

Estimating Multi-Product Production Functions: What Can We Learn Without Demand Assumptions?

Nicole Scholz *

October 30, 2025

Abstract

I prove that, when the demand side is unrestricted, production functions for multi-product firms are unidentified, except in population if the conditional time-series variance of inputs is unbounded. I develop a novel identification strategy that does not rely on demand-side assumptions. Instead, by imposing the weaker assumption that the productivity distribution is in a stationary equilibrium, I show that the production function parameters are set-identified. Using simulations I show that the estimator is robust to non-stationarity of the productivity processes in short panels and provide evidence that the identified set is highly informative about the data-generating parameters. My approach avoids the need for instruments or numerical solvers, providing a widely applicable method for estimation.

*University of Warwick; nicole.scholz@warwick.ac.uk

1 Introduction

Multi-product firms are central to modern economies, producing the vast majority of manufacturing output despite making up a minority of firms. In the US they represented 37% of firms between 1987 and 1997, yet produced 87% of manufactured goods (Bernard et al., 2010) with similar figures observed in Europe (see, e.g., Dhyne et al., 2022). Their importance in international trade is even more striking with multi-product firms being responsible for 99% of US exports in 2000 (Bernard et al., 2007).

However, our tools for estimating productivity, markups, and markdowns – which have recently received increasing interest in the IO-labour literature (e.g. Autor et al., 2020; Montag, 2024) – fail when input allocations within multi-product firms are unobserved, as is the case for most datasets.

The literature has taken two paths. One approach estimates production frontiers: By relating total inputs to the feasible set of outputs, they are flexible enough to allow for joint production where inputs are shared across product lines. Yet this theoretical flexibility is not matched in practice; the functional forms typically used for estimation cannot accommodate non-joint production unless all products are produced by the same constant elasticity of substitution (CES) technology (Cairncross et al., 2023). With a lack of methods allowing for more flexible technologies, this effectively makes jointness an assumption.

This paper focuses on the second path: Estimating product-specific production functions. This approach, which imposes non-jointness, is highly attractive for its tractability in policy counterfactuals, but forces the econometrician to confront the unobserved input allocation problem directly. Several solutions have been developed, most notably by Orr (2022) and Valmari (2023) whose approach uses the firm’s first-order conditions to back out input allocations. However, the validity of existing methods relies on overly restrictive productivity processes, or often untestable assumptions on the market environment, firm conduct, and the specific form of demand.

These assumptions are likely to induce misspecification bias as economists have reached starkly different conclusions about conduct by making slightly different demand assumptions, even in the same industry in the same country at the same time. For example, both Miller et al. (2021) and De Loecker and Scott (2024) study the

US beer industry using the demand estimates of [Miller and Weinberg \(2017\)](#). While [Miller et al. \(2021\)](#) construct a model of price leadership and strongly reject Bertrand pricing, [De Loecker and Scott \(2024\)](#) focus on the vertical brewer-supermarket relationships and find that Bertrand pricing with a competitive retail sector agrees with their production-based markup estimates. In their Table 1, [De Loecker and Scott \(2024\)](#) further show that marginal cost estimates can vary widely under different ownership matrices. Even beyond demand conduct, the basic assumption of profit maximisation has been challenged in the wake of better data on managerial expectations and decision-making ([Keiller et al., 2024](#)).

This paper confronts this identification challenge, examining its extent and providing an alternative approach. My first contribution is to demonstrate that the existing estimators critically rely on profit maximisation and demand-side assumptions; without them, in any finite sample, all parameter vectors can be rationalised and identification is only achieved 'at infinity', when time-series variation in inputs becomes arbitrarily large. Building on this insight, my main contribution is to provide a novel identification strategy that is based on the weaker assumption of stationarity of the productivity distribution. My approach is easily implementable using moment inequalities and delivers set identification. The identified set converges to the data generating vector of parameters as the time-series variation in inputs increases. As the identification method does not rely on profit maximisation, the identified set can be used to test auxiliary assumptions on demand and conduct.

While there has been no direct empirical study of the stationarity of productivity, there is evidence by [Jaimovich et al. \(2023\)](#) that firm revenues are reasonably approximated by a stationary AR(1) distribution. The assumption also has precedence in the economics literature, having been used by [Blundell and Bond \(2000\)](#) to estimate production functions for single-product firms, and is more likely to hold for mature industries ([Hopenhayn, 1992](#)).

I use simulations to show that the moment inequalities provide meaningful restrictions on the set of potential data-generating parameters under realistic input distributions and are robust to non-stationarity of the productivity process in short panels. I show that the identified set shrinks as periods are added, making the estimator partic-

ularly useful with long horizons or high frequencies of observation.

This paper significantly expands the toolset for applied economists interested in estimating productivity, markups and markdowns. Of course, whether stationarity or knowledge of conduct and demand structure are more plausible in any given setting is an empirical question. For example, superconductor and computer chip manufacturing is rapidly evolving, making stationarity unlikely to hold. On the other hand, carbonated soft drinks and beer, or the refinement and sale of petroleum are established industries where economists disagree about the underlying conduct, making stationarity preferable to relying on demand side assumptions.

The next section situates this paper within the existing literature. Section 3 presents the formal assumptions underlying the identification results, while section 4 examines identification under an unrestricted demand side. Section 5 employs stationarity of the productivity distribution to deliver set identification. Section 6 extends these results. In section 7 I use simulations to show that the identified set places meaningful restrictions on the potential data generating production function parameters, and section 8 concludes.

2 Related Literature

Production function estimation is an essential tool to measure productivity and output elasticities; it underpins research in fields from macroeconomics and international trade to industrial organisation and labour economics. Seminal contributions by [Olley and Pakes \(1996\)](#), [Levinsohn and Petrin \(2003\)](#), and [Akerberg et al. \(2015\)](#) established the control function approach to address the endogeneity of firms' input choices to their productivity, which is unobserved by the econometrician. This methodology has since enabled a vast literature studying topics from aggregate productivity and resource misallocation to market power. These tools are also at the core of a growing literature on firms' input market power, enabling the estimation of labour markdowns and monopsony power (e.g. [Autor et al., 2020](#); [Montag, 2024](#)).

However, there has been a foundational challenge to the widespread application of these methods to measure markups. While the theory is based on output elasticities,

most datasets only report revenues. This issue is at the heart of a recent debate in the literature. [Bond et al. \(2021\)](#) show that, when deriving markups from revenue elasticities, researchers should mechanically find that firms are pricing at marginal cost. [De Ridder et al. \(2024\)](#) confirm that markup levels are not identified with revenue data, but show that [Bond et al.’s \(2021\)](#) result only holds when all firms charge the same markup. Whenever there is heterogeneity in markups, their trends are well-approximated.

Similarly, a large and growing literature has emphasized the importance of physical productivity (TFPQ) over revenue-based productivity (TFPR). Since TFPR bundles technical efficiency with demand and price variation, it can severely distort inference about firm performance, heterogeneity, and resource allocation. For example, [Foster et al. \(2008\)](#) show that TFPR-based measures overstate the relative role of productivity in driving firm survival, instead finding demand to be the key determinant, while [Hsieh and Klenow \(2009\)](#) show that between-firm resource misallocation can reverse the positive correlation between revenue- and quantity-based TFP measurements.¹ Together, this evidence demonstrates the importance of tools to estimate quantity-based production functions to recover economically meaningful measures of technical efficiency.

The shift towards physical output data, however, introduces additional econometric challenges: Most datasets only contain total inputs at the firm level; their split across product lines is typically unobserved. The recent literature has diverged into two main approaches to address this issue, either estimating production frontiers to avoid the input allocation problem or imposing additional assumptions to recover product-level inputs.

[Dhyne et al. \(2022\)](#) estimate separable production frontiers by extending the control function approach of [Akerberg et al. \(2015\)](#) to the multi-product setting. This makes their method hard to scale as it requires one flexible input per output. [Caselli et al. \(2025\)](#) instead use the firm’s first-order conditions to back out productivities after performing a demand estimation. While this allows them to separate product quality and productivity, it also opens them up to misspecification bias. They also have to restrict themselves to separable CES frontiers. Separability is a strong assumption, as

¹They find that Chinese exporting plants have 46% higher TFPQ but 13% lower TFPR than other Chinese plants, likely reflecting preferential treatment by Chinese authorities, while US exporters have similar TFPQ advantages over their US competitors, but also have 6% higher TFPR.

shown by [Cairncross et al. \(2023\)](#). They prove that the frontier derived from non-joint production functions is only separable if all goods are produced using the same CES production function. This limits the range of questions that can be answered using existing production frontier methodologies.

On the other hand, several papers have developed methods to estimate production functions. [De Loecker et al. \(2016\)](#) assume that there are no returns to scope and no within-firm productivity differences. This allows them to use single-product firms to estimate production functions and impute input allocations for multi-product firms in a final step. [Gong and Sickles \(2021\)](#) also restrict the productivity process, assuming that productivity can be decomposed into a firm-product fixed effect that is constant across time, and a time fixed effect that is constant across firms and products. Neither assumption is likely to be realistic.

More recent papers instead recover input allocations from the demand side. [Orr \(2022\)](#) does this sequentially by first running a demand estimation and then using the marginal costs to recover inputs under the assumption that the technology is homogeneous and common across all products. Meanwhile, [Valmari \(2023\)](#) estimates the production function and demand-side parameters simultaneously under the assumption that firms operate in monopolistically competitive markets. [Chen and Liao \(2022\)](#) follow a similar approach to [Valmari \(2023\)](#) with slightly different assumptions on demand. They also show that the production function is non-parametrically unidentified when input allocations are unobserved.

I contribute to this literature in several ways. First, I show that, when the demand side is left unrestricted, the production function is unidentified except in population as the time-series variance of inputs goes to infinity. This extends the non-identification result of [Chen and Liao \(2022\)](#). Second, I provide a novel set-identification result that uses stationarity of the productivity distribution while leaving the productivity process unspecified. This result contributes to the literature estimating production frontiers as it provides another way to estimate production without needing to recover input allocations. It also contributes to the literature estimating production functions by providing alternative assumptions with which to estimate the production parameters.

As the set-identification result leaves the demand side unspecified, it can also be used to construct a test of demand conduct by checking whether the parameter estimates recovered under a specific set of demand assumptions, like those of [Valmari \(2023\)](#), lie within the identified set. This jointly tests the demand assumptions under the null that only the assumptions made in this paper are correct. As such, this paper adds to a large literature on conduct testing (e.g. [Backus et al., 2021](#); [Magnolfi and Sullivan, 2022](#); [Duarte et al., 2024](#)).

Finally, this paper contributes to a growing literature using moment inequalities for identification. In industrial organisation, [Pakes et al. \(2025\)](#) study the US healthcare insurance market utilising consumers’ binary switching choices. Their implementation adopts a Bayesian approach following [Kline and Tamer \(2016\)](#), which is part of a large econometric literature on inference for partially identified models. Two other papers from this literature are particularly noteworthy. First, [Kaido et al. \(2019\)](#) provide a method for constructing confidence intervals for individual elements of the parameter vector. Second, [Romano et al. \(2014\)](#) show how to incorporate information on the set of unsatisfied moment inequalities to increase the estimator’s power. Both methods can be adapted to implement the moment inequalities in this paper.

In the next section, I will discuss the theoretical framework before I formally establish the identification results in the following two sections.

3 Theoretical Framework

3.1 Notation

Script letters will denote sets, while arrows denote vectors and lowercase letters denote the natural logarithm of their upper case counterparts. For example:

- \mathbb{X} is the set of inputs;
- \vec{X}_t , a vector, is the realisation of inputs at time t , and $X_{a,t}$ is the a -th element of \vec{X}_t ;
- \vec{x}_t is the vector of the log of realised inputs at time t , $(x_{1,t}, x_{2,t}, \dots, x_{\dim(\vec{X}),t})$.

Similarly, $f(\cdot) = \ln(F(\cdot))$ for any function $F(\cdot)$.

For an index range $a \leq b$ I write $\vec{X}_{b:a}$ for the history of elements from period a to period b . For example, $\vec{X}_{t-1:0} = \{\vec{X}_{t-1}, \vec{X}_{t-2}, \dots, \vec{X}_0\}$ is the entire history of total inputs until the previous period.

$g_{a|b}(\cdot)$ will denote the probability density function of a conditional on b . So $g_{\vec{X}|t}(\cdot)$ is the joint probability density of total inputs in period t .

Additionally, I will use an asterisk to denote the true value of a variable. Therefore, $\vec{\theta}^*$ will denote the true vector of production function parameters and \vec{X}^* the true input vector.

I will use $\text{ess sup}_{Z \in \mathcal{Z}} F(Z)$ to denote the essential supremum of $F(\cdot)$, that is the supremum over all but a measure zero subset of Z . ess inf , the essential infimum, is similarly defined.

3.2 Production

In each of $T + 1$ periods, J firms² sell any number of I differentiated products.^{3,4} Each product is produced using the common production technology

$$Y_{i,j,t} = F(\vec{\theta}^*, \vec{X}_{i,j,t}) e^{\beta_0^* + \omega_{i,j,t}^*}, \quad (1)$$

where i denotes a product, j denotes a firm, t denotes the period, and $Y_{i,j,t}$ denotes the quantity of good i produced by firm j in period t . Taking Hicks-neutral productivity, e^{β_0} , out of the production function means that I am normalizing $F(\cdot)$. Without loss of generality I will assume $F(\vec{\theta}, \vec{1}) = 1$ for all $\vec{\theta}$.

I will assume that the researcher observes the entire distribution of possible realisations of observables, that is, the limiting case as $J \rightarrow \infty$. Also, I will assume that there are M inputs and two outputs. The results are easy to generalise to $I > 2$ outputs, but are restricted here to ease notation. The rest of this section will further clarify the

²While I will be using 'firm' throughout the text to be consistent with the literature, the assumptions are most likely to hold at the plant level.

³Using this terminology, Kellog's Frosted Flakes and Cap'n Crunch would be the same differentiated product. The exact definition of a product is up to the researcher and, generally, a function of the data.

⁴One could alternatively define v markets in which J_v firms sell any number of I differentiated products. Then observing a large number of markets is equivalent to observing a large number of firms in my framework.

underlying assumptions of the model.

Assumption 1 (Homogeneous production technology). The production technology, $F(\cdot)$, is continuous and differentiable, equal to zero if any of its arguments are equal to zero, strictly increasing in all arguments, quasi-concave, and homogeneous of degree $\phi > 0$.

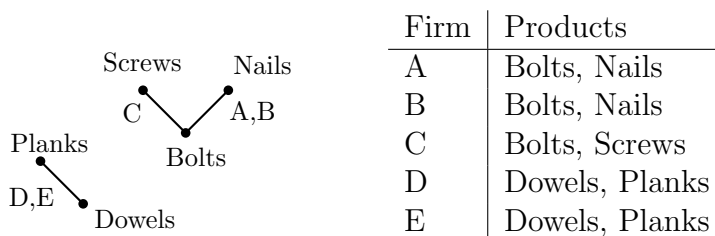
For ease of exposition and to simplify and drive intuition for the main results, I will strengthen assumption 1 in the main body of this paper, replacing it with assumption 1a. I will discuss the generalisation of the set-identification result in section 6 and of the identification at infinity result in section A.

Assumption 1a (Cobb-Douglas technology). The production technology is Cobb-Douglas with $\vec{\theta} = (\beta_{X_1}, \dots, \beta_{X_M})$ and returns to scale $\phi = \sum_{m=1}^M \beta_{X_m}$,

$$F(\vec{\theta}, \vec{X}) = \prod_{m=1}^M X_m^{\beta_{X_m}}.$$

Assumption 2 (Common technology). The production technology, $F(\cdot)$, is the same for all products produced in a firm and for all firms producing the same product. This is also true of β_0 . However, production within a firm and across firms can differ as a result of differences in the product-firm-time-specific productivity, $\omega_{i,j,t}$.

Figure 1: Production network



Nodes represent products. There is an edge between two nodes if at least one firm produces both of these products.

Assumption 2 appears more restrictive than it is in practice. It does not necessitate that every product in the dataset is produced by the same production function. Take the economy in figure 1. Assumption 2 means that screws, nails, and bolts (produced by firms A-C) share one production function, $F_{snb}(\cdot)$, and planks and dowels (produced

by firms D-E) share another production function $F_{pd}(\cdot)$. Assumption 2) does not have any implication on the relationship between these production functions. Orr (2022) also points out that, in practice, most datasets only have enough observations to estimate production functions at the level of a sector or industry, so that this assumption would have to be made implicitly regardless. Either way, the existence of plants that produce across different sectors and industries is an empirical question.

Assumption 3 (No measurement error). There is no measurement error in the observed output, $Y_{i,j,t}$.

To be able to solve for the input allocation using demand information Orr (2022) and Valmari (2023) must assume the absence of measurement errors in outputs, which introduce another $I \cdot J \cdot (T + 1)$ unknowns. This increase in the degrees of freedom is also the reason I am making this assumption here, although it can be generalised for the main result as discussed in section 6.1. Similarly, assumption 4 is important to avoid having to deal with a set of hidden state variables and costs.

Assumption 4 (Costless transfer of inputs). Inputs can be costlessly transferred between production lines.

Assumption 5 rules out that the marginal product of any good depends on how much of any other good is produced. It is the key differentiating assumption between the production function and production frontier literature.⁵

Assumption 5 (Non-joint production). The set of inputs does not contain any products produced within the same firm. Additionally, total inputs are completely attributable to product lines. That is, for every firm, j , time, t , and input, m , the true inputs can be recovered by allocating the observed total inputs:

$$X_{m,i,j,t}^* = S_{i,j,t}^m X_{m,j,t},$$

where $S_{i,j,t}^m \in [0, 1]$ and $\sum_i S_{i,j,t}^m = 1$.

⁵While this distinction is generally true for the existing literature, it is possible to model joint production explicitly in a production function framework. An example of this is the model of production with shared inputs described in chapter 15 of Baumol et al. (1982) and adopted by Cairncross et al. (2023) in their appendix.

I will denote the set of pre-determined inputs as \mathbb{K} , the set of inputs that can be adjusted subject to adjustment costs as \mathbb{L} , and the set of static inputs as \mathbb{M} .

Assumption 6 (Static cost minimisation). Firms choose total quantities of static and dynamic inputs, as well as input allocations to solve the static cost minimization problem:

$$\min_{\{X_{m,1,j,t}, \dots, X_{m,I,j,t}\}_{m=1}^M} \sum_{X_m \in \mathbb{M}} P_{j,t}^{X_m}(X_{m,j,t}) X_{m,j,t} + \sum_{X_m \in \mathbb{L}} C_{j,t}^{X_m}(X_{m,j,t}, X_{m,j,t-1}) \quad (2)$$

subject to

$$\begin{aligned} \sum_i X_{m,i,j,t} &= X_{m,j,t} \quad \forall m \\ X_{m,i,j,t} &\geq 0 \quad \forall m, i \\ Y_{i,j,t} &= F(\vec{\theta}^*, \vec{X}_{i,j,t}) e^{\beta_0^* + \omega_{i,j,t}^*} \quad \forall i, \end{aligned}$$

where $P_{j,t}^{X_m}(\cdot)$ is a price schedule, which may be constant,⁶ and $C_{j,t}^{X_m}(\cdot)$ is the adjustment cost for input m – for example hiring and firing costs – both of which are known to the firm.⁷

These assumptions allow me to use a simplifying result from [Orr \(2022\)](#). [Lemma 1](#) clarifies that a consequence of having a common and homogeneous production technology is that input shares are independent of the input – if a firm uses 30% of its materials in the production of a product it will also use 30% of any other input. The proof is contained in appendix A of [Orr \(2022\)](#).

Lemma 1 (Lemma 1 – [Orr \(2022\)](#)). *If assumptions 1-2 and 4-5 hold then there exists a solution to the firm's conditional cost minimization problem satisfying $X_{m,i,j,t} = S_{i,j,t} X_{m,j,t} \quad \forall X_m \in \mathbb{X}$, where $S_{i,j,t} \in [0, 1]$ and $\sum_i S_{i,j,t} = 1$.*

4 Identification at Infinity

For the purpose of this section I will restrict the class of productivity processes to be auto regressive of order 1 (AR(1)). While most identification treatments only assume a

⁶A constant price schedule means that the firm has no market power.

⁷The particular form of the law of motion for dynamic inputs does not change the results in this paper. So it is possible to accommodate functional forms like that of [Orr \(2022\)](#).

general first order Markov process,⁸ most empirical applications use an AR(1) process. Imposing more structure also makes identification easier to achieve. That is, if $\vec{\theta}$ is not identified under the more restrictive AR(1) assumption, it is also not identified under the more general assumption of a first order Markov process.

Assumption 7 (AR(1) productivity).

$$\omega_{i,j,t} = \rho_i \omega_{i,j,t-1} + \zeta_{i,j,t},^9$$

where $\rho \in (0, 1)$ and the random error, $\zeta_{i,j,t}$, is independently distributed across firms and time, but may be correlated across products within a firm.

It is also important to clarify the timing of input choices, and whether the firm has input market power.

Assumption 8 (Timing of firm decisions). The firm chooses the total amount of each pre-determined input, $X_{m,t}$ for all $X_m \in \mathbb{K}$, before observing productivity, $\omega_{i,j,t}^*$. All other inputs are chosen or can be adjusted subject to costs after observing productivity.

The firm may also buy any number of its static inputs on markets where it does not have market power. These do not have to be perfectly competitive markets. For example, it could be a market where a monopolistic supplier faces a continuum of producers. [Orr \(2022\)](#) and [Valmari \(2023\)](#) need at least one such input to exist to recover marginal costs from demand. I allow the set to be empty.

Definition 1 (Input market power). Denote the set of inputs bought on markets without buyer market power as \mathbb{C} . That is, $\forall X_m \in \mathbb{C}, P_{j,t}^{X_m}(X) = P_{j,t}^{X_m}(X') \forall X, X' \in \mathbb{R}_+$. Denote the vector of input prices for inputs in \mathbb{C} by $\vec{P}_{\mathbb{C},t} = \{P_{X_m,t} : X_m \in \mathbb{C}\}$.

With this information we can define the set of variables that the innovation in productivity, $\zeta_{i,j,t}$, needs to be independent of, the information set at time t .¹⁰

$$\vec{Z}_t = (\vec{K}_t, \vec{X}_{t-1:0}, \vec{P}_{\mathbb{C},t:0}, \vec{Y}_{t-1:0}).$$

⁸Productivity is first order Markov if $\omega_{i,j,t} = E[\omega_{i,j,t} | \omega_{i,j,t-1}] + \zeta_{i,j,t}$, where $\zeta_{i,j,t}$ is mean-zero and independent across firms and time.

⁹Note that, if the productivity distribution is in its stationary equilibrium, β_0 could be absorbed into the productivity distribution instead with $\omega_{i,j,t} = \beta_0(1 - \rho) + \rho\omega_{i,j,t-1} + \zeta_{i,j,t}$.

¹⁰The exact nature of these variables are unimportant for the identification result. As long as they are known to be unrelated to the true innovation in productivity, $\zeta_{i,j,t}$, any variable can be included. Of course, to be useful, they need to be correlated with the firm's inputs.

Further, denote its support by $\mathbb{Z}_t = \{\vec{Z} : g_{\vec{K}_t, \vec{X}_{t-1,0}, \vec{P}_{C,t,0}, \vec{Y}_{t-1,0}|t}(Z) > 0\}$.

Lastly, I will introduce two assumptions to simplify the problem. As I explain in more detail in section A in the appendix, in the absence of these assumptions there are cases where the production function is neither always identified nor always unidentified.

Assumption 9. All moments of the conditional log input differences exist and are bounded. For every $\rho \in (0, 1)$

$$\sup_{i,t} \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} E \left[(s_{i,t}^* - \rho s_{i,t-1}^*)^d | \vec{Z}_t \right] < \infty, \quad \forall d \in \mathbb{N},$$

where $\mathbb{Z}_t = \{\vec{Z}_t : \vec{Z}_t \in \text{supp}(g_{\vec{Z}_t}(\cdot))\}$ is the set of observed controls.

Assumption 10. All the moments of last period's productivity exist and are bounded:

$$\sup_{i,t} \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} E[|\omega_{i,t-1}^*|^d | \vec{Z}_t] < \infty \quad \forall d \in \mathbb{N}$$

With these assumptions we can derive the below identification theorem. The intuition is similar to that of a simple regression. When the expected log change in inputs, $E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t]$, is unbounded it dominates the expected productivity, making it unbounded, unless $\beta_{X_m} = \beta_{X_m}^*$. The unboundedness in $E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t]$ essentially controls for the unobserved input shares and productivities when assumptions 9-10 hold. This also explains why there is a discontinuity at infinity, if $E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t]$ is bounded, the expected innovation changes continuously with β_{X_m} , so that the true parameter cannot be recovered. Similarly, when $E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t]$ and $E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t]$ are jointly unbounded – that is, when their ratio approaches a constant – this has the same effect as collinearity in a regression framework; their independent effects on the expected innovation in productivity can no longer be separated, unless there is variation in their relative rates of unboundedness.

Theorem 4.1. Under assumptions 1a-10 β_{X_m} is identified if and only if

$$\exists \{\vec{Z}_r\}_{r=1}^{\infty} \text{ s.t. } \lim_{r \rightarrow \infty} |E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t = \vec{Z}_r]| = \infty$$

and, for all $m' \neq m$, one of the following holds:

- $\exists \{\vec{Z}_r\}_{r=1}^\infty$ such that

$$\lim_{r \rightarrow \infty} \frac{E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t = \vec{Z}_r]}{E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t = \vec{Z}_r]} = c,$$

for $c \in \{-\infty, 0, \infty\}$

- $\exists \{\vec{Z}_r\}_{r=1}^\infty, \{\vec{Z}'_r\}_{r=1}^\infty$ such that

$$\lim_{r \rightarrow \infty} \frac{E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t = \vec{Z}_r]}{E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t = \vec{Z}_r]} = c,$$

$$\lim_{r \rightarrow \infty} \frac{E[x_{m',t'} - \rho' x_{m',t'-1} | \vec{Z}_t = \vec{Z}'_r]}{E[x_{m,t'} - \rho x_{m,t'-1} | \vec{Z}_t = \vec{Z}'_r]} = c',$$

for $c \neq c'$.

Proof. I will start with some definitions and notation.

Definition 2. Let

$$\omega(\vec{\theta}, S, \beta_0, Y, \vec{X}) = \ln(Y) - f(\vec{\theta}, S, \vec{X}) - \beta_0$$

be the productivity when a firm with total input vector \vec{X} produces a quantity of Y using a Cobb-Douglas production technology with parameter vector $\vec{\theta}$, Hicks-neutral productivity of β_0 , and an input share of S .¹¹

Then we can write

$$\begin{aligned} E[\zeta_{i,j,t} | \vec{Z}_t] &= \sum_{m=1}^M (\beta_{X_m}^* - \beta_{X_m}) E[x_{j,t}^m - \rho x_{j,t-1}^m | \vec{Z}_t] + \phi^* E[s_{i,j,t}^* - \rho s_{i,j,t-1}^* | \vec{Z}_t] \\ &\quad - \phi E[s_{i,j,t} - \rho s_{i,j,t-1} | \vec{Z}_t] + (\rho^* - \rho) E[\omega_{i,j,t-1}^* | \vec{Z}_t] + (1 - \rho^*) \beta_0^* - (1 - \rho) \beta_0, \end{aligned} \quad (3)$$

since $E[\zeta_{i,j,t}^* | \vec{Z}_t] = 0$.

¹¹That is,

$$f(\vec{\theta}, S, \vec{X}) = \sum_{m=1}^M \beta_{X_m} x_m + \phi s.$$

Definition 3. Let $\vec{Z}_{T+1} := (\vec{X}_T, \vec{Z}_T)$ be the vector of observables over all time periods in the data. Then denote the input allocation rule by

$$\mathcal{S} : [0, 1]^{2(T+1)} \times \mathbb{Z}_{T+1} \rightarrow \mathbb{R}_+,$$

where $\int_{[0,1]^{2(T+1)}} \mathcal{S}(\vec{S}|\vec{Z})d\vec{S} = 1$ and $\mathcal{S}(\vec{S}, \vec{Z}) = 0$ whenever $\sum_i S_{i,t} \neq 1$ for any t .

Intuitively, the input allocation rule defines what percentage of firms is mapped to each input share given their observables and past input allocations. Formally, it is a conditional distribution. I will also let $\mathcal{S}_{0.5}$ denote the allocation rule that splits all inputs equally regardless of the observables, ie $\mathcal{S}_{0.5}(\vec{S}, \vec{Z}) = 0$ whenever $\vec{S} \neq 0.5$.

Then the relevant distribution for the conditional expectation is the joint distribution of inputs, outputs, and input allocations induced by the input allocation rule and the joint distribution of observables:

$$g_{\vec{s}_{t:0}, \vec{y}_t, \vec{x}_t, \vec{z}_t | \mathcal{S}}(\vec{s}, \vec{y}, \vec{x}, \vec{z}) = g_{\vec{y}_t, \vec{x}_t, \vec{z}_t | \mathcal{S}}(\vec{y}, \vec{x}, \vec{z}) \int_{\vec{s}' \in \mathbb{S}(\vec{s})} \int_{\vec{z}' \in \mathbb{Z}(\vec{y}, \vec{x}, \vec{z})} \mathcal{S}(\vec{s}', \vec{z}'),$$

where $\mathbb{S}(\vec{s}) = \{\vec{S}_{T:0} : \exists \vec{S}_{T:t+1} \text{ s.t. } (\vec{S}_{T:t+1}, s) = \vec{S}_{T:0}\}$ and $\mathbb{Z}(\vec{y}, \vec{x}, \vec{z})$, defined similarly, are the set of vectors that agree with the respective histories up to time t . I will further write the expected conditional innovation as

Definition 4.

$$\bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}, \beta_0, \rho | \vec{Z}_t, \vec{S}_{t-1:0}) = E_{g_{\vec{s}_{t:0}, \vec{y}_t, \vec{x}_t, \vec{z}_t | \mathcal{S}}} \left[\omega(\vec{\theta}, S_{i,t}, \beta_0, \vec{Y}_{i,t}, \vec{X}_t) - \rho \omega(\vec{\theta}, S_{i,t-1}, \beta_0, \vec{Y}_{i,t-1}, \vec{X}_{t-1}) | \vec{Z}_t, \vec{S}_{t-1:0} \right]$$

For this proof I will first fix $(\vec{\theta}, \rho)$ and show that there exist \mathcal{S} and β_0 such that the moment conditions hold almost everywhere, i.e. $\sup_{i,t} \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} |\bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}, \beta_0, \rho | \vec{Z}_t)| = 0$, if and only if

$$\sup_{i,t} \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} \bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t) < \infty. \quad (4)$$

Then, in part 2, I will discuss under what conditions equation (4) holds and how to combine these insights to get point identification.

Part 1

By inverting $\omega(\vec{\theta}, S_{-i}, \beta_0, \vec{Y}, \vec{X})$ and using that $S_i = 1 - S_{-i}$ we can write the productivity of good i as a function of the productivity of good $-i$,

$$\omega_i(\vec{\theta}, \omega_{-i}, \beta_0, \vec{Y}, \vec{X}) = \ln(Y_i) - \phi \ln \left(1 - e^{(\ln(Y_{-i}) - \omega_{-i} - \sum_m \beta_{X_m} \ln(X_m) - \beta_0)/\phi)} - \sum_m \beta_{X_m} \ln(X_m) - \beta_0 \right). \quad (5)$$

That is, ω_i is a convex function of ω_{-i} . Therefore, it is possible to arbitrarily increase $E[\omega_{i,t}|\vec{Z}_t]$ while keeping $E[\omega_{-i,t}|\vec{Z}_t]$ constant by changing the distribution of \vec{S}_t . However, this is not symmetrical. That is, $E[\omega_{i,t}|\vec{Z}_t]$ can only be decreased by a finite amount by changing the distribution of \vec{S}_t since it is bounded below by $E[\ln(Y_i) - \sum_m \beta_{X_m} \ln(X_m) - \beta_0]$. However, as long as (4) holds, there is a simple approach to finding an input allocation rule and Hicks-neutral productivity such that the moment equalities hold.

First, choose β_0 ¹² high enough for the expected innovation is negative for almost all \vec{Z}_t , even if all inputs were assigned to both outputs. For example,

$$\beta_0 = \sup_{i,t} \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} \frac{\bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0|\vec{Z}_t) - \ln(0.5)}{(1 - \rho)}.$$

This makes it possible to increase $\omega_{i,t}$ in the next step without having to worry about $\bar{\zeta}_{i,t+1}$.

I am going to start at $t = 1$, at each step constructing an input allocation rule \mathcal{S}_t by changing only the conditional distribution of \vec{S}_t leaving the conditional distribution of \vec{S}_τ unchanged for all $\tau \neq t$,¹³ to make

$$\bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_t, \beta_0, \rho|\vec{Z}_t, \vec{S}_{t-1:0}) = 0 \text{ a.s.},$$

while ensuring

$$\bar{\zeta}_{i,t+1}(\vec{\theta}, \mathcal{S}_t, \beta_0, \rho|\vec{Z}_{t+1}, \vec{S}_{t:0}) \leq 0 \text{ a.s.}$$

Conditioning on $(\vec{Z}_t, \vec{S}_{t-1:0})$ means conditioning on last period's productivity,

¹²While it would be clearer to write $\beta_0^{\vec{\theta}, \rho}$ to make explicit that a different β_0 is chosen for different values of $\vec{\theta}$ and ρ , I drop those superscripts to avoid notational clutter.

¹³That is $g_{\vec{S}_{T:t+1}, \vec{S}_{t-1:0}, \vec{Z}_T|\mathcal{S}_t}(\cdot) = g_{\vec{S}_{T:t+1}, \vec{S}_{t-1:0}, \vec{Z}_T|\mathcal{S}_{t-1}}(\cdot)$.

$\omega(\vec{\theta}, S_{i,t-1}, \beta_0, Y_{i,t-1}, \vec{X}_{t-1})$. Construct \mathcal{S}_t so

$$g_{\vec{Z}_{T+1}|\vec{Z}_t, \vec{S}_{t:0}, \mathcal{S}_t}(\cdot|\vec{Z}, \vec{S}) = g_{\vec{Z}_{T+1}|\vec{Z}_t, \vec{S}_{t:0}, \mathcal{S}_t}(\cdot|\vec{Z}, \vec{S}') \quad \forall \vec{S}, \vec{S}' \in \text{supp}(\mathcal{S}_t).$$

That is, the distribution over future paths of observables is independent of the history of input allocations. This is important because it means that the distribution of future (counterfactual) productivities is also independent of the history of input allocations:

$$g_{\vec{\omega}_\tau|\vec{S}_{\tau-1:0}, \vec{Z}_\tau, \mathcal{S}_t}(\vec{S}, \vec{Z}) = g_{\vec{\omega}_\tau|\vec{S}_{\tau-1:0}, \vec{Z}_\tau, \mathcal{S}_t}(\vec{S}', \vec{Z}) \text{ for all } \vec{Z}_t \text{ and all } \tau > t.$$

Therefore

$$\bar{\zeta}_{i,t+1}(\vec{\theta}, \mathcal{S}_t, \beta_0, \rho|\vec{Z}_{t+1}, \vec{S}_{t:0}) = \bar{\zeta}_{i,t+1}(\vec{\theta}, \mathcal{S}_{0.5}, \beta_0, \rho|\vec{Z}_{t+1}) + \ln(S_{i,t}) - \ln(0.5) < 0,$$

ensuring that the second condition holds. Additionally, since

$$\bar{\zeta}_{i,\tau}(\vec{\theta}, \mathcal{S}_t, \beta_0, \rho|\vec{Z}_\tau, \vec{S}_{\tau-1:0}) = \bar{\zeta}_{i,\tau}(\vec{\theta}, \mathcal{S}_{0.5}, \beta_0, \rho|\vec{Z}_\tau) \quad \forall \tau > t + 1, \quad \vec{S}_{\tau-1:0} \in \text{supp}(\mathcal{S}_t),$$

the problem of finding a suitable input allocation in future periods is unaffected by the changes made to allocations in period t .

For each $(\vec{Z}_t, \vec{S}_{t-1:0})$, choose three input allocations, $(\text{vec}S_u^{\vec{Z}_t, \vec{S}_{t-1:0}}, S_{0.5}^{\vec{Z}_t, \vec{S}_{t-1:0}}, S_d^{\vec{Z}_t, \vec{S}_{t-1:0}})$, with $S_{u,1}^{\vec{Z}_t, \vec{S}_{t-1:0}} > 0.5$, $S_{0.5,1}^{\vec{Z}_t, \vec{S}_{t-1:0}} = 0.5$, and $S_{d,1}^{\vec{Z}_t, \vec{S}_{t-1:0}} < 0.5$, and weights, $(a_u, a_{0.5}, a_d)$ summing to one, so that

$$a_u \ln(S_{u,i}^{\vec{Z}_t, \vec{S}_{t-1:0}}) + a_{0.5} \ln(S_{0.5,i}^{\vec{Z}_t, \vec{S}_{t-1:0}}) + a_d \ln(S_{d,i}^{\vec{Z}_t, \vec{S}_{t-1:0}}) = \bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{t-1}, \beta_0, \rho|\vec{Z}_t, \vec{S}_{t-1:0}) + \ln(0.5).$$

Note that the right hand side is smaller than zero almost surely. Then by the concavity of $\ln(S)$ it is always possible to choose such a set of weights and input shares that the above equation is satisfied for both good i and $-i$.¹⁴

¹⁴That is, it is possible to pick a set of input shares, $(S_u^{\vec{Z}_t, \vec{S}_{t-1:0}}, S_{0.5}^{\vec{Z}_t, \vec{S}_{t-1:0}}, S_d^{\vec{Z}_t, \vec{S}_{t-1:0}})$, and weights, $(a_u, a_{0.5}, a_d)$, so that

$$\bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_t, \beta_0, \rho|\vec{Z}_t, \vec{S}_{t-1:0}) = \bar{\zeta}_{-i,t}(\vec{\theta}, \mathcal{S}_t, \beta_0, \rho|\vec{Z}_t, \vec{S}_{t-1:0}) = 0$$

almost everywhere.

Next, I show that, when $\text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} \bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t) = \infty$, there is no way to make it finite by changing the distribution of \vec{S}_{t-1} . By decreasing $S_{i,t-1}$ we can make $\zeta_{i,t}(\vec{\theta}, \vec{S}_{t:t-1}, \beta_0, \rho, \vec{Y}_{t:t-1}, \vec{X}_{t:t-1})$ arbitrarily small, but inevitably increase $\zeta_{-i,t}(\vec{\theta}, \vec{S}_{t:t-1}, \beta_0, \rho, \vec{Y}_{t:t-1}, \vec{X}_{t:t-1})$ albeit by a finite amount. However, by assumptions 9-10,

$$\sup_t \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} |\bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t) - \bar{\zeta}_{-i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t)|^d < \infty \quad \forall d \in \mathbb{N}. \quad (6)$$

Intuitively, the only source of unboundedness comes from the expected change in total log inputs, and log inputs have the same effect on both outputs. Since the moment conditions control for past input allocations it is impossible to increase $E[\omega_{i,t-1} | \vec{Z}_t, \vec{S}_{t-1:0}]$ and $E[\omega_{-i,t-1} | \vec{Z}_t, \vec{S}_{t-1:0}]$ simultaneously. Additionally, because of equation (6), the productivities jointly approach infinity no matter how we split the observations. Formally, for any $(h_{\vec{X}_t, \vec{Y}_t | \vec{Z}_t}^0(\cdot), h_{\vec{X}_t, \vec{Y}_t | \vec{Z}_t}^1(\cdot))$, such that

$$h_{\vec{X}_t, \vec{Y}_t | \vec{Z}_t}^0(\cdot) + h_{\vec{X}_t, \vec{Y}_t | \vec{Z}_t}^1(\cdot) = g_{\vec{X}_t, \vec{Y}_t | \vec{Z}_t}(\cdot),$$

we can find a function $V : \mathbb{Z}_t \rightarrow \{0, 1\}$ such that, for some $v \in \{V(\vec{Z}_t), |V(\vec{Z}_t) - 1|\}$,

$$\sup_t \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} |\bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t, h^v) - \bar{\zeta}_{-i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t, h^v)| < \infty$$

and

$$\sup_t \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} |\bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t, h^v) = \infty,$$

where

$$\bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t, h^v) = E_{h_{\vec{Y}_t, \vec{X}_t, \vec{Z}_t}^v} \left[\omega(\vec{\theta}, 0.5, 0, \vec{Y}_{i,t}, \vec{X}_t) - \rho \omega(\vec{\theta}, 0.5, 0, \vec{Y}_{i,t-1}, \vec{X}_{t-1}) \right].$$

Since changes to $g_{\vec{S}_t | \vec{Z}_t, \vec{S}_{t-1:0}}(\cdot)$ can at most reduce $\sup_i \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} \bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}, \beta_0, \rho | \vec{Z}_t)$ by a finite amount, this proves that, for a given $(\vec{\theta}, \rho)$ the moment conditions can be made to hold almost surely if and only if

$$\sup_{i,t} \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} \bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t) < \infty.$$

Part 2

I will now describe what parameters can satisfy the moment conditions and then show

how to use the shape of this set to recover point identification. To build intuition I will first consider the case where the essential supremum is finite for all inputs except X_m :

$$\operatorname{ess\,sup}_{\vec{Z}_t \in \mathbb{Z}_t} |E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t]| < \infty \text{ for all } m' \neq m.$$

Then, from equation (3) we can see that $\operatorname{ess\,sup}_{\vec{Z}_t \in \mathbb{Z}_t} \bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t) < \infty$ if and only if one of the following holds:

- $\operatorname{ess\,sup}_{\vec{Z}_t \in \mathbb{Z}_t} E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t] = \infty$ and $\beta_{X_m}^* \leq \beta_{X_m}$
- $\operatorname{ess\,sup}_{\vec{Z}_t \in \mathbb{Z}_t} E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t] = -\infty$ and $\beta_{X_m}^* \geq \beta_{X_m}$
- $-\infty < \operatorname{ess\,sup}_{\vec{Z}_t \in \mathbb{Z}_t} E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t] < \infty$

However, with more unbounded inputs, it is possible that $\operatorname{ess\,sup}_{\vec{Z}_t \in \mathbb{Z}_t} \bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t) < \infty$ even if the above conditions fail to hold. I will now add a second input, m' , for which the expectation diverges.

Take a sequence of controls, $\{\vec{Z}_r\}_{r=1}^\infty$, such that

$$\lim_{r \rightarrow \infty} E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t = \vec{Z}_r] = \infty$$

and

$$\lim_{r \rightarrow \infty} \frac{E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t = \vec{Z}_r]}{E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t = \vec{Z}_r]} = c,$$

for some $c \in \mathbb{R} \cup \{-\infty, \infty\}$. Then we can rewrite the limit of the expected innovation in productivity as

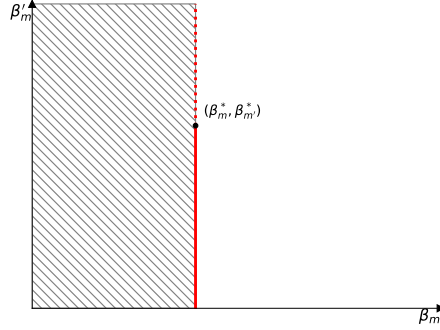
$$\lim_{r \rightarrow \infty} \bar{\zeta}_{i,t}(\vec{\theta}, \mathcal{S}_{0.5}, 0, \rho | \vec{Z}_t) = \left((\beta_{X_m}^* - \beta_{X_m}) + c(\beta_{X_{m'}}^* - \beta_{X_{m'}}) \right) \lim_{r \rightarrow \infty} E[x_{m,i,t} - \rho x_{m,i,t-1} | \vec{Z}_t = \vec{Z}_r] + C,$$

for some $C \in \mathbb{R}$. Therefore, the expected innovation is bounded above if and only if:

$$c(\beta_{X_m}^* - \beta_{X_m}) + (\beta_{X_{m'}}^* - \beta_{X_{m'}}) \leq 0.$$

Figure 2 shows that $\beta_{X_m}^*$ can be recovered if $c \in \{-\infty, \infty\}$. It is also noteworthy that we can do this recursively. That is, once we know $\beta_{X_m}^*$ we can recover $\beta_{X_{m'}}^*$ even

Figure 2: Sequential recovery



Shaded area denotes set of $\vec{\theta}$ satisfying the moment conditions for $\lim_{r \rightarrow \infty} E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t = \vec{Z}_r] = \infty$ with $c = \infty$. The dotted line is not part of this set.

if

$$\lim_{\vec{Z}_t \rightarrow Z} \frac{E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t]}{E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t]} = 0$$

for all $\rho, t, \{\vec{Z}_r\}_{r=1}^{\infty}$ such that $\lim_{\vec{Z}_t \rightarrow Z} E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t] = \infty$. This covers the case for $c = 0$.

Lastly, from figure 3 one can see that, when $c \notin \{-\infty, 0, \infty\} \forall (\rho, t, \{\vec{Z}_r\}_{r=1}^{\infty}), \beta_{X_m}^*$ can be recovered if and only if there exist (ρ, t, Z) and (ρ', t', Z') such that the ratio of expected log input changes approaches c and c' respectively, and $c \neq c'$.

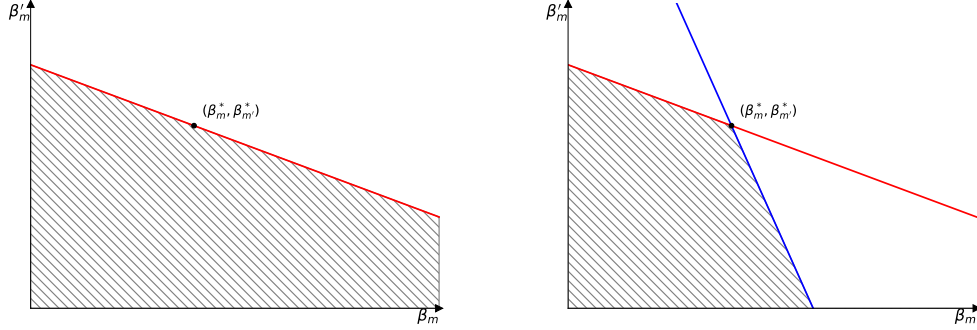
□

While the identification result is of independent interest, a key takeaway is that, as the cross-sectional sample size increases, there is a discontinuity at infinity. With any finite sample

$$\sup_t \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} |E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t]| < \infty$$

so that any $\vec{\theta}$ can be rationalised. This shows that, in practice, additional assumptions need to be made and motivates my result in the next section, where I propose a novel identification strategy based alternative assumptions.

Figure 3: Recovering point identification



Red line: $c = 0.5$. Blue line: $c = 1$. Shaded area is set of $\vec{\theta}$ satisfying the moment conditions for $\lim_{r \rightarrow \infty} E[x_{m,t} - \rho x_{m,t-1} | \vec{Z}_t = \vec{Z}_r] = \infty$.

5 Set Identification

In this section I show that we can recover set identification by restricting the productivity process to be in a stationary equilibrium.

Assumption 11 (Stationarity Equilibrium). The true joint probability density of productivity, denoted by $g_{\omega|t}^*$, is stationary; that is,

$$g_{\omega|t}^*(\cdot) = g_{\omega|t'}^*(\cdot) \quad \forall t, t'.$$

Recall from definition 3 that \mathcal{S} denotes the joint distribution of product-specific inputs. Let $\vec{\omega}(\vec{\theta}, \vec{S}, \beta_0, \vec{Y}, \vec{X}) = (\omega(\vec{\theta}, S_1, \beta_0, Y_1, \vec{X}), \omega(\vec{\theta}, S_2, \beta_0, Y_2, \vec{X}))$ be the vector of productivities as defined in definition 2.

Definition 5. The joint distribution of (counterfactual) productivities is defined by the density

$$g(\vec{\omega}; \vec{\theta}, \mathcal{S}|t) = \int_{\mathbb{Y} \times \mathbb{X}} \int_{[0,1]^n} \delta(\vec{\omega} - \vec{\omega}(\vec{\theta}, \vec{s}, 0, \vec{y}, \vec{x})) \mathcal{S}_t(\vec{y}, \vec{x}, \vec{s}) g_{\vec{Y}, \vec{X}|t}(\vec{y}, \vec{x}) d\vec{s} d\vec{y} d\vec{x},$$

where $\delta(\cdot)$ is the Dirac delta function. Intuitively, $g_t(\vec{\omega}; \vec{\theta}, \mathcal{S})$ aggregates the probability mass of all triples $(\vec{y}, \vec{x}, \vec{s})$ that map into the productivity vector $\vec{\omega}$.

Define the space of possible parameters as $\Theta = \{\theta : \theta \in (0, \infty]^M\}$. Denote the identified set by Θ_I and its complement by $\Theta_I^C = \Theta \setminus \Theta_I$.

The following theorem and its proof are written for the case of a Cobb-Douglas production technology. As the result relies on stationarity of the productivity distribution and β_0 shifts all productivities equally it is not identified unless additional assumptions on the productivity process are made, for example, by normalising $E[\omega_i^*] = 0$.

Theorem 5.1. *If assumptions 1a-6 and 11 hold and there exist at least two periods t, t' and input $X_m \in \mathbb{X}$ such that its distribution differs between periods t and t' , $g_{X_m|t}(\cdot) \neq g_{X_m|t'}(\cdot)$, then the set*

$$\Theta_I^C = \left\{ \vec{\theta} : g(\cdot; \vec{\theta}, \mathcal{S}, \beta_0|t) \neq g(\cdot; \vec{\theta}, \mathcal{S}, \beta_0|t') \text{ for all } \mathcal{S} \right\}$$

is non-empty. That is, under the distribution of observables, some parameter values are inconsistent with productivity being in stationary equilibrium.

Proof. If it exists, denote by $h_{\vec{\theta}}(\cdot)$ the stationary distribution of productivities that rationalises the data given $\vec{\theta}$. More formally, if there exists a set of input allocations, \mathcal{S} , such that $g(\cdot; \vec{\theta}, \mathcal{S}|t) = g(\cdot; \vec{\theta}, \mathcal{S}|t') \forall t, t'$, let $h_{\vec{\theta}}(\cdot) := g(\cdot; \vec{\theta}, \mathcal{S}|t)$.¹⁵

From assumption 11 we know that $h_{\vec{\theta}^*}(\cdot)$ exists. Therefore, the identified set is non-empty. It must be possible to re-allocate inputs in each period to change the counterfactual joint density of productivities $g(\cdot; \vec{\theta}^*, \mathcal{S}_{0.5}|t)$ into $h_{\vec{\theta}^*}(\cdot)$. In general, this problem can be thought of as finding a family of $T + 1$ transport density functions $\{\pi_t\}_{t=0}^T$ on $\mathbb{R}^2 \times \mathbb{R}^2$, where $\pi_t(\vec{\omega}|\vec{\omega}')$ represents the conditional density of transporting mass from productivity tuple $\vec{\omega}'$ to productivity tuple $\vec{\omega}$, such that the marginal consistency condition:

$$h_{\vec{\theta}}(\vec{\omega}) = \int_{\mathbb{R}^2} \pi_t(\vec{\omega}|\vec{\omega}') g(\vec{\omega}'; \vec{\theta}, \mathcal{S}_{0.5}|t) d\vec{\omega}'$$

and the normalization condition hold:

$$\int_{\mathbb{R}^2} \pi_t(\vec{\omega}|\vec{\omega}') d\vec{\omega} = 1 \quad \text{for all } \vec{\omega}',$$

¹⁵I do not prove uniqueness, so there may be multiple stationary distributions of productivities induced by different sets of input allocations. If there are, the statements I make hold true for all of them so it is without loss of generality to pick one at random.

where $\vec{\omega}' = (\omega'_i, \omega'_{-i})$ and $\vec{\omega} = (\omega_i, \omega_{-i})$, and where the density $\pi_t(\vec{\omega}|\vec{\omega}')$ is constrained to have support only on the feasible set of productivities, $\mathbb{W}(\vec{\omega}')$, defined below. That is, $\pi_t(\vec{\omega}|\vec{\omega}') = 0$ for all $\vec{\omega} \notin \mathbb{W}(\vec{\omega}')$.¹⁶

To derive the productivity pairs that could be achieved by reallocating inputs, $\mathbb{W}(\vec{\omega}')$, use equation (5). By substituting in

$$\omega(\vec{\theta}, 0.5, 0, Y, \vec{X}) = \ln(Y) - \sum_m \beta_{X_m} \ln(X_m) - \phi \ln(0.5)$$

we obtain the feasible set of counterfactual productivities:

$$\mathbb{W}(\vec{\omega}') = \left\{ (\omega_i, \omega_{-i}) : \omega_{-i} = \omega'_{-i} + \phi \ln(0.5) - \phi \ln \left(1 - \exp \left(\frac{\omega'_i - \omega_i + \phi \ln(0.5)}{\phi} \right) \right) \right\}. \quad (7)$$

The shape of these sets have important implications for how different $g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t)$ and $h_{\vec{\theta}}(\cdot)$ can be. It is easy to verify that the set of feasible counterfactual productivity pairs is non-intersecting, $\mathbb{W}(\vec{\omega}) \cap \mathbb{W}(\vec{\omega}') = \emptyset$, if and only if $\vec{\omega} \geq \vec{\omega}'$ or $\vec{\omega}' \geq \vec{\omega}$ with $\vec{\omega} \neq \vec{\omega}'$.¹⁷ Let $\mathbb{W}_{\vec{\omega}'}^U = \{\vec{\omega} : \vec{\omega} \geq \vec{\omega}'' \text{ for some } \vec{\omega}'' \in \mathbb{W}(\vec{\omega}')\}$ be the upper contour set of the feasible productivity pairs and define $G_{\vec{\omega}'}^U = \{\vec{\omega} : \vec{\omega} \geq \vec{\omega}'\}$. Similarly, define $G_{\vec{\omega}'}^L = \{\vec{\omega} \in \mathbb{W}_{\vec{\omega}'}^L : \vec{\omega} \leq \vec{\omega}'\}$. This implies that

$$\Pr(\mathbb{W}_{\vec{\omega}'}^U | h_{\vec{\theta}}(\cdot)) \leq 1 - \Pr(G_{\vec{\omega}'}^L | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t)), \quad (8)$$

because all feasible productivities for $\vec{\omega} \leq \vec{\omega}'$ lie weakly outside $\mathbb{W}_{\vec{\omega}'}^U$. Therefore, any feasible transport plan that moves mass to $\mathbb{W}_{\vec{\omega}'}^U$ has to take this mass from outside $G_{\vec{\omega}'}^L$.

Similarly,

$$\Pr(\mathbb{W}_{\vec{\omega}'}^L | h_{\vec{\theta}}(\cdot)) \geq \Pr(G_{\vec{\omega}'}^U | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t)), \quad (9)$$

as any feasible transport plan can only move mass from points in $G_{\vec{\omega}'}^U$ to points in $\mathbb{W}_{\vec{\omega}'}^L$.

¹⁶There are some measurability concerns I do not address here. Particularly, whether $\pi_t(\vec{\omega}|\vec{\omega}')$ is integrable with respect to $\vec{\omega}'$. However, the precise definition of the reallocation problem is not crucial. The results derive from the constraint of the potential set of destinations, $\mathbb{W}(\vec{\omega}')$. The transport map characterisation is only included to drive intuition.

¹⁷Where $(x_1, x_2) \geq (y_1, y_2)$ means $x_i \geq y_i$ for $i \in \{1, 2\}$.

Since equations (8) and (9) have to hold for all t we can respectively take the minimum and maximum across time:¹⁸

$$\max_t \Pr(G_{\vec{\omega}'}^U | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t)) \leq \Pr(\mathbb{W}_{\vec{\omega}'}^U | h_{\vec{\theta}}) \leq \min_t \left(1 - \Pr(G_{\vec{\omega}'}^L | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t)) \right).$$

Which can be rearranged to give

$$\max_t \Pr(G_{\vec{\omega}'}^U | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t)) + \max_t \Pr(G_{\vec{\omega}'}^L | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t)) \leq 1. \quad (10)$$

I will now show that, given t, t' such that $g_{X_m|t}(\cdot) \neq g_{X_m|t'}(\cdot)$, there exists $\vec{\theta}$ such that $\max_{s \in \{t, t'\}} \Pr(G_{\vec{\omega}'}^U | h_{s, \vec{\theta}}) + \max_{s \in \{t, t'\}} \Pr(G_{\vec{\omega}'}^L | h_{s, \vec{\theta}}) > 1$. That is, Θ_I^C is non-empty. Furthermore, since $F(\vec{\theta}, \vec{X})$ is continuous in $\vec{\theta}$, proving this also proves that Θ_I^C has non-zero measure given the Lebesgue measure on $\mathbb{R}^{\dim(\vec{\theta})}$.

The Cobb-Douglas production technology is strictly increasing in $\vec{\theta}$. Therefore, given two different vectors of inputs, \vec{X}, \vec{X}' , we can find a sequence of parameter vectors $\{\vec{\theta}_r\}_{r=1}^\infty$ such that:

$$\lim_{r \rightarrow \infty} |F(\vec{\theta}_r, \vec{X}) - F(\vec{\theta}_r, \vec{X}')| = \infty.$$

Without loss of generality assume that $g_{X_M|t} \neq g_{X_M|t'}$. Now take a sequence, $\{\vec{\theta}_r\}_{r=1}^\infty$, such that $\frac{\beta_{X_m, r}}{\beta_{X_M, r}} = c_{M, m}$, and $\lim_{r \rightarrow \infty} \beta_{X_M, r} = \infty$. Denote the linear combination of log inputs by $x^{comb}(\vec{X}) := x_M + \sum_{m=1}^{M-1} c_{M, m} x_m$. Then, as we go along the sequence, the (normalized) difference in counterfactual productivities gets dominated by the difference in x^{comb} :

$$\lim_{r \rightarrow \infty} \frac{\omega(\vec{\theta}_r, 0.5, 0, Y, \vec{X}) - \omega(\vec{\theta}_r, 0.5, 0, Y', \vec{X}')}{\beta_{X_M, r}} = x^{comb}(\vec{X}) - x^{comb}(\vec{X}'). \quad (11)$$

Now consider the cumulative distribution function (CDF) of $x^{comb}(\cdot)$ induced by $g_{\vec{Y}, \vec{X}|t}$ and denote it $G_{x^{comb}|t}(\cdot)$. Pick any x such that $G_{x^{comb}|t}(x) \neq G_{x^{comb}|t'}(x)$, I will prove below that this is always possible. Assume, without loss of generality, that $G_{x^{comb}|t}(x) <$

¹⁸Nonetheless, for a given ω' , these are properly viewed as $\binom{T+1}{2}$ moment inequalities, one for each pair of time periods.

$G_{x^{comb}|t'}(x)$. Then

$$\lim_{r \rightarrow \infty} \Pr \left(G_{\vec{\omega}_r}^U | g(\cdot; \vec{\theta}_r, \mathcal{S}_{0.5}|t) \right) + \Pr \left(G_{\vec{\omega}_r}^L | g(\cdot; \vec{\theta}_r, \mathcal{S}_{0.5}|t') \right) = 1 - G_{x^{comb}|t}(x) + G_{x^{comb}|t'}(x) > 1,$$

for $\vec{\omega}_r = (x\beta_{X_M,r}, x\beta_{X_M,r})$.

The last thing to show is that, whenever $g_{X_M|t}(\cdot) \neq g_{X_M|t'}(\cdot)$, we can find a linear combination of log inputs so that there exists an x such that $G_{x^{comb}|t}(x) \neq G_{x^{comb}|t'}(x)$.¹⁹ However, this is easy as we can simply put a weight of 1 on X_M and 0 on $X_m \forall m \neq M$, concluding the proof. □

The set-identification result has some appealing properties. Firstly, it only relies on computing empirical probabilities, which is computationally cheap and does not require any exogenous instruments. This has many advantages. [no need for competitive or fully flexible goods]

There are, however, also downsides when using inference. Since there exists an inequality for each $\vec{\omega}' \in \mathbb{R}^2$ there are an uncountable number of potential inequalities. Additionally, even when choosing a finite number of points, equation (10) actually summarizes $\binom{T+1}{2}$ moment inequalities for each $\vec{\omega}'$ meaning that the number of moment inequalities grows rapidly.

6 Set-Identification Generalisations

It is worth discussing how some of the assumptions underlying the set-identification result can be relaxed.

6.1 Measurement error

If outputs are measured with log-additive error so that $Y_{i,j,t} = Y_{i,j,t}^* e^{\epsilon_{i,j,t}}$ then the results in each section hold as long as the respective assumptions hold for the composite

¹⁹The assumption that $g_{X_m|t}(\cdot) \neq g_{X_m|t'}(\cdot)$ for some $X_m \in \mathbb{X}$ is sufficient but not necessary for the existence of such a point. For example, if the joint distributions of total inputs, $g_{\vec{X}|t}$, $g_{\vec{X}|t'}$, have support only on $[1, \infty)^M$ then all log inputs would be positive so a version of the Cramer-Wold theorem with respect to the positive orthant tells us that $G_{x^{comb}|t}(x) = G_{x^{comb}|t'}(x) \forall x \in \mathbb{R}$ if and only if $g_{\vec{X}|t}(\cdot) = g_{\vec{X}|t'}(\cdot)$, a strictly weaker condition than $g_{X_M|t}(\cdot) \neq g_{X_M|t'}(\cdot)$ if there are at least two inputs.

productivity term $\tilde{\omega}_{i,j,t}^* = \omega_{i,j,t}^* + \epsilon_{i,j,t}^*$. This allows for a wide range of conceivable error structures. For example, it could allow more productive firms to be more accurate when reporting output or for larger firms to make bigger absolute errors in reporting. This is a major advantage of the methods in this paper over methods using demand side inversions, like [Orr \(2022\)](#) or [Valmari \(2023\)](#), which cannot handle measurement error in outputs as it affects the firm’s first order conditions in interdependent and non-linear ways, making the inversion difficult.²⁰

6.2 Productivity trends

The productivity literature has found strong evidence of the average productivity in an industry increasing over time. The set identification result generalises to a situation where the distribution is stationary except for a time trend. Adding a trend gives additional degrees of freedom, decreasing the informativeness of the identified set. Allowing for a linear trend only introduces a single extra degree of freedom. On the extreme end if we allow a different β_0 for each t it becomes very hard to reject stationarity. If $\Pr(G_{\vec{\omega}}^U | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t)) + \Pr(G_{\vec{\omega}}^L | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t')) > 1$ increasing $\beta_{0,t}$ or decreasing $\beta_{0,t'}$ enough will make the equality hold. This can be repeated until all inequalities hold unless $\Pr(G_{\vec{\omega}}^U | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t)) + \Pr(G_{\vec{\omega}}^L | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t')) > 1$ and $\Pr(G_{\vec{\omega}'}^U | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t')) + \Pr(G_{\vec{\omega}'}^L | g(\cdot; \vec{\theta}, \mathcal{S}_{0.5}|t)) \geq 1$ or there exists a cycle of periods for which an equivalent set of inequalities hold.

6.3 Non-common technology

Non-common technologies mean that lemma 1 no longer applies. This does not materially change anything for Cobb-Douglas production functions as they are homogeneous in each input separately. However, in general, it means that the set of feasible productivity pairs, \mathbb{W} , is now not just a function of $\vec{\omega}$ but also of \vec{X} . This means that simple arguments about the shape of the feasible set no longer work and only allows identification if the dependence on \vec{X} is restricted enough.

²⁰[Caselli et al. \(2025\)](#) allow for an error to observed output prices that is independently and identically distributed across firms, time, and products. By taking care to choose instruments that are independent of measurement errors their method can incorporate measurement errors in quantity as well.

7 Simulations

While theorem 5.1 shows that the set of production function parameters that can be rejected is never empty, it does not clarify how informative the identified set is in practice. It can be arbitrarily informative, but also arbitrarily uninformative depending on the distribution of observables.

In this section, I use simulations to show that, even with a small number of periods, the identified set places significant constraints on the potential data-generating parameter vectors and becomes increasingly informative as more periods are added.

To give a first approximation of informativeness, I calculate point estimates of the identified set without inference.²¹ I simulate data for two firms that each produce two goods using a Cobb-Douglas production technology with labour, capital, and materials as inputs. The firm pair competes on two separate duopolistic markets with logit demand. For ease of calculation, I assume that the firms are colluding.

Definition 6 (Logit demand). The demand for good i produced by firm j is given by

$$Y_{i,j,t} = N \frac{\exp(d_{i,j} + \delta_{i,j,t} - aP_{i,j,t})}{\exp(d_{i,j} + \delta_{i,j,t} - aP_{i,j,t}) + \exp(d_{i,-j} + \delta_{i,-j,t} - aP_{i,-j,t})},$$

where N is the market size, $d_{i,j}$ is the mean utility of good i produced by firm j and $\delta_{i,j,t}$ is a mean-zero random normal shock.

I assume that total labour and capital are fixed at $t-1$ while firms buy materials on a perfectly competitive market. In line with the evidence by Cabral and Mata (2003), the initial values of labour and capital are drawn from a log-normal distribution. For the baseline specification, I have chosen a mean of 1, variance of 2, and covariance of 0.3. Log capital and log labour then evolve according to a random walk. That is,

$$k_{j,t} = k_{j,t-1} + \epsilon_{j,t}^k,$$

$$l_{j,t} = l_{j,t-1} + \epsilon_{j,t}^l,$$

²¹This means I calculate the identified set using the moment inequalities from equation (10) but do not calculate standard errors for the set.

where the random components, $(\epsilon_{j,t}^k, \epsilon_{j,t}^l)$, are drawn from a multi-variate normal distribution with a standard deviation of 0.25 and correlation of 0.3.

The material input price is drawn from a uniform distribution with support $[0.1, 1]$. It also follows a random walk with a standard deviation of 0.25. I follow the empirical evidence of Foster et al. (2008) and draw productivity from a stationary distribution with persistence parameter $\rho = 0.8$ and standard deviation of 0.25.

I simulate the economy for 5 periods, drawing a total of 2,500 observations each period - 50 draws of productivities for each of 50 draws of labour and capital. All firms in the same period face the same demand shocks to ensure these are not driving the results. For each specification of the simulation parameters I generate 50 datasets.

For the baseline specification I take the highest product specific productivity for every period and evaluate the moment inequalities at all combinations of these values.²² When I demonstrate that the estimator performs better as more periods are added, I instead take a set of percentiles of the productivity distribution in the dataset with all 20 periods, and evaluate the moments at their combination to avoid having the number of moments explode.

I assume that the econometrician knows that $\vec{\theta} \in [0, 1]^3$. To approximate the identified set, I use a grid search using increments of 0.1 for each parameter giving a total of 1,331 parameter combinations. The mean number of grid points in the identified set is around 50 to 80 in most specifications,²³ with medians much smaller around 40 points. These points are not symmetrically distributed around the true parameter vector, as can be seen from figure 4.

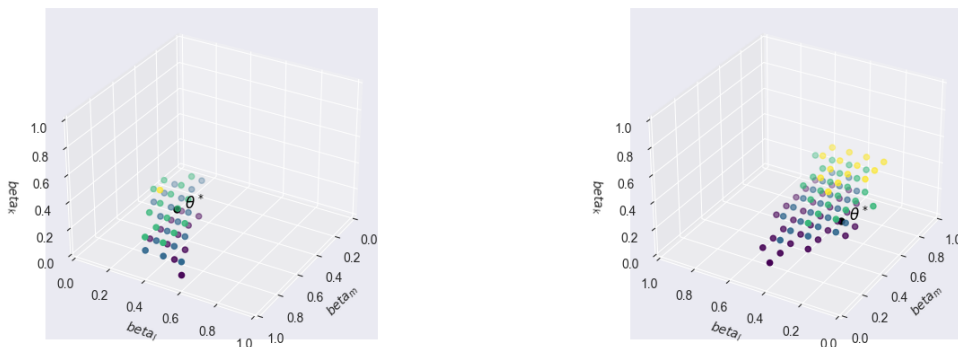
The true parameter vector is contained in the identified set for all simulated datasets.²⁴ This is reassuring, but expected as the moment inequalities give necessary not sufficient conditions for the existence of an input allocation rule that makes the productivity distribution stationary, leading to over-acceptance.

²²For example, if I had two periods with $\max_j \omega_{1,j,0} = 1$, $\max_j \omega_{1,j,1} = 2$, $\max_j \omega_{2,j,0} = 3$, and $\max_j \omega_{2,j,1} = 4$, then I would evaluate the moment inequalities at $(1, 3)$, $(1, 4)$, $(2, 3)$, and $(2, 4)$.

²³88.4 for the main specification.

²⁴With 50 datasets per specification and 8 specifications this means the true parameter vector was covered by the identified set for 400 different samples.

Figure 4: Identified Set - Baseline



Lighter colours reflect higher values of β_k .

The estimator is robust to misspecification in short samples. When drawing the initial productivity from a non-stationary AR(1) distribution²⁵ the identified set continues to contain the true parameter vector in all 50 datasets. However, the identified set tends to be noticeably smaller with a mean of 67.8 and a median of 26 grid points. Therefore, it is likely that for time spans significantly longer than 5 periods and for much smaller values of persistence the estimator becomes unreliable.

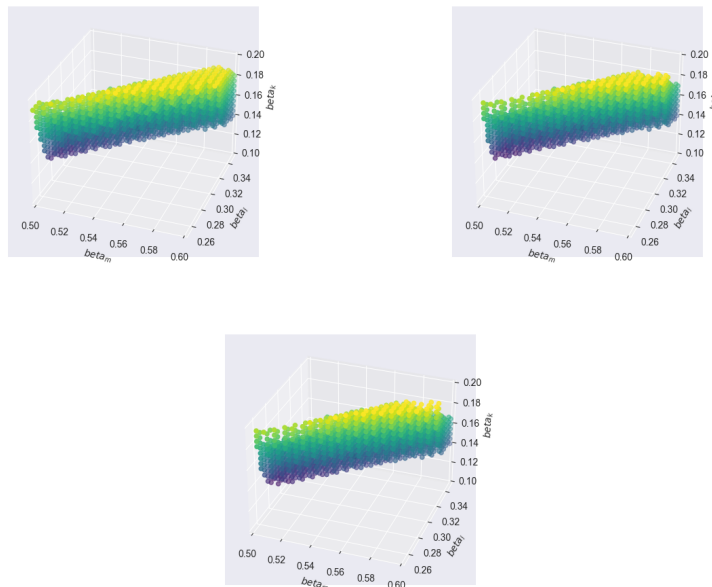
Figure 5 shows how the identified set shrinks as more time periods are added. To make the data generating process unfavourable, I model the labour and capital evolution as an AR(1) process rather than a random walk. This means that the sharper identified set as more periods are added is not simply a result of unnaturally large labour and capital values,²⁶ but it also means that smaller and smaller changes to the identified set can be expected as periods are added. This is also what can be seen in the graph.

The identified set in this exercise is significantly smaller than the set in the baseline specification, even with 5 periods. This exemplifies the importance of choosing the

²⁵I use an initial standard deviation of 0.25 rather than $0.25/(1-\rho)^2$.

²⁶Notably, a unit of labour today is not equivalent to a unit of labour in 1990, and similarly for capital. Therefore, even if the distribution of labour and capital amongst firms was the same today as in 1990, this would signify a growth in deflated inputs. As a result, the AR(1) process likely has less variation than can be expected in real-world data.

Figure 5: Identified Set - Longer Horizons



Identified set calculated for $T = 5$ (upper left), $T = 10$ (upper right), and $T = 20$ (bottom) using the same points to evaluate the moment inequalities. Lighter colours reflect higher values of β_k .

optimal values at which to evaluate the moments, an open question at this stage.

8 Conclusion

An increasing number of papers have been using assumptions on market structure, firm conduct, and demand to identify the supply side. In this paper I show that without relying on demand-side information and profit maximisation, the production function is only identified as the time-series variance in inputs becomes unbounded. More critically, in any finite sample, without further assumptions any parameter vector can be rationalised. As misspecifications of demand can severely bias the supply side estimates, I propose an alternative strategy, imposing stationarity of the productivity distribution to recover set identification.

My approach significantly expands the toolkit for researchers studying productivity, markups, and market power. It is particularly valuable for empirical work in mature industries, such as beverages or petroleum refining, where market conduct is often dis-

puted, making the stationarity assumption more palatable than relying on a specific demand model. Additionally, the proposed framework can be used to test market conduct and assumptions on the shape of demand under the null hypothesis that firms are cost minimising and productivity is in stationary equilibrium.

A major benefit of my approach is that it does not require any instrumental variables and only requires the researcher to compute empirical probabilities. This is computationally cheap and straightforward. It also makes the estimator robust to misspecifications of adjustment costs, input market power, or timing assumptions.

This work opens several avenues for future research. The set-identification approach generates an infinite number of moments and future work could study how to optimally select a finite subset. Furthermore, while this paper uses stationarity to achieve set identification, it may be possible to recover point identification by imposing more structure on the productivity process – for instance, by modelling it as a first-order Markov process – while weakening stationarity. These extensions promise to further refine our ability to understand the production technologies of multi-product firms, which dominate modern economies.

References

- Daniel A. Akerberg, Kevin Caves, and Garth Frazer. Identification Properties of Recent Production Function Estimators. *Econometrica*, 2015.
- David Autor, David Dorn, Lawrence F Katz, Christina Patterson, and Jon Van Reenen. The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 2020.
- Matthew Backus, Christopher Conlon, and Michael Sinkinson. Common ownership and competition in the ready-to-eat cereal industry. *NBER Working Paper Series*, 2021.
- William J Baumol, John C Panzar, Robert D Willig, Elizabeth E Bailey, Dietrich Fischer, and Herman C Quirmbach. *Contestable markets and the theory of industry structure*. New York: Harcourt Brace Jovanovich, 1982.
- Andrew B. Bernard, J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. Firms in international trade. *Journal of Economic Perspectives*, 2007.
- Richard Blundell and Stephen Bond. Gmm estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 2000.
- Steve Bond, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics*, 121:1–14, July 2021. ISSN 0304-3932. doi: 10.1016/j.jmoneco.2021.05.004.
- Luís Cabral and José Mata. On the evolution of the firm size distribution: Facts and theory. *The American Economic Review*, 2003.
- John Cairncross, Peter Morrow, Scott Orr, and Swapnika Rachapalli. Multi-Product Markups. 2023.
- Mauro Caselli, Arpita Chatterjee, and Shengyu Li. Productivity and quality of multi-product firms. 2025.
- Zhiyuan Chen and Moyu Liao. Identification and estimation of production functions for multiproduct firms. 2022.

- Jan De Loecker and Paul T. Scott. Markup estimation using production and demand data: An application to the us brewing industry. 2024.
- Jan De Loecker, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. Prices, markups, and trade reform. *Econometrica*, 84(2):445–510, 2016. ISSN 0012-9682. doi: 10.3982/ecta11042.
- Maarten De Ridder, Basile Grassi, and Giovanni Morzenti. The hitchhiker’s guide to markup estimation: Assessing estimates from financial data. 2024.
- Emmanuel Dhyne, Amil Petrin, Valerie Smeets, and Frederic Warzynski. Theory for extending single-product production function estimation to multi-product settings. 2022.
- Marco Duarte, Lorenzo Magnolfi, Mikkel Sølvsten, and Christopher Sullivan. Testing firm conduct. *Quantitative Economics*, 2024.
- Lucia Foster, John Haltiwanger, and Chad Syverson. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 2008.
- Binlei Gong and Robin C. Sickles. Resource allocation in multi-divisional multi-product firms. *Journal of Productivity Analysis*, 55(2):47–70, February 2021. ISSN 1573-0441. doi: <https://doi.org/10.1007/s11123-020-00595-5>.
- Hugo A. Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 1992.
- Chang-Tai Hsieh and Peter J. Klenow. Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, 2009.
- Nir Jaimovich, Stephen J. Terry, and Nicolas Vincent. The empirical distribution of firm dynamics and its macro implications. *NBER Working Paper Series*, 2023.
- Hiroaki Kaido, Francesca Molinari, and Jörg Stoye. Confidence intervals for projections of partially identified parameters. *Econometrica*, 2019.
- Agnes Norris Keiller, Áureo de Paula, and John Van Reenen. Production function estimation using subjective expectations data. *NBER Working Paper Series*, 2024.

- Brendan Kline and Elie Tamer. Bayesian inference in a class of partially identified models. *Quantitative Economics*, 2016.
- James Levinsohn and Amil Petrin. Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 2003.
- Lorenzo Magnolfi and Christopher Sullivan. A comparison of testing and estimation of firm conduct. *Economics Letters*, 2022.
- Nathan H. Miller and Matthew C. Weinberg. Understanding the price effects of the millercoors joint venture. *Econometrica*, 2017.
- Nathan H. Miller, Gloria Sheu, and Matthew C. Weinberg. Oligopolistic price leadership and mergers: The united states beer industry. *American Economic Review*, 2021.
- Felix Montag. Mergers, foreign competition, and jobs: Evidence from the u.s. appliance industry. 2024.
- G. Steven Olley and Ariel Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 1996.
- Scott Orr. Within-firm productivity dispersion: Estimates and implications. *Journal of Political Economy*, 2022. doi: [.org/10.1086/720465](https://doi.org/10.1086/720465).
- Ariel Pakes, Jack Porter, Mark Shepard, and Sophie Calder-Wang. Unobserved heterogeneity, state dependence, and health plan choices. 2025.
- Joseph P. Romano, Azeem M. Shaikh, and Michael Wolf. A practical two-step method for testing moment inequalities. *Econometrica*, 2014.
- Nelli Valmari. Estimating production functions of multiproduct firms. *Review of Economic Studies*, 90(6):3315–3342, January 2023. ISSN 1467-937X. doi: [10.1093/restud/rdad005](https://doi.org/10.1093/restud/rdad005).

A Identification at infinity – Generalisations

The comments from the main text about measurement error still apply here. That is, the results go through if all moments of the error exist and are bounded. However, there are other important generalisations that I will discuss in this appendix. I will first discuss general functional forms and then turn to a discussion of why bounding the moments for the true input shares and unobserved productivities

A.1 Non-common production

Assumption 12 (Boundedness for finite inputs). For all $\vec{\theta}$ and all closed sets $C \subset \mathbb{R}_+^{\dim(\vec{X})}$, $\sup_{\vec{X} \in C} F(\vec{\theta}, \vec{X}) < \infty$

The non-identification result is also easy to generalise to non-common, non-homogeneous technologies. While an equal split of inputs across outputs is no longer necessarily cost minimizing, for each pair of production function parameters, $(\vec{\theta}_i, \vec{\theta}_{-i})$, we can instead pick an arbitrary cost minimizing allocation and replace $\mathcal{S}_{0.5}$ with the input allocation rule that always chooses this arbitrary point.

The input share will generally depend on the identity of the input, that is $S_{m',i,j,t} \neq S_{m',i,j,t}$; however, there is still only one degree of freedom, that is we can write the firm's input share for each input as a function of the input share of the input share of a given good:

$$S_{m',i,j,t} = S_{m'}(\vec{X}_{j,t}, S_{m,i,j,t}).$$

Assuming that $F(\vec{\theta}, \vec{X})$ remains continuous and differentiable, equal to zero if any of its arguments are equal to zero and strictly increasing in \vec{X} , it is still true that productivity is bounded below and unbounded above:

$$y - f(\vec{\theta}_i, \vec{X}) \leq \omega(\vec{\theta}_i, S_m, \beta_0, Y, \vec{X}) \leq \infty.$$

Therefore, under assumption 12, an analogous argument to the proof in the main text can be made to show that no part of $\vec{\theta}$ it can be identified unless

$$\exists \{\vec{Z}_r\}_{r=1}^{\infty} \text{ s.t. } \lim_{r \rightarrow \infty} |E[x_{m',t} - \rho x_{m',t-1} | \vec{Z}_t = \vec{Z}_r]| = \infty.$$

Showing that a specific element of $\vec{\theta}_i$ can be identified is not possible in general as the identification argument relies on the shape of the set of parameter vectors for which

an input allocation rule exists so that the moment conditions are satisfied.

A.2 Unbounded inputs/productivity

The key step in the proof of theorem 4.1 is showing that $\sup_i \text{ess sup}_{\vec{Z}_t \in \mathbb{Z}_t} \bar{\zeta}(\vec{\theta}, \mathcal{S}_{0.5}, 0 | Z_t, \vec{S}_{t-1:0})$ cannot be made unboundedly smaller by changing the distribution of input shares at time $t - 1$. Assumptions 9-10 ensure this is the case. In this section I will describe what can go wrong if these assumptions are dropped. I will illustrate some of the issues with stylised examples.

First I will focus on assumption 9. Take a two period economy where a set of firms produce two outputs with one input using a Cobb-Douglas technology. Every firm has the same productivity for every output; fix it at 0. For simplicity, let

$$F(\beta^*, X) = X.$$

In period 0 let X be uniformly distributed on $(1, 2)$ and let firms allocate their inputs equally between outputs so that $Y_{i,j,t} = Y_{-i,j,t} = \frac{1}{2}X_{j,t}$. Then all the information is contained in the total input in the previous period, ie $\vec{Z}_1 = X_0$. Now let $X_1 = 2 - X_0$. Then $\lim_{X_0 \rightarrow 2} E[x_1 - \rho x_0 | Z_1] = -\infty$, so β is identified if assumption 9 holds. However, if the distribution of data-generating input allocations at $t = 1$ violates assumption 9, β may be unidentified. If $\lim_{X_0 \rightarrow 2} \frac{x_1}{s_{1,j,1}^*} = 0$ while $\lim_{X_0 \rightarrow 2} s_{1,j,1}^* = -\infty$, then

$$\lim_{X_0 \rightarrow 2} (\beta^* - \beta) E[x_1 - \rho x_0 | X_0] + \beta^* E[s_{1,1}^* - \rho s_{1,0}^* | X_0] = -\infty \forall \beta,$$

while

$$\lim_{X_0 \rightarrow 2} (\beta^* - \beta) E[x_1 - \rho x_0 | X_0] + \beta^* E[s_{2,1}^* - \rho s_{2,0}^* | X_0] = \begin{cases} \infty & \text{if } \beta < \beta^* \\ -\beta^* \ln(0.5) & \text{if } \beta = \beta^* \\ -\infty & \text{if } \beta > \beta^*. \end{cases}$$

Therefore, for the moment conditions to hold \mathcal{S} needs to satisfy

$$\lim_{X_0 \rightarrow 2} E[s_{1,1} - \rho s_{1,0}] = \infty$$

and

$$\lim_{X_0 \rightarrow 2} E[s_{2,1} - \rho s_{2,0}] = \begin{cases} \infty & \text{if } \beta < \beta^* \\ -\beta^* \ln(0.5) & \text{if } \beta = \beta^* \\ -\infty & \text{if } \beta > \beta^*. \end{cases} \quad (12)$$

If $\beta < \beta^*$, this can be accomplished by letting the counterfactual input share in period 0 be a function of X_0 so that $\lim_{X_0 \rightarrow 2} s_{2,0}(X_0) = -\infty$, and choosing a distribution over $s_{1,1}$ such that $\lim_{X_0 \rightarrow 2} E[s_{1,1}|X_0] = -\infty$. If $\beta \geq \beta^*$ it is sufficient to let $s_{2,0} = s_{1,0} = \ln(0.5)$ and following the method in the main text to choose a suitable distribution over $(s_{1,1}, s_{2,1})$.

While such a strategy works here, and can continue to work even if more periods are added, whether there exists any \mathcal{S} that satisfies the moment equalities for a given $\beta \neq \beta^*$ in general is highly dependent on the set of possible paths of observables, \mathbb{Z}_{T+1} . This is the case because, after changing the distribution of the counterfactual input shares in period $t - 1$, while leaving the input shares at 0.5 for all $\tau > t$ – call this input allocation rule \mathcal{S}_t – it will not be true in general that

$$\bar{\zeta}_{i,\tau}(\vec{\theta}_i, \mathcal{S}_t, \beta_0, \rho | \vec{Z}_\tau, \vec{S}_{\tau-1:0}) = \bar{\zeta}_{i,\tau}(\vec{\theta}_i, \mathcal{S}_{0.5}, \beta_0, \rho | \vec{Z}_\tau) \quad \forall \tau > t, \vec{S}_{\tau-1:0} \in \text{supp}(\mathcal{S}_t),$$

making it unclear whether there exists a distribution of input allocations in period τ that can make the moments hold.