial in $k_{it}$ and $m_{it}$, $m_{it}$ being either intermediate input demand or investment. The idea is that $c_{i,t-1}\lambda$ can control for $\omega_{it}$, and therefore will lead to consistent estimates of $\beta_k$ and $\beta_l$. In fact, the function controlling for $\omega_{it}$ is just a polynomial of a polynomial, and therefore can be expanded out to yield a 2SLS estimation. The residual variation in $k_{it}$ and $v_{it}$, after conditioning on $k_{it-1}$ and $m_{it-1}$ identifies the relevant parameters. If we instead replace $c_{it-1}$ with $q_{it-1} - k_{it-1}\beta_k - v_{it-1}\beta_l$, we get the following model

$$q_{it} = \beta_0 + k_{it}\beta_k + v_{it}\beta_l + \rho_1 (q_{it-1} - k_{it-1}\beta_k - v_{it-1}\beta_l) + ... + \rho_q (q_{it-1} - k_{it-1}\beta_k - v_{it-1}\beta_l)^G + u_{it}$$

which, at the true parameter values, equals

$$q_{it} = \beta_0 + k_{it}\beta_k + v_{it}\beta_l + \rho_1 (\omega_{it} + \epsilon_{it}) + ... + \rho_q (\omega_{it} + \epsilon_{it})^G + u_{it}$$

and using the binomial theorem, yields

$$q_{it} = \beta_k k_{it} + \beta_v v_{it} + \sum_{|g|\leq G} \alpha_g \omega_{it}^{g_1} \epsilon_{it}^{g_2} + u_{it} \tag{2}$$

[8] I change the notation slightly to be consistent with this paper. Also note that the parameters of the productivity process are not identified from this equation.

After projecting this onto the space of the instruments and using the assumed independence of $\epsilon_{it}$, we get the following estimating equation

$$E[q_{it}|Z_{it}] = k_{it}\beta_k + E[v_{it}|Z_{it}]\beta_l + \sum_{|g|\leq G} \alpha_g E[\epsilon_{it}^{g_2}]E[\omega_{it}^{g_1}|Z_{it}] \tag{3}$$

Because $E[\epsilon_{it}^{g_2}]$ doesn't depend on $Z_{it}$, it is absorbed into the coefficients $\alpha_g$ and therefore does not confound the estimation of $\beta_k$ and $\beta_l$, and these parameters are identified. A key identifying assumption in Wooldridge (2009) is that

$$E[\epsilon_{it}|k_{it}, v_{it}, m_{it}, k_{it-1}, v_{it-1}, m_{it-1}, ...] = 0 \tag{4}$$

for all $t$. This is clearly weaker than assuming independence of $\epsilon_{it}$, so it is worth considering why we would want to make the much stronger assumption. The entire control function literature is built on the assumption that there exists an input $m_{it}$ (generally either investment or materials) whose demand can be inverted to proxy for $\omega_{it}$. The literature has shown that this assumption holds for perfectly competitive markets, i.e. Levinsohn and Petrin (2003), under Assumption 4. However, recent papers have demonstrated that it does not hold generally when Assumption 4 is violated and when markets are imperfectly competitive.[9] As such, we are not just replacing the weaker assumption (4) with the stronger Assumption 5. We are also replacing the strong Assumption 4. Given the restrictions on potential economic environments that is implicit in the scalar unobservables assumption, making this trade-off appears reasonable.

## 3.1 Estimation Algorithm

Equation (3) suggests the following estimation routine, which reduces to a series of linear regressions. In an initial step, use 2SLS to regress output on a polynomial in the inputs $k_t$

[9] See Gandhi et al. (2020), Flynn et al. (2019), Bond and Söderbom (2005), etc.

and $v_t$, as well as a polynomial in $q_{t-1}$, $k_{t-1}$ and $v_{t-1}$.[10] This consistently recovers estimates of $\beta$ and an estimate of $q_t - f(k_t, v_t; \hat{\beta}^0) \approx \hat{\omega}_t = \omega_t + \epsilon_t$. However, these estimates will tend to have large bias in small samples due to the number of parameters being estimated and the high degree of collinearity between regressors. However, because $\hat{\omega}_t$ is a consistent estimate of $\omega_t + \epsilon_t$, we can use it in a second estimation step to reduce the small sample bias. We now regress $q_t$ on the time $t$ inputs and a polynomial of $\hat{\omega}_t$. This results in a new estimate of $\hat{\omega}_t$, and so the process can be repeated until convergence. At each step we use a consistent estimate of $\hat{\omega}_t$, and so consistency of the entire procedure is ensured. Estimation has the following steps

**Iterated Least Squares**

- Step 0: Regress $q_t$ on the polynomials $p_t = p_K(k_t, v_t)$ and $p_{t-1} = p_J(y_{t-1}, k_{t-1}, v_{t-1})$, using 2SLS
- Step 1: Form the residual $\hat{\omega}_t^m = q_t - f(k_t, v_t; \hat{\beta}^m)$ and generate the polynomial $p_{t-1} = p_J(\hat{\omega}_{t-1})$
- Step 2: Regress $q_t$ on $p_t$ and $p_{t-1}$, using 2SLS, to recover $\hat{\beta}^{m+1}$. Repeat from Step 1 until convergence.

**Discussion** Using Iterated Least Squares avoids the need for nonlinear GMM, which is commonly used in the literature. This can greatly simplify estimation, especially for more complex data generating processes. For instance, when $K = 1$ the model is Cobb-Douglas and nonlinear GMM only requires a search over a two-dimensional parameter space. However, increasing $K$ to 3, requires searching over a ten-dimensional space. This significantly increases the computational complexity of estimation. By comparison, the increasing the size of $K$ does not present any additional computational complexity when using Iterated Least Squares.

[10] See the discussion on choosing instruments at [ZZ].

Step 0 ensures that the initial estimate of $\hat{\omega}_t$ is consistent. Each additional step then uses a new consistent estimate of $\hat{\omega}_t$ during estimation. Iterating in this way can reduce the small sample variance of the estimator. Including $k_{t-1}$ and $v_{t-1}$ allows the moments of $\epsilon_t$ to depend on time $t$ variables, allowing for heteroskedasticity. In this way, full statistical independence is not crucial for the method to work. All we need is to ensure that time $t$ variables provide no information about moments of $\epsilon_{it-1}$ given knowledge of time $t-1$ variables.

Assuming independence allows us to write $E[\epsilon_{it-1}^m | Z_{it}] = E[\epsilon_{it-1}^m]$. Its clear that why we need the moments of $\epsilon_{it}$ to be finite in order for this method to work. If the Markov process $g$ is representable as a polynomial of finite degree then we only need the moments to exist up to the order of the polynomial. However, for semi-parametric estimation we need the degree approximating polynomial to go to infinity, and so we need to ensure that all moments of $\epsilon_{it}$ are finite. A simple assumption is that ensures this is that $\epsilon_{it}$ is bounded or that $\epsilon_{it}$ is sub-Gaussian. This seems reasonable as it is unlikely a firm will report positive or negative output greater than world GDP.

When $f$ is parametric we get $\sqrt{n}$-consistency and asymptotic normality. When $f$ is taken to be non-parametric, the rate of convergence will depend on the intrinsic dimension of $f$. In general, the "control" function $h = E[g|Z_{it}]$ will have a slow rate of convergence, but that won't impact the estimates of $f$.[11] Increasing the dimension of $f$ will increase the arguments that enter the polynomial approximation of $h$, requiring more data for accurate estimation. To see this, note that the number of terms in a complete polynomial is given by

$$\binom{n+d}{d}$$

[11] See the discussion in [ZZ].

where $n$ is the number of variables and $d$ is the degree of the polynomial. With just two factors, this means that a third-degree polynomial will have 20 terms (as we need to include $y$ for a total of 3 regressors). Adding an additional parameter (say an intermediate input $m$) would increase the number of terms to 35. In this sense, the linear estimator does not scale well with dimension. We can ameliorate this somewhat by using regularization or sparse regression techniques. In this way, we can specify a high order polynomial with many terms and then use a LASSO or Elastic Net to select relevant terms. However, by construction, this will only help if $g$ is a high order polynomial, but most of the terms are nearly zero.

## 3.2 Testing Endogeneity

Imposing restrictions on coefficients improves the efficiency of our estimator. However, it may also lead to inconsistent estimates. Consider, for instance, the case where $\epsilon_{it} = 0$ for all $i$ and all $t$. Then the nonlinear least-squares estimator will generally be more efficient than the 2SLS estimator as it restricts the functional relationship between the terms in $h$. Additionally, it treats $y_{it-1}$ as exogenous, and therefore does not rely on instruments to control for the measurement error. We can use a Durbin-Wu-Hausman test to test for the presence of measurement error in the case of Equation (1) or jointly test the validity of the proxy assumptions. In the appendix, I provide Monte Carlo results using a static model to show the performance of the Durbin-Wu-Hausman test. The results indicate that for small sample sizes the test rejects the null hypothesis in too many samples when the null hypothesis is true. For 300 observations, 75% percent of the samples were rejected using an $\alpha$ of 10%. By the time this increases to 30,000%, the null hypothesis is being reject 11% of the time, close to the true significance level.

Testing for $\epsilon_t = 0$ or for the assumptions of the control variable approach is an important step because both of these cases recover estimates of $\omega_{it}$ while the proposed estimator only recovers $\omega_{it} + \epsilon_{it}$. If we want to study an individual markups, then it is necessary

to recover $\omega_{it}$ directly, as firms will choose their input shares based on this component of productivity alone. If we are content with studying an entire industry, then $\omega_{it} + \epsilon_{it}$ may provide sufficient information as averaging over each term will cancel out the $\epsilon_{it}$ in expectation.

## 3.3 Generating Instruments

One advantage of Assumption 5 is that it encourages persistent, unobserved variation in the data generating process for all variables. This is in contrast to the control function approach which requires the presence of an input whose demand is an invertible function of a *single* unobservable, productivity. Therefore, the estimation routine with Assumption 5 can be directly extended to situations with imperfect competition, and in fact, this extension can aid in identification by providing relevant instruments. I discuss several choices of instruments next.

**Input Prices** If input prices were observed and were determined exogenously, then they would be excellent instruments for estimating the parameters in $f$. In the standard control function approach, it would be necessary to include these prices in one's estimate of the control function. This reduces their relevance as instruments because they need to control for their own variation in the estimation routine. Because they do not directly enter the proposed estimator, they remain valid instruments.

**Lagged Inputs** Of course, input prices are rarely observed so various authors have proposed using lags of the the inputs as instruments. The idea is that these lags are orthogonal to the innovation in productivity, but contain information about time $t$ inputs through persistent variation (such as persistence in input prices or the productivity process). Again, unobserved, heterogeneous input prices confound the control function approach but cause lagged inputs to be valid instruments in this approach.

**State-variables of competitors** In an imperfectly competitive market the state-variables of competitors has a direct impact on a firm's output decisions. Therefore, the state-variable of competitors will influence the input demands independently of the innovation in productivity and would make strong instruments. For instance, we can consider instruments akin to Berry et al. (1995) and use the sum of competitor's capital within a market as an instrument for labor demand. In the unlikely scenario that a competitor's investment decision does not depend on the firm's productivity, these would be good instruments to use for 2SLS and would likely be more efficient than using a control function. However, if firm's respond to each other, then competitor's capital will be correlated with a firm's productivity and will therefore not be exogenous. This would then require a timing assumption, such as the competitor's capital is chosen at time $t-1$ as well. Note that when a competitor's capital is a state variable for a firm it needs to be included in the OP/LP/ACF control function. However, Assumption 5 allows us to just include it as an instrument if observed and ignore it if not. I now turn to some Monte Carlo experiments that help elucidate the properties of the proposed estimator.

# 4 Monte Carlos

For the dynamic production problem, I assume a translog production function, similar to my empirical application

$$f(k_t, l_t; \beta) = \beta_l l_t + \beta_{l^2} l_t^2 + \beta_k k_t + \beta_{k^2} k_t^2 + \beta_{kl} l_t k_t$$

Each period, log-output is given by

$$q_{it} = f(k_t, l_t; \beta) + \omega_t + \epsilon_t$$

and $\omega_t$ is assumed to be Markovian in $\omega_{t-1}$. The unanticipated shock is distributed $\epsilon_t \sim$

$N(0, \sigma_\epsilon^2)$. Not that $\epsilon_t$ is independent of the inputs and all moments exist and are finite. Each period, productivity evolves according to

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

and $\eta_t$ is distributed $\eta_t \sim N(0, \sigma_\eta^2)$. I let $g$ be a fourth order polynomial in $\omega_t$[12]

$$\omega_t = \sum_{i=1}^{4} c_i(\omega_{t-1})^i + \eta_t$$

Capital, $k_t$, is chosen at $t-1$, and so is independent of $\eta_t$ while $l_t$ is chosen with knowledge of $\eta_t$. I specify the following nonlinear reduced form equations

$$l_t = h_1(\omega_t, k_t, w_t)$$

$$k_t = h_2(\omega_{t-1}, k_{t-1}, r_{t-1})$$

where $w_t$ and $r_t$ are stochastic (unobserved) input prices. The presence of $w_t$ ensures that the scalar unobservable assumption does not hold. I also assume that $h_1$ is not monotonic in $\omega_t$ so that the equation cannot be inverted. Therefore, this set-up is not consistently estimated by traditional dynamic panel methods or by a control function approach. Table 1 reports results for $n = 5,000$ observations and $N = 50,000$. It estimates the models using both the linear approach, e.g. two-stage least squares, and iterated least squares (ILS). I also estimate the model using an Arellano-Bond estimator for comparison.

[12]I choose the polynomial functional form for simplicity. One might wonder if this overstates the performance of the estimator as $g$ is in the sieve space. I have explored models where $g$ is given by various functions with convergent Taylor series and found nearly identical rates of convergence.

Table 1: Monte Carlo Results: Production Function

| | | Linear | | ILS | | Arellano-Bond | |
|---|---|---|---|---|---|---|---|
| | True Value | Estimate | RMSE | Estimate | RMSE | Estimate | RMSE |
| N = 5,000 | | | | | | | |
| $\beta_l$ | 0.50 | 0.553 | 0.092 | 0.506 | 0.021 | 1.126 | 0.630 |
| $\beta_{l^2}$ | -0.15 | -0.151 | 0.004 | -0.150 | 0.001 | -0.173 | 0.024 |
| $\beta_k$ | 0.50 | 0.466 | 0.056 | 0.491 | 0.023 | 0.131 | 0.370 |
| $\beta_{k^2}$ | 0.25 | 0.257 | 0.009 | 0.252 | 0.004 | 0.299 | 0.056 |
| $\beta_{kl}$ | -0.50 | -0.499 | 0.004 | -0.500 | 0.002 | -0.477 | 0.025 |
| N = 50,000 | | | | | | | |
| $\beta_l$ | 0.50 | 0.504 | 0.023 | 0.501 | 0.007 | 1.164 | 0.664 |
| $\beta_{l^2}$ | -0.15 | -0.150 | 0.001 | -0.150 | 0.000 | -0.178 | 0.028 |
| $\beta_k$ | 0.50 | 0.498 | 0.015 | 0.499 | 0.006 | 0.128 | 0.373 |
| $\beta_{k^2}$ | 0.25 | 0.250 | 0.002 | 0.250 | 0.001 | 0.283 | 0.034 |
| $\beta_{kl}$ | -0.50 | -0.500 | 0.001 | -0.500 | 0.001 | -0.472 | 0.028 |

The linear specification has a much higher RMSE than the ILS estimates, reflecting the larger share of variables included in that regression. These variables tend to be highly collinear, increasing the variance in the linear model's estimates. The linear method also appears to have a higher bias in small samples, but increasing the observations to $50,000$ removed much of the bias. This highlights an important issue for finite-sample estimation. While both estimators display $\sqrt{n}$-consistency, the worst-case error can be substantially different in small sample sizes. Using the linear specification to generate a consistent starting point and then iterating appears to be a promising strategy to reduce small-sample bias. Unsurprisingly, the Arellano-Bond estimator does a poor job of estimating the model because the productivity process is not autoregressive. In the following section, I compare the proposed estimator to standard methods in the production function estimation literature, using an empirical example.

# 5 Empirical Application

In my empirical application, I estimate a production function for the oil pipeline industry using several different methods that are standard in the literature. I then compare the performance of these estimators, noting where they are similar and where they differ. The oil pipeline industry provides an interesting case study as there is significant variation in the regulatory environment of each pipeline. Interstate pipelines are regulated by the Federal Energy Regulatory Commission (FERC) and the Pipeline and Hazardous Materials Safety Administration (PHMSA) at the federal level, but must also adhere to individual state regulations. As such, there is considerable variation in the regulatory environment that each pipeline faces, variation that impacts their production decisions. In principal, we could introduce a separate control function for each combination of states that we observe a pipeline operating in, but we quickly run out of observations to estimate these separate control functions. Therefore, we might be concerned that there is an additional unobservable, i.e. regulatory environment, that would invalidate the control function approach. As always, we are also concerned about the correlation between variable inputs and unobserved productivity as well as the presence of transitory errors.

I use the dataset outlined in Ponder (2021). In this dataset, I see physical units of output - barrel-miles - as well as detailed data on capital stocks and operating expenses used specifically for the transportation of oil. For this exercise, I assume that the production function is translog

$$q_{it} = v_{it}\beta_l + k_{it}\beta_k + v_{it}^2\beta_{ll} + k_{it}\beta_{kk} + k_{it}v_{it}\beta_{lk} + \omega_{it} + \epsilon_{it}$$

and that productivity follows a Markov process that is approached using a polynomial of degree 3.[13] Table (2) presents the results. In column (1), the production function is estimated using OLS. In column (2), the production function is estimated using my preferred

[13]Note that the results presented in this paper differ from those in Ponder (2021) primarily because I assume productivity depends only on its prior value and no other state variables.

approach (PCF). In column (3), I make the assumption that $\epsilon_{it} = 0 \; \forall i, t$ as in Ackerberg and Hahn (2015) (henceforth AH). Finally, column (4) shows the estimates using the control function using operating fuel as the flexible input (henceforth CF).

The parameter estimates using OLS are often quite different from the other estimation routines. For instance, $\beta_v$ is estimated to be almost twice as large as the same coefficient using PCF or CF and the estimate for $\beta_k$ is 60% as large as the other estimators. PCF and CF give largely similar results, with two exceptions. First, $\beta_{v^2}$ is considerably larger using CF and the difference is statistically significant. Similarly, $\beta_{k^2}$ is estimated to be twice as large in absolute terms and to be negative. This ultimate results in a higher estimated average elasticity for the variable input and a lower average elasticity for the fixed input. Interestingly, the estimates under the AH assumption are broadly similar to those using PCF. The two exceptions are for the coefficients $\beta_v$ and $\beta_{k^2}$, however these differences are not statistically significant. In fact, a Durbin-Wu-Hausman test does not reject the null hypothesis that $\epsilon_{it} = 0$ while it rejects the null hypothesis that the CF approach is correctly specified.

It is interesting to note that the control function approach yields the highest average variable elasticity and the lowest average capital elasticity. This is in contrast to the proposed approach, which yields the lowest variable elasticity and highest fixed input elasticity. The higher capital elasticity makes sense in the context of the oil pipeline industry where production is generally capital intensive. However, PCF, AH, and CF all yield an estimated average return-to-scale of 1.1 while OLS estimates a return to scale (RTS) of 1.2. This higher RTS results from a large variable elasticity while the capital elasticity is not significantly attenuated. Pipelines are the classic example of an increasing RTS technology, as doubling their diameter quadruples their capacity. As such, we might expected that the RTS would be larger than 1.1. A potential explanation is that I see more variation in pipeline length than in pipeline diameter in the data. Doubling the length of a pipeline (and

its associate costs) would merely double output when measured in terms of barrel-miles, potentially explaining the low estimated returns to scale.

Table 2: Production Function Estimates

| Estimation Procedure | OLS | PCF | AH | CF |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $\beta_v$ | 1.149 | 0.687 | 0.800 | 0.625 |
| | (0.066) | (0.118) | (0.194) | (0.214) |
| $\beta_k$ | 0.315 | 0.512 | 0.458 | 0.581 |
| | (0.062) | (0.067) | (0.085) | (0.08) |
| $\beta_{v^2}$ | 0.055 | 0.064 | 0.064 | 0.105 |
| | (0.011) | (0.013) | (0.019) | (0.017) |
| $\beta_{k^2}$ | 0.037 | 0.014 | 0.008 | -0.023 |
| | (0.011) | (0.013) | (0.019) | (0.017) |
| $\beta_{vk}$ | -0.122 | -0.074 | -0.068 | -0.045 |
| | (0.021) | (0.026) | (0.039) | (0.035) |
| Avg. Opex Elast. | 0.808 | 0.619 | 0.758 | 0.878 |
| Avg. Capex Elast. | 0.402 | 0.477 | 0.379 | 0.258 |
| Local RTS | 1.210 | 1.096 | 1.137 | 1.135 |
| Observations | 2,295 | 2,295 | 2,295 | 2,295 |

# 6 Conclusion

This paper proposed a new method by which we can estimate production functions that does not depend on the assumption of scalar unobserveables and monotonicity. This estimation routine is internally consistent with a broader range of competitive environments than previous approaches and therefore provides a promising means by which to carry out the cost-minimization approach to recover firm or plant level markups. Crucially, the estimator is no more complex (and in some cases less complex) that other estimators in the literature and has been shown to have good small sample properties. The principal drawback is that the model is only partially identified in the sense that we can consistently

recover the parameters of the production function but not those of the productivity process. However, under additional assumptions, such as assuming that the Markov process governing productivity can be represented as a polynomial, the parameters of $g$ can also be consistently estimated. Additional work must be done to extend this approach to multi-product production, as well as to provide general conditions on $\epsilon_{it}$ under which the model can be consistently estimated.

# References


Ackerberg, D., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Ackerberg, D. and J. Hahn (2015). Some non-parametric identification results using timing and information set assumptions. *Unpublished manuscript*.

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*(6), 1795–1843.

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*(2), 277–297.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Blundell, R. and S. Bond (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric reviews 19*(3), 321–340.

Bond, S., A. Hashemi, G. Kaplan, and P. Zoch (2020). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *National Bureau of Economic Research* (w27002).

Bond, S. and M. Söderbom (2005). Adjustment costs and the identification of cobb douglas production functions. *IFS Working Papers 05/04*.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics 6*, 5549–5632.

Chen, X., H. Hong, and D. Nekipelov (2011). Nonlinear models of measurement errors. *Journal of Economic Literature 49*(4), 901–937.

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

Flynn, Z., A. Gandhi, and J. Traina (2019). Identifying market power in production data. *SSRN* (3358472).

Gandhi, A., S. Navarro, and D. Rivers (2020). On the identification of gross output production functions. *Journal of Political Economy 128*(8).

Griliches, Z. and J. Mairesse (1995). Production functions: The search for identification. *National Bureau of Economic Research* (w5067).

Hausman, J., W. Newey, H. Ichimura, and J. Powell (1991). Identification and estimation of polynomial errors-in-variables models. *Journal of Econometrics 50*(3), 273–295.

Hu, Y. and S. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*(1), 195–216.

Kim, K. i. and A. Petrin (2013). A new control function approach for non-parametric regressions with endogenous variables. *National Bureau of Economic Research* (w16679).

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*(2), 317–341.

Newey, W. and J. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

Olley, S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Ponder, M. (2021). The welfare impact of incentive based regulation: An analysis of the midstream oil market. *Working Paper*.

Schennach, S. (2012). Measurement error in nonlinear models: A review. *cemmap working paper* (CWP41/12).

Schoenberg, I. (1973). *Cardinal Spline Interpolation*. SIAM.

Wooldridge, J. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics letters 104*(3), 112–114.

# A Appendix

## A.1 Proof of Proposition (1)

I focus on the case where the production function is Cobb-Douglas for simplicity. Extending these results to more complex production functions is straightforward. Additionally, I focus on the case where $T$ is fixed as it simplifies the discussion. This method can be extended in a straightforward method when we take $T \to \infty$. If we observed $\omega_{it}$ then we could consistently recover $\beta$ using two-stage least squares using

$$q_{it} = \beta_k k_{it} + \beta_v v_{it} + \sum_{|i|=1}^{T} \alpha_i y_{it-1}^{*i_1} k_{it-1}^{i_2} v_{it-1}^{i_3} + \eta_{it} + \epsilon_{it} \tag{5}$$

where $y_{it-1}^* = \beta_v v_{it-1} + \beta_k k_{it-1} + \omega_{it-1}$. This follows from equation (2) and the definition of $y_{it-1}^*$. By assumption, $E[\epsilon_i | Z_i] = 0$, so regressing $E[y_{it}|Z_{it}]$ on $\{E[v_{it}|Z_{it}]\}$ and $\{E[y_{it-1}^{*j_1} v_{it-1}^{j_2} k_{it-1}^{j_3} | Z_i]\}_j$ consistently recovers all of the model parameters. We want to show that we can get asymptotically identical estimates of $\{\beta_m\}_m$ using only $v_{it}$, $k_{it}$, and $y_{it}$. This involves two steps. First, we show under mild regularity conditions that the space spanned by $\{E[y_{it}^{*j}|Z_{it}]\}_j$ is also spanned by $\{E[y_{it}^j|Z_i]\}_j$. Next, we use Frisch-Waugh-Lovell to argument that partialing out $\{E[y_{it}^j|Z_i]\}_j$ yields asymptotically identical estimates of $\{\beta_m\}_m$ as (5). The consistency of (5) then implies the consistency of the estimator using contaminated regressors.

Given a vector of instruments $Z_{it}$, we have that

$$
\begin{aligned}
E[y_{it}^m|Z_{it}] &= E[(y_{it}^* + u_{it})^m|Z_{it}] \\
&= E\left[\sum_{k=0}^{m}\binom{m}{k} y_{it}^{*m-k}u_{it}^k\Big|Z_{it}\right] \\
&= E\left[\sum_{k=0}^{m}\binom{m}{k} y_{it}^{*m-k}E[u_{it}^k|y_{it}^*, Z_{it}]\Big|Z_{it}\right] \\
&= \sum_{k=0}^{m}\binom{m}{k} E[u_{it}^k]E[y_{it}^{*m-k}|Z_{it}]
\end{aligned}
$$

where the third line follows from iterating expectations and the last line follows from independence. Assuming the moments of $u_{it}$ exist and are finite, the moments $E[y_{it}^m|Z_i]$ are a linearly combination of the moments $E[y_{it}^{*k}|Z_i]$ with $0 \leq k \leq m$. Let $Y = \{1, y_{it}, y_{it}^2, ..., y_{it}^m\}$ and $Y^* = \{1, y_{it}^*, y_{it}^{*2}, ..., y_{it}^{*m}\}$. Then these two matrices are related by the

$$Y = Y^*C$$

where $C$ is $m+1 \times m+1$ and upper triangular. No element on the principal diagonal of $C$ is 0, so $C$ is invertible. This implies that $Y$ and $Y^*$ span the same space which further implies that the projection matrices for $\{E[y_{it}^j|Z_{it}]\}_j$ and for $\{E[y_{it}^{j*}|Z_{it}]\}_j$ are identical. Importantly, this means that the annihilator matrices are identical as well. This concludes the first step.

The second step follows directly, since Frisch-Waugh-Lovell implies that to estimate $\{\beta_m\}_m$ in (5), we can either use OLS directly or we can first partial out $\{E[y_{it}^{*j}|Z_{it}]\}_j$. Let $p_K(x)$ denote the matrix made up of all monomials up to $K$ of $x$. Defined the projection matrix

$$P_{p_K(x)} = p_K(x)(p_K(x)'p_K(x))^{-1}p_K(x)'$$

and the associated annihilator matrix

$$M_{p_K(x)} = I - P_{p_K(x)}$$

Frisch-Waugh-Lovell says that the regression of $M_{E[p_K(y^*_{t-1})|Z]}y$ on $M_{E[p_K(k^*)|Z]}p_J(l)$ produces identical estimates of $\beta$ as OLS. However, we have seen that $M_{E[p_K(y^*_{t-1})|Z]} = M_{E[p_K(y_{t-1})|Z]}$, so using $k$ instead of $k^*$ must also produce consistent estimates of $\beta$. This result is of limited use when full identification is desired, as it does not consistently estimate the constant or the parameters $\{\delta_j\}$. However, if these parameter are unimportant, then we don't need to resort to more complex convolution methods. One such case might be a dynamic panel model.

## A.2 Identifying Productivity Coefficients

Identification of the production function is possible given a fairly flexible process for the transitory shock to output. However, more structure is required to identify the evolution of the productivity process. We can identify all the model parameters if the evolution of productivity is given by a polynomial function and the transitory errors are independent of a set of instruments. We can jointly estimate the moments of the transitory shocks and the polynomial coefficients governing the evolution of productivity. The proof of this fact follows closely from the work of Hausman et al. (1991), however I relax the assumption that the instrumental variable equation is causal. Note that

$$E[\omega_t + \epsilon_t|Z] = \sum_{n=0}^{N} \gamma_n E[\omega^n_{t-1}|Z]$$

and

$$E[\omega_t + \epsilon_t | Z] = \sum_{n=0}^{N} \delta_n^1 E[(\omega_{t-1} + \epsilon_{t-1})^n | Z] \tag{6}$$

$$= \sum_{n=0}^{N} \delta_n^1 E[\sum_{k=0}^{n} \binom{n}{k} \sigma^{n-k} \omega_{t-1}^k | Z] \tag{7}$$

Rearranging terms yields

$$\sum_{n=0}^{N} \sum_{k=n}^{N} \binom{k}{n} \sigma^k \delta_{k-n}^1 E[\omega_{t-1}^n | Z]$$

This implies the coefficients of regression [ZZ] satisfy

$$\gamma_n = \sum_{k=n}^{N} \binom{k}{n} \sigma^k \delta_{k-n}^1, \ n = 0, ..., N$$

There are $N + 1$ equations and $2N$ unknowns (where we have used the fact that $\sigma^0 = 1$ and $\sigma^1 = 0$ by assumption). We can further consider the regression equations

$$E[(\omega_t + \epsilon_t)\omega_{t-1} | Z] = \sum_{n=1}^{N+1} \gamma_n E[\omega_{t-1}^n | Z]$$

and

$$E[(\omega_t + \epsilon_t)(\omega_{t-1} + \epsilon_{t-1}) | Z] = \sum_{n=0}^{N+1} \delta_n^2 E[\sum_{k=0}^{n} \binom{n}{k} \sigma^{n-k} \omega_{t-1}^k | Z]$$

By a similar argument, we have

$$\gamma_{n-1} = \sum_{k=n}^{N} \binom{k}{n} \sigma^k \delta_{k-n}^2, \ n = 0, ..., N$$

where $\gamma_{-1} = 0$. This provides an additional $N + 1$ equations. However, note that two of

these equations are asymptotically equivalent to two of the previous ones, because both $\delta_N^1 = \delta_N^2$ and $\delta_{N-1}^1 = \delta_{N-1}^2$. Therefore, we add $N-1$ new equations and are only just identified. This process may be repeated for higher moments, e.g. we can next run the regression

$$E[(\omega_t + \epsilon_t)\omega_{t-1}^2|Z] = \sum_{n=1}^{N+1} \gamma_n E[\omega_{t-1}^n|Z]$$

and use the subsequent $N-1$ conditions to over-identify the parameters. However, there is a trade-off between adding over-identifying restrictions and the reduction in precision that results from estimating these higher-moment equations.[14]

## A.3 Estimating Conditional Expectations

The proposed estimation routine requires estimating conditional expectations of endogenous variables given the value of instruments, e.g. $E[l_t^i|Z_t]$. To this end, I use truncated, complete polynomial series in $Z_t = \{Z_t^k\}_k$ to approximate this conditional expectation. I project the endogenous variables[15] onto the polynomial in $Z_t$, so that

$$E[x|Z_t] = \sum_j \hat{\gamma}_j E[p_j(Z)|Z] = \sum_j \hat{\gamma}_j p_j(Z) \tag{8}$$

where $j$ represents a multi-index. Rather than use raw polynomials, I use orthogonal polynomials to reduce the collinearity between the various terms. Specifically, I construct a basis using complete polynomials and then use the Gram-Schmidt procedure to orthgonalize the basis for $Z$. Additionally, I use Tikihonov regularization to ensure that the function approximating $g$ resides in a compact space. That is, I add a penalty term $\lambda$ when estimating

[14]To understand why, note that the new error terms are now given by $\eta_t\omega_{t-1}^2$ in the "true" regression. So we increase the variance of the error term while the variance of the true regressors does not increase.

[15]The exogenous variables, such as $k_t$ are included in the set of instruments and therefore have a conditional expectation equal to themselves.

(8) that penalizes the magnitude of $\hat{\gamma}_j$. This leads to two hyper-parameters that need to be chosen: $\lambda$ and $T$, the degree of the polynomial. Both of these hyper-parameters are chosen by cross-validation. A typical cross-validation scheme would choose leave out a fraction $x$ of the sample and estimate the model on the remaining $1-x$ share of observations. It would re-estimate the model for each given choice of hyper-parameters and then choose the pair that has the best out-of-sample predictive power. This works well for the Cobb-Douglas production function, because the model is essentially linear and there are few local minima in the GMM objective function. Unfortunately, this runs into problems using a translog production function due to the larger number of local minima. We essentially need to do a careful search over all the local minima to identify the global minimum. This is a time consuming process that is prone to error. Instead, I first estimate the model using a $\lambda = 0$ and $T = 3$. This provides a consistent, albeit inefficient, initial estimate of $\hat{u}_t \approx \omega_t + \epsilon_t$. I then use directly use cross-validation on the $\hat{u}_t$ to select optimal hyper-parameters. This approach requires less nonlinear minimization (only the initial consistent estimates require nonlinear minimization), and therefore is more computationally reliable. Because $\hat{u}_t$ is projected on $\hat{u}_{t-1}$ using $Z_t$ as instruments, this effectively reduces to cross-validating a 2SLS estimate. As new hyper-parameters are chosen, one can repeat the process to generate more internally consistent estimates. However, there is no guarantee that this process will converge.

## A.4 Basis Functions

The proposed approach is not limited to polynomial basis functions. In fact, better small-sample performance can often be attained using B-splines. A classic result of Schoenberg (1973) is that the cardinal B-splines of order $L$ with equidistant knots can perfectly reproduce any polynomial of degree $l = 0, ..., L-1$. This implies that the set of basis functions given by

$$\tilde{B}_{k,l}(x) = \int B_{k,l}(x + \epsilon) dF(\epsilon)$$

can uniformly approximate any continuous function over a compact set. We have established $\{\int (x+\epsilon)^l dF(\epsilon)\}_{l=0}^{L-1}$ spans the same space as $\{x^l\}_{l=0}^{L-1}$. The Weierstrass Approximation Theorem says that any continuous real-valued function with compact support can be uniformly approximated by polynomials. Since any polynomial can be expressed in terms of the functions $\tilde{B}_{k,l}(x)$, this implies the expected value of the B-splines can uniformly approximate any continuous function as well.

## A.5 Monte Carlos

### A.5.1 A Static, Separable Model

To fix ideas, I start with a simpler problem. Here, production is separable in $l_i$ and $k_i$, but capital is measured with error. That is

$$y_i = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{1i}^2 + \sum_j \delta_j x_{2i}^{*j} + \epsilon_i$$

$$x_{2i} = x_{2i}^* + u_i$$

and $u_i$ is independent of all regressors. I assume that both $x_{1i}$ and $x_{2i}^*$ are determined by the random vector $z_i$, such that

$$x_{1i} = \sum_k \gamma_{1k} z_i^k + \eta_{1i}$$

and

$$x_{2i}^* = \sum_k \gamma_{2k} z_i^k + \eta_{2i}$$

with $z_{ik} \sim N(\mu_k, \sigma_k^2)$. Some elements of $z_i$ are observed and can therefore act as instruments for $x_{2i}^*$. In this set-up, $x_{1i}$ is exogenous and therefore can act as its own instruments. I estimate the model assuming $u_i = 0$ and $u_i \neq 0$. First, I assume that $u_i = 0$ and use

Table 3: Monte Carlo Results: Measurement Error

| | | $u_i = 0$ | | $u_i \neq 0$ | |
|---|---|---|---|---|---|
| | True Value | OLS | 2SLS | OLS | 2SLS |
| N = 300 | | | | | |
| $\beta_1$ | 0.50 | 0.504 | 0.504 | 0.586 | 0.523 |
| $\beta_2$ | -0.15 | -0.149 | -0.149 | -0.173 | -0.166 |
| DWH (%) Reject | | | 75% | | 93% |
| N = 3,000 | | | | | |
| $\beta_1$ | 0.50 | 0.501 | 0.501 | 0.597 | 0.506 |
| $\beta_2$ | -0.15 | -0.150 | -0.150 | -0.174 | -0.154 |
| DWH (%) Reject | | | 30% | | 99% |
| N = 30,000 | | | | | |
| $\beta_1$ | 0.50 | 0.500 | 0.500 | 0.596 | 0.501 |
| $\beta_2$ | -0.15 | -0.150 | -0.150 | -0.172 | -0.150 |
| DWH (%) Reject | | | 11% | | 99% |

both OLS and 2SLS. Both methods recovers all of the model parameters. Next, I assume that $u_i \neq 0$ (I draw it from $N(0, \sigma_u^2)$). I estimate the model using OLS and I estimate the model using 2SLS. For instruments, I assume that $z_{i1}$ is observed and $x_{1i}$ instruments for itself. I generate a matrix of instruments by generating a complete polynomial in $x_{1i}$ and $z_{ik}$ of degree $d$. Then I regress two additive polynomials of $x_{1i}$ and $x_{2i}$ on the matrix of instruments to generate the first-stage predicted values. I further regress the dependent variable $y_i$ on the instrument matrix to generate $E[y_i|Z_i]$. Finally, I do OLS on these predicted matrices to generate consistent estimates of $\alpha_1$ and $\alpha_2$. Table 3 presents the results. Note that OLS is efficient when $u_i = 0$ but inconsistent when $u_i \neq 0$. The 2SLS estimates are inefficient but consistent in both cases. Therefore, I run a Durbin-Wu-Hausman test on the coefficients $\alpha_1$ and $\alpha_2$ to test for the presence of measurement error, using 2 degrees of freedom and $\alpha = 0.1$.

The estimates are not very accurate for small sample sizes. Intuitively, the span of the basis functions only coincide asymptotically, so there will always be a finite sample bias. When the sample size is small, this bias can be quite large. As we increase the sample size from 300 to $30,000$ we see standard parametric rates of convergence. The Durbin-Wu-Hausman test rejects the null hypothesis of no measurement error for too many samples. Three-quarters are incorrectly rejected when $N = 300$ and this doesn't come down to roughly 10% until $N = 30,000$. Clearly the asymptotic covariance matrix is not well approximated when $N$ is small. Conversely, the test almost always rejects the null hypothesis when this is in fact measurement error.