Nonparametric Identification Using Timing and Information Set Assumptions with an Application to Non-Hicks Neutral Productivity Shocks*

Daniel Ackerberg[†] Jinyong Hahn [‡] Qingsong Pan[§]

June 11, 2026

Abstract

Recent studies in empirical industrial organization, both in production function and demand estimation, address endogeneity using assumptions regarding the time when agents chose endogenous variables and their information sets at those times. Using a control function framework, we show these assumptions can identify a nonparametric model with a nonseparable unobservable term. Moreover, the model's structure allows a relaxation of the strong support condition typical of control function approaches, which is empirically important in production function contexts. Our empirical application identifies nonseparable (non-Hicks-neutral) shocks in widely-used production datasets, revealing biased technological change consistent with prior literature, but in a distinct manner.

1 Introduction

When estimating structural models in empirical industrial organization, addressing potential endogeneity problems is often an important consideration. There are a number of approaches for addressing such endogeneity problems. This paper studies a non-parametric generalization of one specific approach that has seen wide application in the recent production function literature,

*This paper is a combination of two earlier drafts "Some Nonparametric Identification Results using Timing and Information Set Assumptions" by the first two coauthors (Ackerberg and Hahn (2015)), and "Support Conditions in Control Function Approaches" by the third coauthor (Pan (2021)). All errors are our own.

[†]Department of Economics, UT-Austin, daniel.ackerberg@utexas.edu

[‡]Department of Economics, UCLA, hahn@econ.ucla.edu

[§]School of Economics, Shandong University, qingsongpan@utexas.edu

e.g. Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015), to identify coefficients on "fixed" inputs like capital. The approach requires panel data, and similar to classic fixed effects models decomposes the unobservable into two components - one assumed orthogonal (e.g. independent, mean independent, or uncorrelated) to explanatory variables x_{it}, and one permitted to be correlated with the explanatory variables x_{it}. However, while the classic fixed effect approach involves a time varying component and a time varying component (the former assumed orthogonal to explanatory variables), the decomposition in the literature we consider is different.

More specifically Olley and Pakes (1996) (and the literature based on it) assume that the unobservable causing the endogeneity problem, ω_{it}, follows a nonparametric first order Markov process, i.e., $\omega_{it} = g(\omega_{it-1}) + \xi_{it}$, where $E[\xi_{it}|\omega_{it-1}] = 0$. To identify the production function coefficient on capital k_{it}, they use the assumption that ξ_{it} (but not ω_{it-1}) is mean independent of k_{it}. Loosely speaking, this allows firms' choices of k_{it} to depend on ω_{it-1}, but not ξ_{it}. Ackerberg et al. (2007) describe these as *timing and information set* assumptions, i.e., as economic assumptions regarding 1) the point in *time* at which the agent chooses x_{it}, and 2) the agents' *information sets* at that point in time. Specifically, one interpretation of this assumption is that k_{it} is chosen by firms at time $t-1$ (i.e. a time-to-build assumption) and that ξ_{it} is not in firms' information sets at time $t-1$ (while ω_{it-1} is permitted to be in the firms' information sets at $t-1$).

These timing and information set assumptions of Olley and Pakes (1996) have been used in thousands of research papers in the recent production function literature, and the same general identification strategy is increasingly being used in other contexts. For example some recent work on estimation of demand systems, e.g., Berry et al. (1995), Sweeting (2013), Grennan (2013), Lee (2013), Sullivan (2017), and Dearing (2025), have used timing and information set assumptions to address the problem of endogenous product characteristics and/or prices. Bajari et al. (2012) utilize them in hedonic pricing models, and Pan (2022) uses them to estimate an input demand function. So these timing and information set assumptions can be thought of as a general approach to dealing with endogeneity problems.

This literature using these Olley and Pakes timing and information set assumptions has also worked under the assumption that the relationship between y_{it} and x_{it} is parametrically specified, and that there is an additive separable unobservable term. A few exceptions, in particular Gandhi et al. (2020) and Demirer (2020), allow some nonparametric structure, but continue to maintain an additively separable unobservable term. The goal of this paper is to show that, at least under certain assumptions, these timing and information set assumptions also have identifying power in a nonparametric model with a *nonadditively separable* unobserved term, i.e., where the scalar unobserved term enters the model completely flexibly (up to a strict monotonicity restriction). In other words, we show conditions under which these timing and information set assumptions allow us to identify a nonparametric structural relationship $y_{it} = f_t(x_{it}, \omega_{it})$. Note that related timing and information set assumptions have been made in other literatures, e.g. macro (e.g. Hansen and Singleton (1982)), and labor (e.g. Cunha et al. (2010)) though these models are different in terms of structure (and to our knowledge are not fully non-parametric). Also, in the econometrics panel data literature, papers such as Chamberlain (1982), Anderson and Hsiao (1982), Arellano and Bond (1991), Blundell and Bond (1998) and Blundell and Bond (2000) have extended fixed effects models to allow some parts of the time varying component of the unobservable to be correlated with future x_{it}'s. The "sequential exogeneity", or "predetermined" assumption often used in this work is very similar to the timing and information set assumptions discussed above, but these papers again consider a parametric setting. Thus, our work can also be thought of as a non-parametric generalization of a restricted version of Blundell and Bond (2000) - "restricted" because, analogous to the majority of the literature based on Olley and Pakes, our model does not contain a fixed effect.[1]

To show identification of the fully nonparametric model we study, we use a control function approach following, e.g., Heckman (1977), Blundell and Smith (1989), Blundell and Powell (2001), Matzkin (2004), and Imbens and Newey (2009).[2] We first show how the timing and information set assumptions of our model generate a conditional independence result that allows us to put the model in the framework of Imbens and Newey (2009). However, our model imposes more structure than their general model. In particular, in relation to the canonical "triangular" example of Imbens and Newey (2009), what is akin to the "first stage" equation at one t is simultaneously the "outcome equation" at another t. This means that Imbens and Newey's assumption of a (strictly monotone) scalar unobservable in the first stage equation also implies a scalar unobservable in our outcome equation. We show that this additional structure allows us to substantially relax the "common support" condition required by Imbens and Newey (2009), a condition that Imbens and Newey recognize as quite strong. In particular, we show that interesting structural objects can be identified with very "local" support conditions, and that

[1]The models of Arellano and Bond (1991), Blundell and Bond (1998) and Blundell and Bond (2000) all include a fixed effect in addition to a (parametric) AR(1) process,while literature based on Olley and Pakes does not contain a fixed effect, but allows a more general Markov process (entering linearly in the existing literature, but nonparametrically in our model). Our basic model also does not contain an additional unobservable that is uncorrelated across time, while models such as Blundell and Bond (2000) typically do. However, as discussed in Section 2, one can potentially add such an unobservable to our model with additional assumptions/data for a first stage similar to those used in the Olley and Pakes literature.

[2]Note that our model (and our control function approach) best corresponds to the *second stage* of Olley and Pakes type approaches, as this is where the timing and information set assumptions are utilized. This is potentially confusing because Olley and Pakes approaches also typically involve a *first stage* in which an optimal firm decision, assumed monotonic in the productivity shock, is inverted to distinguish the productivity shock from measurement error, and this is sometimes described as a "control function". The point is that the control function in our paper (we call it this because this is how it is referred to in the econometrics literature, e.g. Imbens and Newey (2009)) is distinct from what is sometimes called a control function in reference to the first stage of Olley and Pakes approaches.

the full model can be identified under conditions considerably weaker than the common support condition in Imbens and Newey (2009). Of course, these results do rely on the above scalar and strict monotonicity assumptions on ω_{it}, but this is a limitation of much of the literature on nonparametric identification when one places no parametric restrictions on the structural function (see, e.g., Matzkin (2007)). The power of the assumption that a structural model (of a scalar outcome) is strictly increasing in a scalar unobservable has been illustrated in a number of previous papers, e.g. Chesher (2003), Chesher (2007); Matzkin (2003), Matzkin (2007); Florens et al. (2008); Chernozhukov and Hansen (2005); D'Haultfœuille and Février (2015); Torgovitsky (2015); Vuong and Xu (2017)). This paper illustrates how this is also case in the context of a model where the key economic assumptions generating identification are timing and information set restrictions. The chaining/sequencing arguments we use to show identification with limited support are particularly reminiscent of those used in D'Haultfœuille and Février (2015) and Torgovitsky (2015). However, these papers consider a different context, a standard triangular system with an exclusion restriction on an exogenous "instrumental variable" Z, and their support conditions concern this Z (they also focus mostly on situations where Z has discrete support, e.g. a binary instrument). Our model has no exogenous Z, and our support conditions concern lagged values of (non-exogenous and continuous) variables X and outcome Y.

We then empirically apply our approach to study properties of production functions. We feel our theoretical extension of timing and information set approaches to models that are not additively separable in the unobservables is particularly important here. This is because in a production function context, a model with only an additively separable unobservable (in log output) corresponds to the assumption of a "Hicks neutral" productivity shock. Such shocks are known to be quite restrictive, and there is both direct and indirect evidence that suggests that there are non-Hicks neutral aspects to productivity shocks (e.g., Balat et al. (2016), Kasahara et al. (2015), Doraszelski and Jaumandreu (2018), Raval (2019), Zhang (2019), Demirer (2020), Raval (2020), Oberfield and Raval (2021), Li and Sasaki (2024)). Our methodology relaxes this assumption, and we apply it to study production functions in three large industries in each of Chile and Colombia. Our estimates also imply non-Hicks neutral productivity shocks, and we examine how these shocks enter our production functions. We find that these shocks interact differently with capital and labor inputs, and, interestingly, that the patterns of these interactions appear to be relatively consistent across the industries we consider. For example, heterogeneity in elasticities of output w.r.t. labor are substantially driven by the non-Hicks neutral productivity shock, while heterogeneity in elasticities of output w.r.t. capital are relatively more driven by variation in observed inputs. We also note how the relaxed support conditions derived in our theoretically work are empirically relevant, as the Imbens and Newey support condition seems easily violated in our production datasets.

Other recent papers have also relaxed the assumption of Hicks neutral productivity shocks - including some of the papers mentioned in the above paragraph. However, we do it in a different way. Other approaches have typically added additional shocks within a parametric structure (e.g., Doraszelski and Jaumandreu (2018) add a labor-augmenting shock in a CES production function). In contrast, we keep a scalar productivity shock, but allow it to enter in a nonparametric way. Ideally, one would want both multidimensional shocks and nonparametric structure, but this is likely not possible while preserving point identification. Because our approach to relaxing Hicks neutrality relies on *different* fundamental assumptions than existing approaches, we see it as *complementary* to them. In other words, *we do not claim that the assumptions behind our approach are preferable to those behind alternative approaches*, but that developing different estimators that identify interesting economic phenomena under *different* assumptions is important to assess robustness. For example, similar to Doraszelski and Jaumandreu (2018), we find evidence that our non-Hicks neutral shocks generate substantial capital bias in technological change, which has important implications on labor markets and wages. The fact that we also find this bias, under quite different assumptions as Doraszelski and Jaumandreu (2018), lends further support to their conclusions.

Our theoretical identification results are directly related to at least three other recent papers. Altonji and Matzkin (2005) also study nonparametric identification in panel situations. They consider nonparametric analogues to fixed and random effects estimators. In their setup, the primary endogeneity problem is generated by an unobservable that is fixed over time. This contrasts with our model that follows the spirit of Olley and Pakes (1996), where the problematic unobservable follows a Markov process with timing and information set assumptions like those described above. It is important to note that while these models are different, neither is a generalization of the other. Hu and Shum (2012) and Hu and Shum (2013) also consider nonparametric identification in a panel setting with Markov structure. Like our paper, the problematic unobservable is assumed to be a scalar and follow a finite Mth order Markov process. In contrast to our quantile based, control function approach to identification, these papers use deconvolution approaches. Our data requirements are weaker than these papers. Specifically, we only require the number of observed time periods T to be at least one greater than the dimension of the Markov process (i.e., $T = M + 1$), i.e., we need to observe at least as many lags as the assumed order of the Markov process. In contrast, Hu and Shum's results require $T > M + 1$, in some cases requiring $T = 3M + 2$. So unlike Hu and Shum, we can estimate a model with a first order Markov process using only two periods of data. On the other hand, Hu and Shum's results apply to models broader than ours in that they allow the outcome variable y_{it} to have a dynamic effects (i.e., y_{it-1} can structurally determine y_{it}).[3] We

[3]Kasahara et al. (2015) also use deconvolution techniques to study identification of a production function with extensive time invariant unobserved heterogeneity, but where the productivity shock is still Hicks neutral.

only consider models without such a dynamic effect. Lastly, work by Navarro and Rivers (2018) is related to our work both in theory and in empirics. Independently of the prior version of this paper Ackerberg and Hahn (2015), Navarro and Rivers (2018) take a different approach to identification of a non-separable production function. By utilizing an assumption that firms are price takers in output markets along with an assumption of firm profit maximization, they are able to consider gross output production functions. Like Doraszelski and Jaumandreu (2018), they find evidence of capital biased technological change in these gross output production functions, so our empirical finding of similar patterns in value added production functions is also consistent with theirs.

2 Setup

Our goal is to use panel data on observables $\{x_{it}, y_{it}\}$, $i = 1, ..., N$, $t = 1, ..., T$ to identify the structural equation

$$y_{it} = f_t(x_{it}, \omega_{it}), \tag{1}$$

where $f_t : \mathcal{S}_t^x \times \mathcal{S}_t^\omega \to \mathcal{R}$ is differentiable in (x_{it}, ω_{it}) and strictly increasing in ω_{it}, $\mathcal{S}_t^x \in \mathcal{R}^{d_x}$ is the support of x_{it}, $\mathcal{S}_t^\omega \in \mathcal{R}$ is the support of ω_{it}, x_{it} has a continuous distribution,[4] and ω_{it} is a *scalar* unobservable term that is also continuously distributed.[5]

The scalar and strict monotonicity restrictions on ω_{it} are assumptions that are commonly used in the nonparametric identification literature when one treats a scalar valued structural function f_t completely nonparametrically. They can sometimes be theoretically motivated, for example in the production function literature it is natural to assume that output is strictly increasing in inputs - this could be true with respect to observed inputs of production like capital and labor, or unobserved inputs (e.g. unmeasured managerial ability). Note that we allow the structural functions f_t to change in arbitrary ways over time, but the model is not "dynamic" in the sense that y_{it-1} does not directly determine y_{it}. We consider identification of the structural functions f_t under the assumption that $N \to \infty$ and T is fixed.

We consider a situation where the vector of observables x_{it} is chosen by an economic agent. We start with our key *timing and information set* assumption:

Assumption 1 *(Timing and Information Set) At the time x_{it} is chosen, the agent's information set is $\mathcal{I}_{it-1} = \left\{ \{y_{i\tau}\}_{\tau=1}^{t-1}, \{x_{i\tau}\}_{\tau=1}^{t-1}, \{\omega_{i\tau}\}_{\tau=1}^{t-1}, \{\eta_{i\tau}\}_{\tau=1}^{t-1} \right\}$, where η_{it} are additional unobservables that we describe below.*

They require $T = 4$ for identification of a model where the productivity shock follows a first order Markov process.

[4] With some slight adaptations, our approach also applies to the case where x_{it} is discrete.

[5] Throughout the paper, for the convenience of exposition, we assume all the distributions (joint or marginal) have positive densities over their respective support.

This assumption implies that our economic agents are choosing x_{it} *without* knowledge of the period t structural unobservable ω_{it}, but *with* knowledge of ω_{it-1} (and y_{it-1} and x_{it-1}, and histories of these variables).[6] Since we will allow serial correlation in ω_{it}, x_{it} and ω_{it} can be correlated in this model even though x_{it} is chosen before the agent observes ω_{it}. This is because x_{it} may be chosen as a function of ω_{it-1} and ω_{it-1} may be correlated with ω_{it}.

The agent's information set when choosing x_{it}, $\mathcal{I}_{it-1}$, also includes econometric unobservables η_{it-1}. These are other factors that may affect the agent's payoffs and thus the optimal choice of x_{it}. Note that other than the timing and informational set assumptions, our model is quite general. One nice attribute of our approach is that we will not need to explicitly specify agents' payoffs for our identification results. For example, $x_{it} = h_t(\mathcal{I}_{it-1})$ may be the solution to a dynamic programming problem that would require many other auxiliary assumptions to solve. We will not need to specify h_t, and thus can essentially be agnostic about these auxiliary assumptions.

2.1 Relation to the Empirical Literature

The leading example of these types of assumptions being used in practice is the widely cited and applied Olley and Pakes (1996) approach to estimating production functions. In this context, y_{it} is output (or revenue), x_{it} are inputs chosen by the firm (e.g., capital, labor, R&D) and ω_{it} is an unobservable "productivity" shock. Typically in this literature, at least some of the inputs in x_{it} are assumed to satisfy Assumption (1), i.e., to be chosen *prior* to the firm learning ω_{it}. For example, in Olley and Pakes (1996) the capital input is assumed to satisfy Assumption (1), while in Gandhi et al. (2020) both capital and labor are assumed to satisfy Assumption (1). This is described as a "timing and information set" assumption because, e.g. in Olley and Pakes (1996), it involves both an assumption that firms must commit to their period t capital stock at $t-1$ (a timing assumption)[7] and the assumption that ω_{it} is not observed by firms until period t (an information set assumption). Note that different combinations of timing and information set assumptions can also be consistent with (1). For example, if one assumed that agents do not observe ω_{it} until period $t+1$, then x_{it} could be chosen at t.[8] In the production function context, the unobservable η_{it-1} could represent a multidimensional set of factors affecting input

[6]For simplicity we characterize information sets in terms of random variables rather than using a more formal sigma-algebra description. We discuss how one can potentially relax the timing and information set assumptions in section (5).

[7]This reflects a "time-to-build" assumption on capital or an assumption that labor requires time to adjust. The appropriability of these timing assumptions will depend on the industry being studied and the time frame of the data (e.g., annual vs quarterly vs daily).

[8]Analagously, if ω_{it} was for some reason observed ahead of time at period $t-1$, then x_{it} could need to be chosen at $t-2$. See Ackerberg (2020) for more discussion of this. Also note that the related panel data literature described in the introduction, which makes similar assumptions, might describe this assumption as one of x_{it} being "predetermined".

and output prices (or those prices themselves if they are competitively set). Typically, such factors will impact optimal choices of x_{it}.[9]

In the production function context, Equation (1) is best thought of as a non-parametric version of the *second stage* of Olley and Pakes (1996) and related procedures. Their second stage equation has the form $y_{it} = f_t(x_{it}) + \omega_{it}$, i.e. where the scalar ω_{it} enters additively (instead of our non-additive, nonparametric f). Like in that literature, with auxiliary data/assumptions one could potentially add additional unobservables to our model that are identified in a preliminary stage, i.e. the analogue of the *first stage* of Olley and Pakes (1996). For example, the "Technology Control" approach detailed in Ackerberg and De Loecker (2024) uses additional assumptions and data on a Leontief intermediate input m_{it} in the model $y_{it} = max\{f_t(x_{it}) + \omega_{it}, \gamma_t m_{it}\} + \epsilon_{it}$ where ϵ_{it} is an unanticipated shock to output or measurement error in y_{it} that is uncorrelated with input choices. In this model, one can easily identify ϵ_{it} in a first stage, and the same logic is applicable to the model $y_{it} = max\{f_t(x_{it}, \omega_{it}), \gamma_t m_{it})\} + \epsilon_{it}$. Under the assumption of no unutilized inputs in this latter model (i.e. $f_t(x_{it}, \omega_{it}) = \gamma_t m_{it}$) after such a first stage identifies ϵ_{it}, the model reduces to our Equation (1) with scalar unobservable ω_{it}.[10]

While production functions may be the predominant context for use of timing and information set assumptions to address endogeneity problems, the assumptions have also been utilized in the literature on estimating demand, particularly in addressing potentially endogeneity of product characteristics. For example, Sweeting (2013), Grennan (2013), Lee (2013), and Sullivan (2017) assume that product characteristics take time for a firm to design and change so that they must be decided before the firm observes the period t demand shock. In other words, they make the timing and information set assumption that while period t product characteristics can be chosen as a function of prior periods' demand shocks, they cannot be chosen as a function of the current period demand shock.

In this demand estimation context, the model we study, i.e. $y_{it} = f_t(x_{it}, \omega_{it})$, can be thought of as a simple non-additive, non-parametric demand model, where y_{it} is quantity demanded, x_{it} includes product characteristics, demand shifters, and prices, and ω_{it} is a scalar unobserved demand shock (η_{it-1} might represent cost shocks that affect firms' choices of product characteristics and prices). Thus, our identification results can be directly applied to at least some demand estimation situations, e.g. with a panel of monopoly or monopolistically competitive markets. They can also be directly applied to a multiproduct logit demand model where y_{it}

[9] In this formulation, we are using η_{it-1} to denote the price paid for (or factors influencing the price paid for) inputs x_{it}. But the indexing of η is irrelevant. For example, if one prefers to index these instead by t (i.e. η_{it}), one can simply include η_{it} in $\mathcal{I}_{it-1}$.

[10] Note that while we can "tag on" this Leontief based first stage to allow an additional unobservable into our model, our model and assumptions are not consistent with *all* first stages that have been used in the Olley and Pakes related literature. For example, first stages that use an element of x_{it} to "invert" the productivity shock are not consistent with our model because of our stronger timing assumption on x_{it}, and because we allow η_{it}'s that would violate the scalar unobservable assumption of these inversions.

is defined as the log share ratio for product i in market t and $f_t(x_{it}, \omega_{it})$ represents a non-parametric, non-additive utility function depending on observed product characteristics/prices and a product specific unobserved characteristic/demand shock. Note, however, that our results do not directly apply to more sophisticated demand models, e.g. the BLP model of differentiated product demand. This is because in these aggregated discrete choice models the demand for one product in the market depends in a complicated way on the unobserved demand (or utility) shocks of all competitors in the market, thus violating our assumption that ω is scalar (it also depends on product characteristics of all products, but to the extent these are observed they can be included in x_{it}). Recent work by Dearing (2025) explicitly considers this situation, i.e. he studies identification using timing and information set assumptions in the specific context of a BLP random coefficients demand model. He not only considers the case where both price and product characteristics satisfy Assumption (1), but also cases where only product characteristics are assumed to satisfy (1) while one has access to standard instrumental variables to instrument for endogenous price.[11] This latter situation is similar in spirit our Section 5.2, where we consider relaxing timing assumptions on some elements of x_{it} (while keeping the scalar unobservable assumption). Lastly, other applications using these types of timing and information set assumptions include Bajari et al. (2012), who apply them in hedonic pricing models, and Pan (2022), who uses the techniques here in a situation where equation (1) is an input demand function for a variable input y_{it} conditional on a fixed input x_{it}.

2.2 Additional Assumptions

For our nonparametric identification arguments we make the following additional assumption on the structural unobservable ω_{it}.

Assumption 2 *(Mth Order Markov Process) The distribution of the scalar ω_{it} satisfies $p_t(\omega_{it} \mid \mathcal{I}_{it-1}) = p_t\left(\omega_{it} \mid \ \{\omega_{i\tau}\}_{\tau=t-M}^{t-1}\right)$, where $T \geq M+1$.*

Assumption (2) allows the distribution of ω_{it} to vary across time and be specified nonparametrically. On the other hand, Assumption (2) may be argued to be restrictive because we assume that ω_{it} evolves "exogenously" in the sense that conditional on ω_{it} and past values of ω_{it}, the distribution of ω_{it+1} does not depend on values of the other variables in the model dated t and earlier.[12] We also assume that ω_{it} follows a finite Mth order Markov process.[13] This is

[11]In contrast to our *non-additively separable*, non-parametric model, the results in Dearing (2025) do follow the majority of the BLP literature by relying on an *additive separability* assumption regarding how the unobserved demand shocks enter the utility function.

[12]However, our approach can be generalized to allow for a controlled Markov process as in Doraszelski and Jaumandreu (2013), as long as the control variable is observed.

[13]When $M=0$, we define $\{\omega_{i\tau}\}_{\tau=t-M}^{t-1} = \emptyset$. Obviously this is not a particularly interesting case, because in

a generalization (to $M > 1$) of the assumption typically made on ω_{it} in the Olley and Pakes related literature, but unlike Arellano and Bond (1991), Blundell and Bond (1998), Blundell and Bond (2000), and Altonji and Matzkin (2005), Assumption (2) does not allow there to be a component of ω_{it} that is fixed over time (e.g. a fixed or random effect). [14] On the other hand, we do not require the exchangeability assumption of Altonji and Matzkin (2005).

We only need one more period of data than the order of the Markov process ($T = M + 1$) to obtain identification, i.e., we need to observe a number of lags equal to the assumed order of the Markov process. This is less than what is required by Hu and Shum (2012) and Hu and Shum (2013) in their deconvolution approaches to identification in related models - they require more than $M + 1$ periods (in some cases they require up to $T = 3M + 2$). This is because these deconvolution approaches use restrictions the Markov structure places on correlations between data in time periods that are further apart than the assumed length of the Markov process (e.g. correlations between $t = 1$ and $t = 3$ variables with a first order Markov process). In contrast, in our approach, e.g., if ω_{it} follows a first order Markov process, then we only need two periods of data.

Note that f_t is permitted to vary by t in our model. In our main approach based on control functions, we need to observe all M lags to identify f_t at a particular t. For example, when $M = 1$, we cannot identify f_t for $t = 1$, but we can identify f_t for all the later periods. However, in section (6) we detail a related identification approach that can identify f_t for the initial time periods in the data (i.e., for $t \leq M$).

Given Assumption (2), our model places very few restrictions on the other econometric unobservables, the η_{it}. We do not need to limit the dimension of η_{it}, and the η_{it}'s can be contemporaneously correlated with ω_{it}, and η_{it}'s can be correlated in any way with $\mathcal{I}_{it-1}$ (which includes past values of η). In addition, the distribution of η_{it} can change over time. The key restriction of the model, embodied in Assumption (2), is that the distribution of ω_{it} given $\{\omega_{i\tau}\}_{\tau=t-M}^{t-1}$ does not depend on any past η's. While this assumption may be strong, it is an essential element of basically all the literature stemming from Olley and Pakes (1996). As detailed at length later, we also require support conditions - essentially that there is "enough" variation in η_{it-1} to generate sufficient variation in x_{it} given ω_{it-1}.

Given Assumption (2), we can express

$$\omega_{it} = g_t\left(\{\omega_{i\tau}\}_{\tau=t-M}^{t-1}, \xi_{it}\right), \tag{2}$$

this case, our assumptions imply that ω_{it} is independent of x_{it}, and identification of f_t is trivial using Matzkin (2003).

[14] As discussed in Ackerberg (2020), the tradeoff is that the while the aforementioned literature allows a fixed effect, it typically makes a stronger assumption on the Markov process than does the Olley and Pakes related literature and our model, i.e. it assumes ω_{it} follows a linear AR(1) process.

where WLOG g_t is strictly increasing in ξ_{it}, a scalar unobservable that is independent of $\mathcal{I}_{it-1}$. We make the additional assumptions that

Assumption 3 *g_t is differentiable in its arguments.*

These regularity conditions require the conditional density p_t to be sufficiently smooth - for example, for g_t to be strictly increasing in ξ_{it}, p_t cannot have mass points. Then, by Matzkin (2007), we without loss of generality make the following normalizations:

Assumption 4 *(Normalizations) At each t, either:*
i) ω_{it} and ξ_{it} have uniform marginal distributions on $(0, 1)$,
ii) or $f_t^{-1}(0, y_{it}) = y_{it}$ and ξ_{it} has a standard normal marginal distribution.

One can show that these are equivalent normalizations given the two nonparametric functions f_t and g_t, and we will use i) to show identification and use ii) for estimation.

Before proceeding with our formal identification arguments, we describe the intuition behind identification in this model. This intuition is actually quite simple. Substituting in lagged (2) into (1) results in

$$y_{it} = f_t\left(x_{it}, g_t\left(\{\omega_{i\tau}\}_{\tau=t-M}^{t-1}, \xi_{it}\right)\right).$$

Assumption (1) implies that x_{it} is chosen as a function of only $\mathcal{I}_{it-1}$, and ξ_{it} is a scalar unobservable that, given Assumption (2) is independent of $\mathcal{I}_{it-1}$. Therefore, x_{it} is independent of ξ_{it} (in fact, $(x_{it}, \mathcal{I}_{it-1})$ is jointly independent of ξ_{it}). Because f_t is strictly monotone in ω_{it} for all t, conditioning on M lags of $\{x_{it}, y_{it}\}$ is equivalent to conditioning on M past values of ω_{it}. Hence, conditional on $\{x_{i\tau}, y_{i\tau}\}_{\tau=t-M}^{t-1}$, variation in x_{it} that is independent of ξ_{it} can be used to identify aspects of f_t.

3 Control Function Approach

More formally, focus attention on one particular $t \geq M + 1$. Let x_{it}^1 be the first component of x_{it} and define the random variable

$$\varsigma_{it}^1 = F_{x_{it}^1|\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}}\left(x_{it}^1, \{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right).$$

Now, we consider the second element of x_{it} conditional on $\{y_{i\tau}\}_{\tau=t-M}^{t-1}$, $\{x_{i\tau}\}_{\tau=t-M}^{t-1}$, and ς_{it}^1, i.e., $F_{x_{it}^2|\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}, \varsigma_{it}^1}$. Define the random variable

$$\varsigma_{it}^2 = F_{x_{it}^2|\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}, \varsigma_{it}^1}\left(x_{it}^2, \{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}, \varsigma_{it}^1\right).$$

By iterating this process, we can create $\varsigma_t = (\varsigma_t^1, \varsigma_t^2, \ldots)$.

Theorem 1 *x_{it} is independent of ω_{it} given* $\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$.

Proof. Lemma (6) in the Appendix uses Assumptions (1) and (2) to show that ξ_{it} and ς_{it} are independent of each other given $(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1})$. Next note that x_{it} can be written as a function of $\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$ and ς_{it}, say $x_{it} = \varphi_t\left(\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right), \varsigma_{it}\right)$. [15] Also, since $\omega_{it} = f_t^{-1}(x_{it}, y_{it})$, we can see that $\omega_{it} = g_t\left(\{\omega_{i\tau}\}_{\tau=t-M}^{t-1}, \xi_{it}\right)$ can be written as a function of $\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$ and ξ_{it}, say $\omega_{it} = \phi_t\left(\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right), \xi_{it}\right)$. The Theorem then follows because separate functions of independent random variables are also independent. ■

Theorem (1) establishes that in our model based on timing and information set assumptions, Assumption 1 of Imbens and Newey (2009) holds. This allows us to identify $y_{it} = f_t(x_{it}, \omega_{it})$ using $v_{it-1} = \left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$ as a control function. More specifically, consider the same support condition as Imbens and Newey, i.e.,

Assumption 5 *(Assumption 2 of Imbens and Newey (2009): Common Support) For all x_{it} in the support, the support of v_{it-1} conditional on x_{it} equals the support of v_{it-1}.*

Since f_t is strictly monotone in ω_{it}, to identify $y_{it} = f_t(x_{it}, \omega_{it})$ it suffices to identify the inverse function of f_t, i.e., to identify the ω^0 corresponding to any value of $(x_{it}, y_{it}) = (x^0, y^0)$. With $f_{v_{it-1}}$ denoting the density function of $v_{it-1} = \left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$ we can obtain this ω^0 using the following equation:

$$\begin{aligned} \omega^0 &= Pr(f_t(x^0, \omega_{it}) \le y^0) \\ &= \int Pr(f_t(x^0, \omega_{it}) \le y^0 | v_{it-1} = v) f_{v_{it-1}}(v) dv \\ &= \int Pr(f_t(x^0, \omega_{it}) \le y^0 | x_{it} = x^0, v_{it-1} = v) f_{v_{it-1}}(v) dv \\ &= \int Pr(y_{it} \le y^0 | x_{it} = x^0, v_{it-1} = v) f_{v_{it-1}}(v) dv. \end{aligned} \tag{3}$$

The first equality follows from the normalization $\omega_{it} \sim U(0,1)$. The second equality follows from law of iterated expectation. The third equality follows because conditional on $v_{it-1} = v$, ω_{it} is independent from of x_{it} so we can further condition on $x_{it} = x^0$. The last line follows from the fact that the observed y_{it} is generated by f_t.

Focusing on the last line of (3), the marginal density of v_{it-1}, $f_{v_{it-1}}$, can be directly identified by the data. $Pr(y_{it} \le y^0 | x_{it} = x^0, v_{it-1} = v^0)$ is also directly identified at every point (x^0, v^0) on the joint support of (x_{it}, v_{it-1}). So as long as the Imbens and Newey support condition, i.e.,

[15] This follows by construction of the ς_{it} above, as long as the conditional densities of the x_{it}'s are non-zero on their supports.

Assumption (5), holds, f_t is identified for all $t > M$. It is also clear why this approach doesn't work for $t \leq M$ (e.g., the first period of data when $M = 1$), as for these early time periods we do not observe v_{it-1}.[16]

4 Relaxing Support Conditions

As acknowledged by Imbens and Newey (2009), the support condition of Assumption (5) is quite strong. We feel this is certainly the case in a production function context. Since v_{it-1} includes x_{it-1}, Assumption (5) requires, for example, the support of the distribution of firm's capital stock at t to not depend on that firm's capital stock at $t-1$. This seems counter to the well-known observation that across firm variation in capital stocks is typically much larger than within firm variation. In fact, anticipating our empirical application, Figure (2) in Section (8.3) plots the empirical joint distribution of current vs lagged capital stock for the 6 datasets we consider. Assumption (5) would require these supports to be rectangular, which seems far from the case. Thus, estimation using Equation (3) seems unlikely to be successful since it would require extrapolating far from where there is any data to estimate the probabilities in the last line.

Given this limitation, we now investigate whether the Imbens and Newey support condition can be relaxed in the context of our timing and information set assumption based model. Note that Imbens and Newey investigate bounds on objects of interest when their support condition does not hold. They illustrate how the informativeness of these bounds can vary widely depending on the model and the object of interest. In contrast, we take a different approach. We show that the additional structure of our model allows us to significantly relax the Imbens and Newey support condition yet still obtain point identification of many objects of interest. The additional structure in our model that allows us to do this - that the scalar unobservable ω_{it} follows an Mth order Markov process conditional on $\mathcal{I}_{it-1}$ - is already an integral component of the Olley and Pakes (1996) related literature that we are aiming to extend. In other words, we do this by leveraging assumptions that are often already being made in these contexts. It is also interesting to relate our additional structure to the triangular model that Imbens and Newey (2009) consider as a leading example of their control function methods. In that model, the control function (first stage) equation is assumed to have a scalar unobservable (though the second stage structural equation of interest can have multidimensional unobservables). In our model, the control function is essentially a lagged version of the structural equation of interest, so in a sense a single scalar unobservable assumption results in a scalar unobservable in both the control function and the structural function.

[16] Later we illustrate an alternative identification strategy that can be used to identify f_t for $t \leq M$ in our model.

We denote the joint support of $(x_{it}, v_{it-1}, y_{it})$ as $\mathcal{S}_t^{xvy}$ (similarly for $\mathcal{S}_t^{xv}$, $\mathcal{S}_t^{xy}$, $\mathcal{S}_t^{v}$, etc.), and the conditional support of x_{it} given $v \in \mathcal{S}_t^v$ as $\mathcal{S}_t^{x|v}$ (similarly for $\mathcal{S}_t^{v|x}$, $\mathcal{S}_t^{y|xv}$, etc.) While we relax Imbens and Newey's support condition, i.e., Assumption (5), all the results below use Assumptions (1), (2), (3), and (4) (unless otherwise indicated). Together with the structural equation (1), these five Assumptions constitute our model.

4.1 Partial Identification Result With Relaxed Support Condition

We start with a very simple result that makes very limited, local, support assumptions on the distribution of (x_{it}, v_{it-1}). It is only a partial identification result in that we will not identify the full structural function $y_{it} = f_t(x_{it}, \omega_{it})$ (we do identify the full structural function momentarily). However, aspects of f_t that we do identify are point identified.

Condition 1 *(Small Local Support at (x^0, v^0)) For some $\epsilon > 0$, the conditional distribution of x_{it} given $v_{it-1} = v^0$ has positive density on all x satisfying $\|x - x^0\| < \epsilon$.*

If the joint distribution of the data satisfies support condition (1) at (x^0, v^0), it means that there is some local variation in x_{it} given $v_{it-1} = v^0$ - this is necessary to identify derivatives w.r.t. x_{it}. Denote by $f_t^{-1}(x_{it}, y_{it})$ the inverse of $f_t(x_{it}, \omega_{it})$ w.r.t. its second argument. Recalling that $v_{it-1} = \left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$, we keep the notation simple (with some abuse) by letting $f_{t-1}^{-1}(v_{it-1}) = \{f_\tau^{-1}(x_{i\tau}, y_{i\tau})\}_{\tau=t-M}^{t-1}$. In the baseline case of $M = 1$ this reduces to simply $f_{t-1}^{-1}(v_{it-1}) = f_{t-1}^{-1}(x_{it-1}, y_{it-1})$). We then have

Theorem 2 *If the density of (x_{it}, v_{it-1}) satisfies support condition (1) at some (x^0, v^0), then $\frac{\partial f_t(x_{it}, \omega_{it})}{\partial x_{it}}$ is identified at the points $x_{it} = x^0$ and $\omega_{it} = g_t(f_{t-1}^{-1}(v^0), \xi^0)$ for any $\xi^0 \in (0, 1)$.*

Proof. Plugging in g_t for ω_{it}, we have

$$\begin{aligned}
y_{it} &= f_t(x_{it}, \omega_{it}) \\
&= f_t(x_{it}, g_t(\{\omega_{i\tau}\}_{\tau=t-M}^{t-1}, \xi_{it})) \\
&= f_t(x_{it}, g_t(f_{t-1}^{-1}(v_{it-1}), \xi_{it})) \\
&= \bar{f}_t(x_{it}, v_{it-1}, \xi_{it}).
\end{aligned} \tag{4}$$

This implies that the derivative of $\bar{f}_t$ with respect to x_{it} evaluated at (x^0, v^0, ξ^0) is equal to the derivative of f_t with respect to x_{it} evaluated at $(x^0, g_t(f_{t-1}^{-1}(v^0), \xi^0))$. Since (x_{it}, v_{it-1}) are independent of the scalar ξ_{it}, under our normalization $\xi_{it} \sim U(0, 1)$ we can identify the reduced form function $\bar{f}_t$ at $v_{it-1} = v^0$ and all x satisfying $\|x - x^0\| < \epsilon$. For any $\xi^0 \in (0, 1)$, this identifies the derivatives of $\bar{f}_t$ w.r.t. x_{it} at (x^0, v^0, ξ^0) and hence the derivatives of f_t w.r.t. x_{it} at $\left(x^0, g_t(f_{t-1}^{-1}(v^0), \xi^0)\right)$. ■

It is important to note that this result does not identify $f_t(x_{it}, \omega_{it})$ (or its derivative) at any specific point (x_{it}, ω_{it}). What it is essentially doing is identifying $\frac{\partial f_t}{\partial x_{it}}$ at x^0 and an "unknown" point in the support of ω_{it} - the point $g_t(f_{t-1}^{-1}(v^0), \xi^0)$ (for v^0 and any value of ξ^0). It is an "unknown" point because we do not assume knowledge of the functions f_{t-1}^{-1} or g_t. In other words, we cannot answer some counterfactual questions with this result - e.g., what would y be given $x_{it} = x^0$ and ω_{it} equals some candidate value $\in (0, 1)$ (recall the normalization $\omega_{it} \sim U(0, 1)$).

However, we can answer other interesting counterfactual questions with this result. In particular, it allows us to identify the derivative of the outcome y_{it} with respect to a change in x_{it} for any observation *in the data* who have (x^0, v^0) such that the local support condition holds. In other words, we can answer questions about counterfactual y_{it}'s for observations in the data, if their x_{it} were changed locally.[17] Note that we are able to obtain this result (unlike Imbens and Newey) because the scalar unobservable assumption on ω_{it} allows us to identify ξ_{it} for each observation in the data (as a byproduct of identifying $\bar{f}_t(x_{it}, v_{it-1}, \xi_{it})$). These can be important counterfactuals. For example, in our application to production functions, Theorem (2) implies we can identify the input elasticities of output for each firm in the data under only local regularity conditions. In a demand context it would allow identification of, e.g. price or characteristic elasticities for observations in the dataset.

4.2 Full Identification Results with Relaxed Support Conditions

We now turn to identifying the full $f_t(x, \omega)$ at any specific point (x_{it}, ω_{it}). Again, we will show how, in our model, this can be done with weaker support conditions than used by Imbens and Newey (2009). We start with the following observation that will be useful. Since the scalar ω_{it} can only be identified up to a monotone transformation in our model (hence our normalization $\omega_{it} \sim U(0, 1)$), to identify $f_t(x, \omega)$ it suffices to be able to order any pair of points (x^A, y^A) and (x^B, y^B) in the support $\mathcal{S}_t^{xy}$ in terms of their associated ω, i.e., to be able to compare

$$\omega^A = f_t^{-1}(x^A, y^A) \qquad \text{vs} \qquad \omega^B = f_t^{-1}(x^B, y^B).$$

We formalize this observation in the following lemma based on Debreu (1954).

Lemma 1 *$f_t(x, \omega)$ is identified if and only if for any two points $(x^A, y^A), (x^B, y^B) \in \mathcal{S}_t^{xy}$, we can order $f_t^{-1}(x^A, y^A)$ and $f_t^{-1}(x^B, y^B)$ (i.e. we can identify whether $f_t^{-1}(x^A, y^A) > f_t^{-1}(x^B, y^B)$, $f_t^{-1}(x^A, y^A) < f_t^{-1}(x^B, y^B)$, or $f_t^{-1}(x^A, y^A) = f_t^{-1}(x^B, y^B)$).*

Proof. It is easy to prove the "only if" part. If $f_t(x, \omega)$ is identified, then $f_t^{-1}(x, y)$ is identified. Thus, given any two points $(x^A, y^A), (x^B, y^B) \in \mathcal{S}_t^{xy}$, $f_t^{-1}(x^A, y^A)$ and $f_t^{-1}(x^B, y^B)$ can be

[17] Or more than locally depending on the support of $x_{it}|v^0$.

ordered.

For the "if" part, the proof can be borrowed from the classic proof for the existence of a continuous utility function by Debreu (1954). Since for any two points $(x^A, y^A), (x^B, y^B) \in \mathcal{S}_t^{xy}$ we can order $f_t^{-1}(x^A, y^A)$ vs $f_t^{-1}(x^B, y^B)$, we can identify the binary relation $\succsim = \{((x^A, y^A), (x^B, y^B)) \in \mathcal{S}_t^{xy} \times \mathcal{S}_t^{xy} : f_t^{-1}(x^A, y^A) \geq f_t^{-1}(x^B, y^B)\}$ on $\mathcal{S}_t^{xy}$. It is easy to see that $\succsim$ is complete and transitive. Since $f_t(x, \omega)$ is continuous in (x, ω) and strictly monotone in ω, by the implicit function theorem, $f_t^{-1}(x, y)$ is continuous in (x, y). As a result, the upper and lower contour sets are closed. Finally, note that $\mathcal{S}_t^{xy}$ is a subspace of the Euclidean space, so it is perfectly separable.

Then, by Theorem II of Debreu (1954), there exists a continuous function $M_t(x, y)$ such that $M_t(x^A, y^A) \geq M_t(x^B, y^B) \Leftrightarrow (x^A, y^A) \succsim (x^B, y^B)$. It is straightforward to use the identified $\succsim$ to construct such an $M_t(x, y)$, see e.g., Jaffray (1975) and Rubinstein (2012).[18] We know the identified $M_t(x, y)$ is a monotone transformation of $f_t^{-1}(x, y)$ since $(x^A, y^A) \succsim (x^B, y^B) \Leftrightarrow f_t^{-1}(x^A, y^A) \geq f_t^{-1}(x^B, y^B)$ (by definition of $\succsim$), and therefore $f_t^{-1}(x^A, y^A) \geq f_t^{-1}(x^B, y^B) \Leftrightarrow M_t(x^A, y^A) \geq M_t(x^B, y^B)$. To recover $f_t^{-1}(x, y)$ from $M_t(x, y)$, define $e_{it} = M_t(x, y)$. Since the joint density of (x, y) is identified and $M_t(x, y)$ is identified, the cumulative distribution of e_{it}, i.e., $F_{e_{it}}$, is identified. Thus, given our normalization $\omega_{it} \sim U(0, 1)$, we know $f_t^{-1}(x, y) = F_{e_{it}}(M_t(x, y))$. This identifies $f_t^{-1}(x, y)$, and thus $f_t(x, \omega)$ is identified. ■

With this Lemma in hand we now consider a sequence of successive support conditions, each progressively less restrictive than the previous one, which illustrate various support conditions that ensure that $f_t^{-1}(x^A, y^A)$ and $f_t^{-1}(x^B, y^B)$ can be ordered.

First consider

Condition 2 *There is a x^0 such that for any $v \in \mathcal{S}_t^v$, $x^0 \in \mathcal{S}_t^{x|v}$.*

Support condition (2) weakens Imbens and Newey's Assumption (5). While Imbens and Newey require v_{it-1} to have full support conditional on any x, condition (2) only requires v_{it-1} to have full support at *one particular* x^0. Using the Lemma, one can show

Theorem 3 *Under the assumptions of our model and support condition (2), $f_t(x, \omega)$ is identified for all $t > M$.*

Proof. The Lemma requires us to be able to order $f_t^{-1}(x^A, y^A)$ and $f_t^{-1}(x^B, y^B)$ at any $(x^A, y^A), (x^B, y^B) \in \mathcal{S}_t^{xy}$. We can do that using the point x^0. First, find any v^A and v^B

[18] Since monotone transformations preserve ordering, there is not a unique $M_t(x, y)$ such that $M_t(x^A, y^A) \geq M_t(x^B, y^B) \Leftrightarrow (x^A, y^A) \succsim (x^B, y^B)$. In other words, we have identified just one continuous function $M_t(x, y)$ representative of the binary relation $\succsim$.

s.t. (x^A, y^A, v^A) and (x^B, y^B, v^B) are both in the support of the data $\mathcal{S}_t^{xyv}$. Next, define

$$y^{0A} = F^{-1}_{y_{it}|x^0,v^A}\left(F_{y_{it}|x^A,v^A}\left(y^A\right)\right) \qquad \text{and} \qquad y^{0B} = F^{-1}_{y_{it}|x^0,v^B}\left(F_{y_{it}|x^B,v^B}\left(y^B\right)\right).$$[19]

y^{0A} and y^{0B} are identified by the data because support condition (2) ensures (x^0, v^A) and (x^0, v^B) are in $\mathcal{S}_t^{xv}$, i.e. there is data at (x^0, v^A) and (x^0, v^B) to identify the inverse CDF $F^{-1}_{y_{it}|x,v}$ at these values of the conditioning variables. By construction, the points (x^A, y^A, v^A) and (x^0, y^{0A}, v^A) are associated with the same unobserved ξ's, and since v^A is the same at both points (and $\omega = g_t(f^{-1}_{t-1}(v), \xi)$), the two points must also be associated with the same ω's, i.e. $f_t^{-1}\left(x^0, y^{0A}\right) = f_t^{-1}(x^A, y^A)$. The same is true regarding y^{0B}, i.e. $f_t^{-1}\left(x^0, y^{0B}\right) = f_t^{-1}(x^B, y^B)$. Thus, whether $f_t^{-1}(x^A, y^A) > f_t^{-1}(x^B, y^B)$ depends on whether $f_t^{-1}(x^0, y^{0A}) > f_t^{-1}(x^0, y^{0B})$, which, since f_t^{-1} is strictly monotone in its second argument, corresponds to whether $y^{0A} > y^{0B}$. In sum, the ordering of y^{0A} and y^{0B} (which are identified by the data) determine the ordering of $f_t^{-1}(x^A, y^A)$ and $f_t^{-1}(x^B, y^B)$ ■

Intuitively, support condition (2) allows us to order any $f_t^{-1}(x^A, y^A)$ and $f_t^{-1}(x^B, y^B)$ by using the "special" point x^0, i.e. finding points in the data with $x = x^0$ and the same ω's as the original points. A comparison of the y's at these new points then orders the original points. Clearly, the ability to do the comparisons in the proof again depends crucially on the scalar ω in our model. This is what allows us to find points with the same ω. This idea of finding points with equal ω is crucial for further weakening the support condition, and for the remainder of this section we use the term "iso-omegic" to describe points with the same omega, i.e. in the proof of Theorem (3) (x^0, y^{0A}) and (x^0, y^{0B}) are "iso-omegic" to the points being compared, (x^A, y^A) and (x^B, y^B) respectively.

We next show how identification can be preserved under still more relaxed support conditions. The next four conditions are each a strict relaxation of the prior one (and condition (2)). Though the sequence of conditions is useful to understand the mechanics of identification, to save space we only present a Theorem for the last condition, as this is most general of the four. As noted in the introduction, the chaining/sequencing arguments we use here are reminiscent of those used by D'Haultfœuille and Février (2015) and Torgovitsky (2015) in a different context, as they consider support conditions on an excluded "instrumental variable" z. First consider

Condition 3 *For any $v^0, v^1 \in \mathcal{S}_t^v$, $\mathcal{S}_t^{x|v^0}$ and $\mathcal{S}_t^{x|v^1}$ have a common support point x^{01}.*

Relative to support condition (2), condition (3) allows the common support point (before x^0, now x^{01}) to potentially be different for each pair of (v^0, v^1). One can construct simple

[19] This presumes that the inverse function $F^{-1}_{y_{it}|x,v}$ exists, which should be the case because of our assumptions that 1) $f_t(x, \omega)$ is strictly increasing in ω, 2) $g_t\left(\{\omega_{i\tau}\}_{\tau=t-M}^{t-1}, \xi_{it}\right)$ is strictly monotone in ξ_{it}, and 3) $\xi_{it} \sim U(0, 1)$.

examples of $\mathcal{S}_t^{xv}$ where condition (3) is satisfied but not condition (2). Given the common support point, an argument similar to the above can order any $f_t^{-1}(x^A, y^A)$ and $f_t^{-1}(x^B, y^B)$.

Next, observe that to order any pair $f_t^{-1}(x^A, y^A)$ and $f_t^{-1}(x^B, y^B)$, we do not necessarily need *every* pair (v^A, v^B) (consistent with (x^A, y^A) and (x^B, y^B) respectively) to have a common support point - we only need *some* (v^A, v^B) to have a common support point. Specifically, consider the weaker condition

Condition 4 *For any (x^A, y^A) and $(x^B, y^B) \in \mathcal{S}_t^{xy}$, there exists $v^A \in \mathcal{S}_t^{v|x^A,y^A}$, $v^B \in \mathcal{S}_t^{v|x^B,y^B}$ and some x^{AB} such that $(x^{AB}, v^A), (x^{AB}, v^B) \in \mathcal{S}_t^{xv}$*

This support condition also allows us to order any $f_t^{-1}(x^A, y^A)$ and $f_t^{-1}(x^B, y^B)$ using the "pair-specific" common support points x^{AB}. Note that support condition (4) is dependent on S^{xvy}. This differs from conditions (2), and (3) (and Imbens and Newey) which only put restrictions on S^{xv}. However, condition (4) is implied by the prior conditions and hence weaker.[20]

But we can also do indirect orderings - i.e., order (x^A, y^A) and (x^B, y^B) "through" other points. For example if we can find a point (x^C, y^C) that is iso-omegic to (x^A, y^A) and a point (x^D, y^D) that is iso-omegic to point (x^B, y^B), then instead of comparing (x^A, y^A) to (x^B, y^B), we can compare (x^C, y^C) to (x^D, y^D). To consider this, define the following set of points:

$$\mathcal{W}(x^A, y^A) = \left\{(x, y) : \exists v^0 \text{ s.t. } (x^A, y^A, v^0) \in \mathcal{S}_t^{xyv}, x \in \mathcal{S}_t^{x|v^0}, y = F^{-1}_{y_{it}|v^0,x}\left(F_{y_{it}|v^0,x^A}\left(y^A\right)\right)\right\}.$$

$\mathcal{W}(x^A, y^A)$ is a set of points that is iso-omegic to (x^A, y^A). These points are found by 1) considering all the v^0 that are on the support that are consistent with (x^A, y^A), 2) finding the implied ξ at those values using the identified cumulative distribution $F_{y_{it}|v^0,x^A}\left(y^A\right)$, 3) finding other x's that are on the support that are consistent with v^0, i.e., $x \in \mathcal{S}_t^{x|v^0}$, and 4) using $F^{-1}_{y_{it}|v^0,x}\left(F_{y_{it}|v^0,x^A}\left(y^A\right)\right)$ to compute the y implied by v^0 from step 1), the implied ξ from step 2), and each of those other x's from step 3).

Note that $\mathcal{W}(x^A, y^A)$ does not necesarily contain all the points in $\mathcal{S}_t^{xy}$ that are iso-omegic to (x^A, y^A) - it only contains those we can "find" with v^0's that are observed with (x^A, y^A) and x's associated with those v^0's. How much of the set of iso-omegic points $\mathcal{W}(x^A, y^A)$ contains will depend on the joint support. If the support $\mathcal{S}_t^{v|x^A,y^A}$ is very small, e.g., because the $\mathcal{S}_t^{v|x^A}$ is small, then $\mathcal{W}(x^A, y^A)$ may not capture many of the iso-omegic points.

We can potentially find more iso-omegic points by iteratively applying $\mathcal{W}$. To do this,

[20] Condition (4) is implied by Condition (3), because if (x^A, y^A) and $(x^B, y^B) \in \mathcal{S}_t^{xy}$, there must exist some v^A and v^B s.t. (x^A, y^A, v^A) and $(x^B, y^B, v^B) \in \mathcal{S}_t^{xvy}$, and Condition (3) assures these v^A and v^B have a common x support point.

extend the above operator to work on subsets rather than just points, i.e.,

$$\mathcal{W}(\mathcal{S})=\left\{\begin{array}{c}(x,y):\text{ for some }(x^A,y^A)\in\mathcal{S}\ \exists v^0\text{ s.t. }(x^A,y^A,v^0)\in\mathcal{S}_t^{xyv},x\in\mathcal{S}_t^{x|v^0},\\ y=F^{-1}_{y_{it}|v^0,x}\left(F_{y_{it}|v^0,x^A}\left(y^A\right)\right)\end{array}\right\}$$

where $\mathcal{S}\subseteq\mathcal{S}_t^{xy}$. Then, for example $\mathcal{W}^2(x^A,y^A)=\mathcal{W}(\mathcal{W}(x^A,y^A))$ can find new points that are iso-omegic to (x^A,y^A) (in addition to those in $\mathcal{W}(x^A,y^A)$) . These new points could not be directly linked to (x^A,y^A) through a v, but could be linked indirectly through points in $\mathcal{W}(x^A,y^A)$. One could also iteratively apply $\mathcal{W}$ some number N times, i.e., $\mathcal{W}^N(x^A,y^A)$. But even if this were done infinitely, it would not necessarily contain all points in $\mathcal{S}_t^{xy}$ that are iso-omegic to $\mathcal{W}(x^A,y^A)$ - again, it depends on the support of the data. But we can consider

Condition 5 *For any (x^A,y^A) and $(x^B,y^B)\in\mathcal{S}_t^{xy}$ there is a value x^0 that is in both the sets $\mathcal{W}^N(x^A,y^A)$ and $\mathcal{W}^N(x^B,y^B)$ (for some $N\in\mathbb{N}$).*

Support condition (5) further weakens condition (3) and is also sufficient to order any $f_t^{-1}(x^A,y^A)$ and $f_t^{-1}(x^B,y^B)$. Intuitively, condition (5) implies that for any (x^A,y^A) and (x^B,y^B), we can find iso-omegic sets that have a common support point x^0. Like above, we can then order $f_t^{-1}(x^A,y^A)$ and $f_t^{-1}(x^B,y^B)$ by comparing the y values corresponding to x^0 in those two sets. But this can be generalized as well. It is possible that even if support condition (5) does not hold, we can order $f_t^{-1}(x^A,y^A)$ and $f_t^{-1}(x^B,y^B)$ by finding some (x^C,y^C) for which condition (5) holds pairwise, e.g., $f_t^{-1}(x^A,y^A)<f_t^{-1}(x^C,y^C)$ and $f_t^{-1}(x^C,y^C)<f_t^{-1}(x^B,y^B)$. To utilize this logic, define a sequence of points $(x^0,y^0),...,(x^{J+1},y^{J+1})$ as an *omegically monotone sequence* if either $f_t^{-1}(x^0,y^0)\geq...\geq f_t^{-1}(x^{J+1},y^{J+1})$ or $f_t^{-1}(x^0,y^0)\leq...\leq f_t^{-1}(x^{J+1},y^{J+1})$ is true (for $J\geq 0$). Then consider:

Condition 6 *For any (x^A,y^A) and $(x^B,y^B)\in\mathcal{S}_t^{xy}$ there is an omegically monotone sequence $(x^0,y^0),...,(x^{J+1},y^{J+1})$ in $\mathcal{S}_t^{xy}$ such that each consecutive pair in the sequence, denoted by $((x^j,y^j),(x^{j+1},y^{j+1}))$, is such that $\mathcal{W}^N(x^j,y^j)$ and $\mathcal{W}^N(x^{j+1},y^{j+1})$ contain a common value x^{Cj}, for $j=0,...,J$ and $(x^0,y^0)=(x^A,y^A),(x^{J+1},y^{J+1})=(x^B,y^B)$.*

Support condition (6) is weaker than condition (5) since condition (5) implies that condition (6) holds for all (x^A,y^A) and (x^B,y^B) with $J=0$, i.e., no intermediate points are necessary. Condition (6) may be helpful in relaxing condition (5), especially when (x^A,y^A) and (x^B,y^B) are relatively distant. In this case it might be hard for $\mathcal{W}^N(x^A,y^A)$ and $\mathcal{W}^N(x^B,y^B)$ to overlap, i.e., condition (5) to hold, but condition (6) can still hold as long as there is a "chain" of overlapping points that can connect (x^A,y^A) to (x^B,y^B) indirectly. Lastly, note the need for the sequence in support condition (6) to be omegically monotone - if, e.g., $f_t^{-1}(x^A,y^A)\geq f_t^{-1}(x^1,y^1)$ and $f_t^{-1}(x^1,y^1)\leq f_t^{-1}(x^B,y^B)$, then (x^1,y^1) is not helpful at ordering (x^A,y^A) and (x^B,y^B).

Theorem 4 *Under the assumptions of our model and support condition (6), $f_t(x,\omega)$ is identified for all $t > M$.*

Proof. See Appendix B. ■

Theorem (4) clearly also implies that $f_t(x,\omega)$ is identified under the stronger support conditions (3), (4), and (5). This is useful since the former conditions, while stronger, may be more easily interpretable.

We can also consider other types of support conditions on $\mathcal{S}_t^{xvy}$ that are sufficient for identification. Support conditions (2) and (3) are interesting because they only place restrictions on $\mathcal{S}_t^{xv}$, and assume nothing about $\mathcal{S}_t^{y|xv}$. One can also approach the problem from the "opposite" direction, i.e., starting with more restrictions on $\mathcal{S}_t^{y|xv}$ and less restrictions on $\mathcal{S}_t^{xv}$. While this approach does not generalize Imbens and Newey's support condition, we feel they are also interesting. An additional assumption on the primitives of our model that helps do this is the following:

Assumption 6 *(i) The conditional distribution of the unobservable ω_{it}, i.e., $p_t\left(\omega_{it} \mid \{\omega_{i\tau}\}_{\tau=t-M}^{t-1}\right)$, has support that does not depend on $\{\omega_{i\tau}\}_{\tau=t-M}^{t-1}$. (ii) The boundary of $\mathcal{S}_t^{xy}$ has probability measure zero. (iii) For any $x_{it} \in Int(\mathcal{S}_t^{x})$ (the interior of set $\mathcal{S}$) there exists a v_{it} such that $(x_{it}, v_{it}) \in Int(\mathcal{S}_t^{xv})$.*

Part (ii) and (iii) of Assumption (6) are regularity conditions that make the proof of the following theorem easier as we can work with open sets. Part (i) is more substantive. With our normalization, it implies that ω_{it} has support $(0,1)$ regardless of prior ω_{it}'s (though the distribution over that support will generally depend on prior ω_{it}'s). What this assumption does is restrict $\mathcal{S}_t^{y|xv}$ to not depend on v. Regarding the discussion above, this implies that if $(x^A, y^A) \in \mathcal{S}_t^{xy}$ and $(x^A, v^{new}) \in \mathcal{S}_t^{xv}$ for some v^{new}, then it must be the case that $(x^A, v^{new}, y^A) \in \mathcal{S}_t^{xvy}$. This eases restrictions on $\mathcal{S}_t^{xv}$ required to order any two points (x^A, y^A) and (x^B, y^B), and means we can obtain identification with only a convex support condition, i.e.

Theorem 5 *Under the assumptions of our model and Assumption 6, then if $\mathcal{S}_t^{xv}$ is convex, $f_t(x,\omega)$ is identified for all $t > M$.*

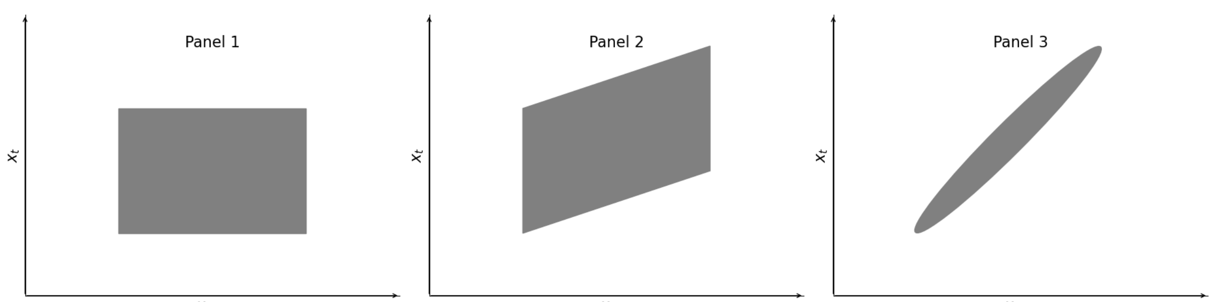
Proof. See Appendix B ■

Theorem (5) shows that under this relatively strong condition on $p_t\left(\omega_{it} \mid \{\omega_{i\tau}\}_{\tau=t-M}^{t-1}\right)$, convexity (together with a regularity condition) is sufficient to identify f_t. Intuitively, Assumption (6) and convexity of $\mathcal{S}_t^{xv}$ allow one to order any (x^A, y^A) and (x^B, y^B) by moving in steps along a straight line from x^A to x^B. For small enough "step-size" along the line (depending on the size of the support of v) we can always find a common v and an iso-omegic point at the next step (using Assumption (6)), eventually arriving at a (x^B, y^{iso}) that is iso-omegic to (x^A, y^A). Then a comparison of y^{iso} to y^B orders the two relevant points.

$\mathcal{S}_t^{xv}$ being convex seems quite weak in relation to Imbens and Newey's support condition. It can hold even if the distribution of $v|x$ (or vice versa) has very small support at each x. For example, it holds if the marginal support of x is an interval $[\underline{x}, \bar{x}]$ and the support of $v|x$ is just $[x-\epsilon, x+\epsilon]$ for any small ϵ. And intuitively, at minimum we clearly need some independent variation in v and x to have any hope for identifying $f_t(x,\omega)$. But again, this is not strictly weaker than Imbens and Newey's condition because it additionally requires Assumption (6). We lastly note that under Assumption (6), convexity is sufficient but not necessary for identification. One needs only to be able to move on some path between any x^A and x^B such that for small enough steps, the support of $v|x$ is large enough to find a sequence of iso-omegic points. Convexity assures that this can be done very simply, i.e., with straight line.

Figure 1 illustrates some of the various support conditions derived above. To keep the graphs in two dimensions, we only consider the x_{t-1} component of the control variable. In a production function context, the graphs could correspond to the joint distribution of current capital stock and lagged capital stock. Panel 1 illustrates the rectangular support required by Imbens and Newey. Panel 2 demonstrates a support that does not satisfy the Imbens and Newey support condition, but is consistent with identification based on Theorem (3). Note that the top of the left boundary of the support is higher than the bottom of right boundary of the support. This ensures there is a x_t for which x_{t-1} (v) has full support. Lastly, Panel 3 shows a convex support consistent with identification based on Theorem (5).[21] .

Figure 1: Graphical Illustration of Support Conditions



[21]Note that what Theorem (4) allows is harder to visualize since, e.g. support condition (6) depends in a complicated way on the distibution of y. However, this condition does allow for "holes" in the support. Also, as noted above, convexity is only a sufficient for identification based on Theorem (5) (with Assumption (6)). One could derive conditions under which holes in the support are permitted by adding more tedious conditions and chaining arguments similar to those used for Theorem (4)

4.3 Partial Identification Revisited

In section (4.1) we showed one type of partial identification results for our model - those related to identifying derivatives of f_t at certain points. With the results in the prior section, we can generate some additional, broader, partial identification results. Support conditions (4), (5), and (6) are stated as holding for any (x^A, y^A) and $(x^B, y^B) \in \mathcal{S}_t^{xy}$. We now consider the situation where (4), (5), or (6) hold over some subset $\widetilde{\mathcal{S}}_t^{xy} \subseteq \mathcal{S}_t^{xy}$. We can show

Theorem 6 *Under the assumptions of our model, if support condition (6) holds for all* (x^A, y^A) *and* (x^B, y^B) *in some subset* $\widetilde{\mathcal{S}}_t^{xy} \subseteq \mathcal{S}_t^{xy}$*, then* $f_t^{-1}(x, y)$ *is identified for all* $t > M$ *on* $\widetilde{\mathcal{S}}_t^{xy}$*, under the normalization that the distribution of* ω_{it} *implied by the distribution of* (x_{it}, y_{it}) *on* $\widetilde{\mathcal{S}}_t^{xy}$ *is* $U(0, 1)$*.*

Proof. Identical to the proof of Theorem 4, with the normalization and identification only on the set $\widetilde{\mathcal{S}}_t^{xy}$. ■

The intuition behind Theorem (6) is that if supoort condition (6) holds on some subset $\widetilde{\mathcal{S}}_t^{xy} \subseteq \mathcal{S}_t^{xy}$, then all points (x^A, y^A) and (x^B, y^B) in that subset can be ordered - in exactly the same way as the prior section. And again analogous to the above, given the ordering on this set, ω_{it}'s are identified up to a normalization on this set. The caveat is that this alternative normalization does not permit one to compare the identified f_t on this set $\widetilde{\mathcal{S}}_t^{xy}$ to f_t at other places on $\mathcal{S}_t^{xy}$ (e.g., perhaps some other partially identified set).

But even with this caveat, Theorem (6) seems economically important because it allows one to identify answers to interesting counterfactual questions within the subset $\widetilde{\mathcal{S}}_t^{xy}$. For any point in the set, (x^A, y^A), it allows us to identify counterfactual outcomes if x^A were changed to x^{Alt}, holding ω^A constant, as long as the resulting $(x^{Alt}, y^{Alt}) \in \widetilde{\mathcal{S}}_t^{xy}$. So, for example, in a production function context one could consider the classic counterfactual reallocation question, i.e., what happens if inputs x are reallocated across firms in alternative ways (holding ω's constant), as long as those reallocations stay within the set. The restriction that the reallocations stay within $\widetilde{\mathcal{S}}_t^{xy}$ is not innocuous, but it is not surprising that one cannot identify outcomes outside of the identified set (and one might be able to put one-sided bounds on outcomes from reallocations that end up outside the set). In any case, this result shows that the identification conditions above have some "localness" to them.

5 Relaxing Timing and Information Set Assumptions

5.1 Nonidentification without Additional Restrictions

Our control function approach to identifying f_t relies crucially on the timing and information set Assumption (1). Because x_{it} is chosen at $t-1$, prior to the realization of ω_{it}, x_{it}

is independent of ξ_{it}, and thus also independent of ω_{it} given the control variables $v_{it-1} = \left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$. These are strong assumptions, though they are frequently used in the empirical literature. For example, much of the production function literature following Olley and Pakes (1996) makes this assumption w.r.t. capital input. Some papers also make this assumptions when x_{it} includes labor input choice, e.g., Gandhi et al. (2020).

Ideally, one might like to relax this timing/information set assumption, i.e. allow x_{it} to be chosen as a function of ω_{it}, in essence making x_{it} "more endogenous". Unfortunately, a non-identification result from Ackerberg et al. (2020) implies that this is not possible without further assumptions. They illustrate that, under this weaker timing and information set assumption, a linear version of the model is not point identified. This implies that a nonparametric model such as ours would also not be identified without additional assumptions. While fully investigating possible additional assumptions that could generate identification in the context of this weaker timing/information set assumption is beyond the scope of this paper, we briefly discuss some possible alternatives.[22] In the parametric linear case, Ackerberg et al. (2020) augment the above model with a sign restriction (on θ_2 and θ_6) to generate identification. As discussed earlier, Hu and Shum (2012) and Hu and Shum (2013) use deconvolution techniques to identify a nonparametric model similar to ours. These deconvolution approaches can accommodate x_{it} depending on ω_{it}, but, unlike our approach, they cannot identify f_t unless one assumes the Markov process is shorter than the number of observed lags (i.e., they require $M < T + 1$). And building on Gandhi et al. (2020) in a production function context, Navarro and Rivers (2018) use a first-order condition approach based on price-taking firms (plus a multiplicative separability assumption) to accommodate a material input that depends on ω_{it}. Similarly, Pan (2022) uses a first-order condition approach to add a variable material input to the current approach under an additional homogeneity assumption. In both these cases, the static first-order condition is only required to hold w.r.t. the variable input, not capital and labor.

5.2 Identification with Additional Restrictions

Another interesting approach concerns a situation where one is only willing to assume some of the elements of x_{it} satisfy the timing/information set Assumption (1), but where one observes traditional "excluded instruments" z_{it} for the elements of x_{it} that are chosen as a function of ω_{it}. These instrumental variables z_{it} determine those latter elements of x_{it}, but satisfy traditional IV exclusion restrictions. In this case, one could think of combining the timing/information set assumptions described above with traditional IV restrictions for identification. This has been

[22] In contrast, it is straightforward to allow x_{it} to be chosen with *less* information, e.g., x_{it} is chosen by the agent at time $t-2$, i.e. where $x_{it} = h_t(I_{it-2})$. In that case, one could use $\left(\{y_{i\tau}\}_{\tau=t-M-1}^{t-2}, \{x_{i\tau}\}_{\tau=t-M-1}^{t-2}\right)$ as control variables (note that using $t-1$ lags as control variables would also work, but this would likely preserve more variation).

done in the parametric case by De Roux et al. (2021) in the production function context. They assume the capital input is chosen at $t-1$, but observe external "input price shifter" instruments for the inputs assumed to be chosen at t. In the demand context, Dearing (2025) considers a similar combination of assumptions in applying timing and information set assumptions to identify a BLP style demand model with multidimensional (but separable) unobservables.

To extend this to our nonparametric, nonseparable context, denote the two types of inputs as x_{it}^F and x_{it}^V. The object of interest is now $y_{it} = f_t(x_{it}^F, x_{it}^V, \omega_{it})$. As in our base model, assume x_{it}^F is chosen as a function of only $\mathcal{I}_{it-1}$ (note that with the additional variable z_{it}, $\mathcal{I}_{it-1}$ is now $\left\{\{y_{i\tau}\}_{\tau=1}^{t-1}, \{x_{i\tau}\}_{\tau=1}^{t-1}, \{\omega_{i\tau}\}_{\tau=1}^{t-1}, \{\eta_{i\tau}\}_{\tau=1}^{t-1}, , \{z_{i\tau}\}_{\tau=1}^{t-1}\right\}$). In contrast, suppose that x_{it}^V is chosen with the additional knowledge of ω_{it}, η_{it}, and z_{it} - according to

$$\begin{aligned} x_{it}^V &= \psi_t\left(x_{it}^F, \omega_{it}, z_{it}, \eta_{it}\right) \\ &= \psi_t\left(x_{it}^F, g_t\left(\omega_{it-1}, \xi_{it}\right), z_{it}, \eta_{it}\right). \end{aligned} \tag{5}$$

[23] Note how this model of x_{it}^V corresponds to a variable, non-dynamic (see Ackerberg et al. (2007)) input in a production function context. In that case, x_{it}^V will generally depend on x_{it}^F and ω_{it} (since they determine the marginal product of x_{it}^V in $f_t(x_{it}^F, x_{it}^V, \omega_{it})$). We require x_{it}^V to depend the observed instruments z_{it} - presumably these are variables related to the price the firm pays for inputs x_{it}^V. We also allow x_{it}^V to depend on unobservables η_{it} - also perhaps related to input markets.

In our base model, our assumptions implied that (x_{it}, v_{it-1}) was jointly independent of ξ_{it}. With x_{it}^V chosen at t, this no longer holds, i.e. we only have that (x_{it}^F, v_{it-1}) is jointly independent of ξ_{it}. Therefore we require the additional assumption that $(x_{it}^F, v_{it-1}, z_{it})$ is jointly independent of ξ_{it}. This is the analogue of a traditional IV restriction w.r.t. z_{it}, though there are a couple of differences. First, note that we require full joint independence, i.e. nothing in the joint distribution between z_{it} and (x_{it}^F, v_{it-1}) can be related to ξ_{it}. Second, note that because of our control variable v_{it-1}, we only need independence of z_{it} from the innovation term ξ_{it}. In other words, z_{it} can be correlated with ω_{it} through ω_{it-1}.[24] Appendix (C) shows that given this setup, we can use the framework of Chernozhukov and Hansen (2005) to identify f_t.[25]

[23] Again in this section, for exposition convenience, we assume $M = 1$, but our approach and arguments extend naturally to the cases where $M > 1$.

[24] As well known in the parametric production function literature, if z_{it} is literally the price of inputs x_{it}^V, then use of z_{it} as instruments will require firms to be price takers in input markets. This wouldn't be required if z_{it} are measured input supply shocks. In the former case, given we only need independence from ξ_{it}, one could speculate that using lagged z_{it} could alleviate requiring this assumption, though we have not fully investigated this possibility.

[25] An alternative to using Chernozhukov and Hansen (2005) would be to re-apply Imbens and Newey (2009) and create another control variable to address the "endogeneity" of x_{it}^V. However, this would require additional restrictions on the η_{it} entering equation (5), e.g. that it is scalar, and additional independence conditions.

Intuitively, this works by again considering the following reduced form expression

$$\begin{aligned} y_{it} &= f_t(x^F_{it}, x^V_{it}, \omega_{it}) \\ &= f_t(x^F_{it}, x^V_{it}, g_t(f^{-1}_{t-1}(x_{it-1}, y_{it-1}), \xi_{it})) \\ &= f_t(x^F_{it}, x^V_{it}, g_t(f^{-1}_{t-1}(v_{it-1}), \xi_{it})) \\ &= \bar{f}_t(x^F_{it}, x^V_{it}, v_{it-1}, \xi_{it}). \end{aligned} \tag{6}$$

and, given independence of $(x^F_{it}, v_{it-1}, z_{it})$ and ξ_{it}, applying the results of Chernozhukov and Hansen (2005) to identify $\bar{f}_t$. Once $\bar{f}_t$ is identified, we can rely on one of our support conditions (Conditions (2), (3), (4), (5), and (6)) to identify f_t. An important caveat is that using this nonparametric IV approach to identify $\bar{f}_t$ requires completeness conditions on the instruments z_{it} that can be hard to interpret in practice.

6 Identification for $t \leq M$

One limitation of the control function approach is that it cannot identify f_1 with a first-order Markov assumption (since for $t = 1$ there is no observed (x_0, y_0) to use for the control function). More generally, when ω_{it} follows an Mth order Markov process, the control function approach only identifies f_t for $t > M$. We now illustrate an alternative approach can be used to identify f_t for $t \leq M$. Like section (4.1), we show that under local regularity conditions, this approach identifies aspects of f_t - in particular, derivatives of f_t w.r.t. x_{it} at certain points.[26] One caveat is that these identification results rely on ω_{it} in fact being serially correlated. While this does not seem like a strong assumption, it highlights that how this identification result is somewhat "indirect", and that the precision of estimates based on it may be sensitive to the level of serial correlation in the model.

For exposition convenience, in this section we assume $M = 1$, i.e. ω_{it} follows a first-order Markov process, but our approach can be straightforwardly extended to the cases where $M > 1$. The intuition of the approach can be illustrated in a simple linear model. Suppose $T = 2$ and that

$$y_{i1} = \theta_1 + \theta_2 x_{i1} + \omega_{i1} \tag{7}$$

$$y_{i2} = \theta_3 + \theta_4 x_{i2} + \omega_{i2} \tag{8}$$

and

$$\omega_{i2} = \rho\omega_{i1} + \xi_{i2} \tag{9}$$

[26] To identify ω_{it} and thus the entire f_t function following this identification strategy, we need similar support conditions as discussed in section (4.2).

Our goal is to identify θ_2 (θ_4 can be identified with the control function method). To do this, substitute the inverted (7) into (9), and this into (8) to get

$$y_{i2} = (\theta_3 - \rho\theta_1) + \theta_4 x_{i2} + \rho y_{i1} + \rho\theta_2 x_{i1} + \xi_{i2} \tag{10}$$

Our timing, information set, and first-order Markov assumptions imply that ξ_{i2} is independent of (x_{i2}, y_{i1}, x_{i1}). So as long as $\rho \neq 0$, we can simply run, e.g., OLS on (10) and recover θ_2 by dividing the coefficient on x_{i1} by that on y_{i1}. Note that the identification here comes from comparing the relative effect of y_{i1} and x_{i1}, through the implied ω_{i1}, on y_{i2}.

We now extend this argument to our nonparametric model. What we will be able to identify is $\frac{\partial f_1}{\partial x_{i1}}$ evaluated at the point x_{i1} and the implied omega corresponding to x_{i1} and y_{i1}. Like in section (4.1), this is a bit of hard result to interpret, as we do not know the actual numeric value of this implied ω. But since x_{i1} and y_{i1} are observed for each data point, it does allow us to identify $\frac{\partial f_1}{\partial x_{i1}}$ for observations in the data evaluated at their existing x_{i1} and ω_{it}. We utilize the following conditions.

Assumption 7 *For some $\epsilon > 0$, the conditional distribution of v_{it-1} given $x_{it} = x_t^0$ has positive support on all v satisfying $\left\| v - v_{t-1}^0 \right\| < \epsilon$.*

Assumption 8 *$\frac{\partial g_t(\omega_{t-1}^0, \xi_t^0)}{\partial \omega_{it-1}}$ is nonzero at $\omega_{t-1}^0 = f_{t-1}^{-1}(v_{t-1}^0) = f_{t-1}^{-1}(x_{t-1}^0, y_{t-1}^0)$ for some $\xi_t^0 \in (0,1)$.*

The local support Assumption (7) is needed to identify derivatives and is very similar to Assumption (1) - the conditioning is reversed since with this strategy we are leveraging variation in y_{it} in response to v_{it-1} *conditional on* x_{it} (whereas the control function approach looks at the reverse conditioning). Assumption (8) is the requirement discussed above that there needs to be some serial correlation in ω_{it} - the analogue of requiring $\rho > 0$ in the simple linear model (9).

Theorem 7 *If the model satisfies Assumptions (7) and (8) at some $\left(x_t^0, v_{t-1}^0\right)$, then $\frac{\partial f_{t-1}(x_{it-1}, \omega_{it-1})}{\partial x_{it-1}}$ is identified at the points $x_{it-1} = x_{t-1}^0$ and $\omega_{it-1} = f_{t-1}^{-1}(v_{t-1}^0) = f_{t-1}^{-1}(x_{t-1}^0, y_{t-1}^0)$.*

Proof. See Appendix B. ■

Theorem (7) uses the fact that the observed ξ_t^0th quantile of y_{it} conditional on $(x_{it}, x_{it-1}, y_{it-1}) = \left(x_t^0, x_{t-1}^0, y_{t-1}^0\right)$ can be written as

$$q_{y_{it}|x_{it}, v_{it-1}}(\xi_t^0 | x_t^0, x_{t-1}^0, y_{t-1}^0) = f_t(x_t^0, g(f_{t-1}^{-1}(x_{t-1}^0, y_{t-1}^0), \xi_t^0)) \tag{11}$$

As a result, we can use the implicit function theorem to identify derivatives of f_{t-1} by taking the ratios of the derivatives of the conditional quantile of y_{it} with respect to x_{it-1} and y_{it-1}.

Note that the variation used in this identification result is quite distinct from that used in the control function approach of Theorem (2). The latter uses variation in y_{it} generated by x_{it} to identify derivatives of f_t, while Theorem (7) uses variation in y_{it} generated by x_{it-1} and y_{it-1} to identify derivatives of f_{t-1}. This also indicates overidentification in this model, e.g., with $T = 3$ (and $M = 1$), aspects of f_2 could potentially be estimated in two distinct ways - 1) using the control function approach with data in periods $t = 1, 2$, and 2) using the alternative approach with data in periods $t = 2, 3$. Also note that crucial to this identification strategy is that the model implies that x_{it-1} and y_{it-1} affect y_{it} through a single index - Hahn et al. (2021) examine other implications of the single index structure of these types of models.

7 Estimation

7.1 A Sieve Maximum Likelihood Estimator

Based on our nonparametric identification results, we propose a relatively easy-to-implement nonparametric estimation procedure. In particular, we bypass estimating equation (3), since utilizing this directly would rely on the Imbens and Newey common support condition (in section (8.3) we show evidence that this common support condition likely fails in our production function application). Instead, our estimation procedure follows equation (4),[27] the basis for our relaxed support condition theoretical results. Inverting equation (4), we have

$$\begin{aligned}\xi_{it} &= g_t^{-1}(\omega_{it-1}, \omega_{it}) \\ &= g_t^{-1}(f_{t-1}^{-1}(x_{it-1}, y_{it-1}), f_t^{-1}(x_{it}, y_{it})). \qquad (12)\end{aligned}$$

where we will use the normalizations that $f_t^{-1}(0, y_{it}) = y_{it}$ and that ξ_{it} has a standard normal marginal distribution. Letting $z_i = (x_{it}, y_{it})_{t=1}^T$ and treating f_t^{-1} and g_t^{-1} as parameters (instead

[27]For exposition convenience, we assume $M = 1$, i.e. ω_{it} follows a first-order Markov process.

of f_t and g_t, see discussion below), we can construct a "partial" log-likelihood for observation i:

$$
\begin{aligned}
l(f_t^{-1}, g_t^{-1}, z_i) &= \sum_{t=2}^{T} log[P(y_{it}|x_{it}, x_{it-1}, y_{it-1})] \\
&= \sum_{t=2}^{T} log[P(\xi_{it}|x_{it}, x_{it-1}, y_{it-1})\frac{\partial \xi_{it}}{\partial y_{it}}] \\
&= \sum_{t=2}^{T} log[\phi(g_t^{-1}(f_{t-1}^{-1}(x_{it-1}, y_{it-1}), f_t^{-1}(x_{it}, y_{it})))\frac{\partial g_t^{-1}(\omega_{it-1}, \omega_{it})}{\partial \omega_{it}}\frac{\partial f_t^{-1}(x_{it}, y_{it})}{\partial y_{it}}], \qquad (13)
\end{aligned}
$$

where ϕ is the pdf of a standard normal distribution. This is a partial likelihood because it only considers the conditional densities of the y_{it}. More specifically, in our model x_{it} are endogenous variables, determined as a potential function of past unobservables (both at $t-1$ and prior). A full likelihood would need to consider the joint likelihood of all the endogenous variables (over time). Our partial likelihood approach does not require fully specifying a model of x_{it}, which is an advantage since in our production function context these are inputs that may be chosen as part of complex dynamic, optimization problems that depend on many other factors and parameters (e.g., Olley and Pakes (1996) and much of the related literature also emphasize this advantage (though in those cases in the context of GMM estimation)). Our partial maximum likelihood estimator is consistent, and satisfies the property of "dynamic completeness" discussed by Wooldridge (2010), allowing us to treat the partial likelihood as a full likelihood when conducting inference.

Our estimator thus takes the form:

$$
\hat{\theta}_n = (\hat{f}_t^{-1}, \hat{g}_t^{-1}) = \underset{(f_t^{-1}, g_t^{-1}) \in (\mathcal{F}_N^{-1}, \mathcal{G}_N^{-1})}{\arg\max} \sum_{i=1}^{N} l(f_t^{-1}, g_t^{-1}, z_i). \qquad (14)
$$

where we use sieves to approximate the inverse functions f_t^{-1} and g_t^{-1} directly. An alternative would be to approximate f_t and g_t (e.g. with sieves) - then inverting them to calculate the partial likelihood. We prefer estimating f_t^{-1} and g_t^{-1} directly because this avoids having to do these inversions for each value of the parameters searched over in our search algorithm. Note, e.g. that given estimates of f_t^{-1}, one can easily obtain the derivatives of f_t using the implicit function theorem, i.e., $\frac{\partial f_t(x,\omega)}{\partial x} = -\frac{\partial f_t^{-1}(x,y)}{\partial x} / \frac{\partial f_t^{-1}(x,y)}{\partial y}$. We will derive the rate of convergence of our estimator in the supplemental material.

8 Application to Production Function Estimation

8.1 Relationship with Literature

We apply these identification results to the estimation of production functions. Our main goal is to examine the implications of relaxing the typical assumption of Hicks neutral productivity shocks. In other words, we relax the assumption of many production function empirical models that (log) output y_i is linear in the unobserved, firm-specific, productivity shock ω_{it}, e.g.

$$y_{it} = f(x_{it}) + \omega_{it}$$

where x_{it} are observed inputs like capital and labor.

We are not the first to do this. A large set of recent papers, including Balat et al. (2016), Fox et al. (2017), Kasahara et al. (2017), Doraszelski and Jaumandreu (2018), Raval (2019), Zhang (2019), Demirer (2020), Raval (2020), Oberfield and Raval (2021), and Li and Sasaki (2024), also relax this restriction. However, we do it in a quite different way. The above papers augment the above model with additional unobservables that enter f, but under specific functional form assumptions. In other words, they consider models that look like

$$y_{it} = f(x_{it}, \omega_{it}^2; \theta) + \omega_{it}^1$$

where ω_{it}^2 represents one (or more) additional unobservable technology shocks and where there are functional form restrictions on f, i.e., f is known up to the finite dimensional parameter vector θ. For example, in Doraszelski and Jaumandreu (2018) ω_{it}^2 is a scalar, "labor-augmenting" shock that directly multiplies the labor input in levels in the context of a CES production function. As another example, in Oberfield and Raval (2021) ω_{it}^2 is a three-vector of factor augmenting shocks in a nested CES production function (with $\omega_{it}^1 \equiv 0$).

Our approach, based on our identification results, makes a very different set of restrictions. We keep the assumption of a scalar ω_{it}, but we are completely flexible with regards to f except for our strict monotonicity restriction, i.e.,

$$y_{it} = f(x_{it}, \omega_{it}).$$

In other words, while we keep the scalar unobservable assumption, we allow ω_{it} to interact in very general ways with the various inputs in the vector x_{it}.[28]

[28] Given the flexibility in which ω_{it} enters the model, it might capture effects that have a somewhat different economic interpretation than standard technology shocks. For example, suppose an input x_{it} is measured in monetary units, that firms face different prices of this input, and that these prices are unobserved to the researcher. Then since our nonparametric model allows ω_{it} to potentially multiply x_{it} in f, ω_{it} could represent

These two approaches to relaxing Hicks neutrality are clearly non-nested - the models in the aforementioned papers make more functional form assumptions, while our model makes a stronger assumption regarding the dimension of the unobservables. We don't see either as being an unambiguously preferable assumption a-priori, and we in no way intend to claim that our model is preferable to the existing ones. But we feel our model provides added value in that it relies on *different* assumptions than the existing ones. In some empirical situations the assumptions of one of the two types of models may be more credible. But perhaps it is more likely that a researcher has no a-priori reason to choose one type of model over the other. In this case it seems reasonable to consider both types models - to the extent ones conclusions are similar across the approaches would be reassuring, i.e. we believe *these approaches/models are complementary to each other*. Ideally one would of course prefer to relax both assumptions, i.e., neither make the functional form restrictions on f (as in the other approaches) nor the scalar dimensionality restriction on ω_{it} (as in our approach). But this would be challenging - as illustrated by the fact that in a simpler model $y = f(x, \epsilon)$ where ϵ is independent of x, one can explain any observed joint distribution $p(y, x)$ with a model with a scalar unobservable ϵ. In other words, models that are both flexible in terms of functional form (like our approach) and flexible in terms of number of unobservables (like the other approach) are going to be challenging to point identify.[29]

Given this complementarity, it seems interesting to see what our approach finds regarding evidence of Hicksian non-neutrality and compare them to the results in the above work. We focus on various aspects of this Hicksian non-neutrality - its implication on production function elasticities, heterogeneity in these elasticities, and the bias (in terms of capital vs labor) of technological change. Interestingly, we find many patterns that are similar to the aforementioned work, which is supportive of their conclusions.

8.2 Data and Model

We use the same Chilean and Colombian data sets as do Levinsohn and Petrin (2003), Gandhi et al. (2020), and many others, and focus on the three largest industries in both countries. The Chilean data set comes from the census of Chilean manufacturing plants conducted by Chile's Instituto Nacional de Estadística.. It covers all firms from 1979 to 1996 with more than 10 employees. The Colombian data set comes from the Colombian manufacturing census, covering all manufacturing plants with more than 10 employees from 1981 to 1991.[30]

this unobserved price (to the extent that it satisfies the assumptions of our model, i.e. a scalar unobservable following a Markov process). Here, "better technology" is a firm somehow getting a lower price for the input.

[29] One could take a partial identification approach to the situation, but we leave this for future work, and think that various point identification results are still informative and useful.

[30] We construct our variables following Gandhi et al. (2020). Specifically, output is measured as deflated value-added. Labor input is measured as a weighted sum of blue- and white-collar workers, with blue-collar

The empirical work in Levinsohn and Petrin (2003) assumes a Cobb-Douglas production function and a Hicks neutral productivity shock, while Gandhi et al. (2020) identify a non-parametric production function, though also with a Hicks neutral productivity shock. Again, our non-Hicks neutral model relaxes these assumptions while controlling for the endogeneity of input choices using the type of timing and information set assumptions that are common in this literature. Note that to utilize our nonparametric framework, we follow Gandhi et al. (2020) and assume that l_{it} is chosen by firms before ω_{it} is realized. The hope is that labor market frictions (e.g. unions, other government policy, training) make this assumption reasonable. This labor timing assumption is stronger than that of Levinsohn and Petrin (2003), who allow l_{it} to be chosen as a function of ω_{it} (all three papers assume that k_{it} is determined before ω_{it} is realized). On the other hand, we are more flexible in regards to other aspects of the labor choice than are Levinsohn and Petrin - for example, their setup rules out the possibility of firms facing unobserved, firm-specific, serially correlated labor price shocks, while we allow such shocks (η_{it} in our general model). Note that we only include capital and labor as inputs in our production function. This is similar to the model in Ackerberg et al. (2015), described by Gandhi et al. (2020) as a "structural" value added production function. In essence, this assumes that if there are variable intermediate inputs (e.g. materials, electricity, fuel), they enter the production function in a Leontief fashion (and hence can essentially be "ignored" in estimation). The reason we do this is because, as detailed in Section 5, our model cannot be used to identify a gross output production function with fully variable inputs (i.e. inputs chosen as a function of ω_{it}). Pan (2022) extends the results here to a gross output production function based on the first-order condition approach of Gandhi et al. (2020).

We estimate our nonparametric nonseparable production function using the sieve MLE approach described in section (7). Specifically, we choose to use polynomials[31] to approximate both the (inverse) production function and the (inverse) Markov process.[32] For inference about some functionals of interest, such as average output elasticities, one can follow Newey (1994) and Ackerberg et al. (2012) and calculate standard errors under the presumption that our polynomial specification constitutes a parametric model. They show that such calculations

workers being weighted by the ratio of the average blue-collar wage to the average white-collar wage. Capital is constructed using the perpetual inventory method, where investment in new capital is combined with deflated capital from period $t-1$ to determine capital in period t.

[31]For both f_t^{-1} and g_t^{-1}, we include 2nd order terms in each of their individual arguments, plus all 3rd order interaction terms between arguments. We also allow the (inverse) production function to depend on a cubic function of time t, but restrict g to be constant over time. This is because much of the related literature allows time varying f but not g.

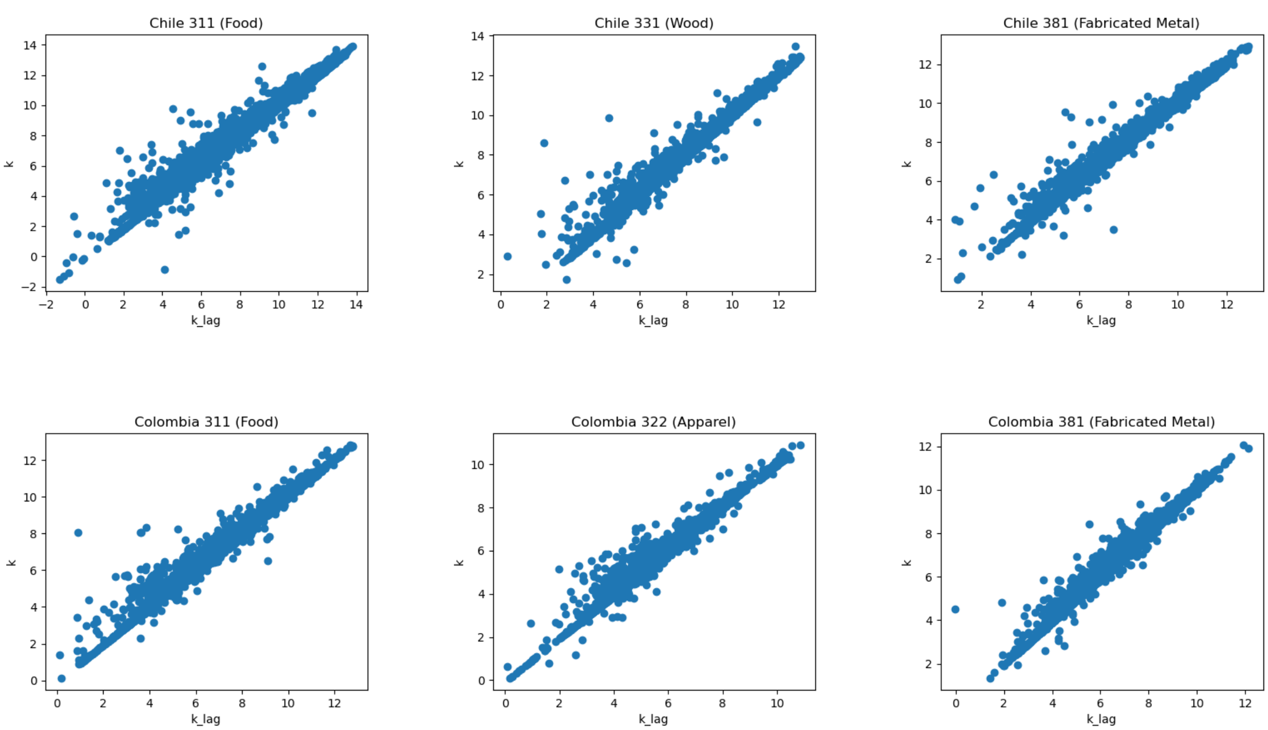
[32]We do not impose the monotonicity constraint explicitly in our estimation, but post-estimation we find that in all cases the monotonicity constraint does hold at our estimated parameters (i.e. even if we had imposed the monotonicity constraint, we would have ended up with the same estimates). If this wasn't the case, we would have needed to re-estimate the model enforcing the strict monotonicity restriction, which can be done using, e.g. Bernstein polynomials.

often produce consistent standard errors taking into account that parts of the problem are nonparametric. But an alternative is to consider what we are doing as a “flexible” parametric model whose identification is ensured by the arguments in the first half of this paper. And given our nonparametric identification arguments, this estimation approach could allow for more flexible production functions, more general first order Markov processes, and higher order Markov processes - to the extent that a data set is sufficiently large.

8.3 Discussion of the Support Condition

We start by showing that in our empirical context, Imbens and Newey (2009)’s common support condition appears unlikely to hold. This is perhaps not surprising in a production function context, since factors of production, especially capital, are highly time-persistent. Note that, under our approach, lagged capital is part of our set of control variables, so we can partially test the support condition by looking at the conditional distribution of current capital given lagged capital. This is done in Figure (2). The support of capital moves upward as lagged capital increases, which suggests a serious breakdown of the common support condition. Thus, imposing the common support condition of Imbens and Newey (2009) and estimating the model based on equation (3) seems unlikely to work well.

Figure 2: Distributions of Current Capital given Lagged Capital



NOTE.—The graphs are scatter plots of firms’ capital (k) given their lagged capital (k_lag).

8.4 Output Elasticities

Given estimates of the structural functions using our approach, Table (1) presents our estimates of average output elasticities of capital and labor, average (local) returns to scale (measured by the sum of the output elasticities of labor and capital), and capital intensity (measured by the ratio of the average output elasticities of capital and labor) for each country-industry pair, respectively. These are averages (across all firms in the data) because unlike in a Cobb-Douglas production function with a Hicks neutral productivity shock, firms in our model have different output elasticities - they depend not only on their capital and labor levels, but also their productivity shocks.[33]

For comparison purposes, in the first column of each panel we report simple OLS estimates of a Hicks neutral translog production function ignoring the endogeneity problem, and in the second column we report the results of estimation of a Hicks neutral translog production function with the endogeneity problem addressed by our timing and information set assumptions. The changes in estimates moving from the first column to the second column, despite in the anticipated direction, are not particularly large. However, when we move to the third column, our full model where we allow the productivity shock to enter in a non-Hicks neutral way, we do see quite large changes. Interestingly, the average output elasticity of labor decreases substantially for all the industries. We also see large decreases in the estimates of returns to scale, even though the average output elasticity of capital goes up in most industries. These results suggest that Hicks neutral models may be misspecified when it comes to estimating output elasticities and returns to scale. It is interesting that average output elasticities change so much (interacting the productivity shock with observed inputs more directly affects variances) - suggesting significant non-linearites. The patterns are qualitatively similar for median elasticities and returns to scale, so this does not appear to be due to tail behavior. Since output elasticities are proportional to marginal products, the results in table (1) also suggest Hicks neutral models may substantially overestimate the marginal productivity of labor, and this could affect measures of labor market power.[34]

Given that results in table (1) suggest Hicks neutral models may be misspecified, we next investigate how much of the heterogeneity in output elasticites is generated by the productivity shocks in our non-Hicks neutral model (in the Hicks neutral models in the first two columns

[33] Since Table (1) only computes elasticities at points observed in the data, the estimates only rely on the small local support condition, i.e. support condition (1). Because they involve calculating counterfactual outcomes, the subsequent Table and Figures rely on support condition (6) (or a convex support plus Assumption (6)). As usual, in empirical practice, the extent to which these support conditions don't hold will generate estimates that may be sensitive to the finite sample parametric structure of our sieves

[34] For a recent literature about identification of labor market power (markdowns) using production function estimation approaches, see, e.g., Dobbelaere and Mairesse (2013), Lu et al. (2019), Kirov and Traina (2021). See also e.g., Azar et al. (2019) and Rubens (2019) for using discrete choice model estimation approaches to identify input market power.

of table (1), *none* of the heterogeneity in output elasticites is generated by the productivity shocks). This can be thought of as quantifying the extent to which, in our non-Hicks neutral model, the productivity shock interacts with capital and labor terms in the production function. We decompose this heterogeneity in Table (2), where we report the mean, standard error, and coefficient of variation of output elasticities for both labor and capital (EL and EK). In the first two columns, we report these estimates fixing labor and capital at their mean and the median levels, respectively. Thus, the non-zero standard deviations and coefficients of variation in the first two columns arise from the non-Hicksian neutral aspects of the productivity shock. In other words, in Hicks neutral models the standard error and coefficient of variation for these two columns would be zero. The results in the first two columns are very similar - the standard errors and coefficients of variation tend to be large in magnitude and significantly different from zero. We conclude that the productivity shock generates a significant amount of heterogeneity in both EL and EK.

In the third column of each panel in Table (2), we report the same distributional statistics evaluated at firms' actual values of labor and capital. Thus, the heterogeneity (measured by the standard deviation (SD) and coefficient of variation (CV)) in this column comes from variation in *both* the productivity shock *and* in input levels across firms. Comparing the SD and CV in this column to the first two columns thus speaks to the extent to which heterogeneity is driven by the productivity shock vs the levels of the observed inputs. Interestingly, the relative increase in the SD's (and CV's) moving from the first two columns to the third is substantially larger for capital in all six industries. This indicates that heterogeneity in EK is relatively more driven by variation in observed inputs than is heterogeneity in EL, and heterogeneity in EL is relatively more driven by the non-Hicks neutral productivity shock. This suggests that Hicks neutral models may do worse at capturing heterogeneity in EL (than in EK), and seems supportive of the specification choice in some of the related literature to parameterize additional shocks as directly impacting labor, e.g., the "labor-augmenting" shocks of Doraszelski and Jaumandreu (2018), Raval (2019), and Demirer (2020).

Table 1: Average Input Elasticities of Output

	Industry (ISIC Code)								
Chile:	Food Products (311)			Wood Products (331)			Fabricated Metals (381)		
	OLS Translog	Endogenous Translog	Endogenous Nonseparable	OLS Translog	Endogenous Translog	Endogenous Nonseparable	OLS Translog	Endogenous Translog	Endogenous Nonseparable
Labor	0.88	0.77	0.59	0.96	0.93	0.86	0.96	0.92	0.77
	(0.01)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.03)	(0.04)
Capital	0.35	0.35	0.37	0.21	0.21	0.24	0.25	0.26	0.31
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Sum	1.22	1.12	0.96	1.17	1.15	1.10	1.21	1.18	1.08
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)	(0.02)	(0.03)
Mean (capital)/mean (labor)	0.40	0.46	0.63	0.22	0.23	0.28	0.26	0.29	0.40
	(0.01)	(0.02)	(0.04)	(0.02)	(0.03)	(0.04)	(0.01)	(0.02)	(0.04)
Colombia:	Food Products (311)			Apparel (322)			Fabricated Metals (381)		
	OLS Translog	Endogenous Translog	Endogenous Nonseparable	OLS Translog	Endogenous Translog	Endogenous Nonseparable	OLS Translog	Endogenous Translog	Endogenous Nonseparable
Labor	0.73	0.72	0.53	0.90	0.85	0.63	0.92	0.87	0.64
	(0.01)	(0.04)	(0.04)	(0.01)	(0.03)	(0.05)	(0.02)	(0.04)	(0.04)
Capital	0.36	0.34	0.42	0.13	0.14	0.23	0.27	0.29	0.37
	(0.01)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.03)	(0.03)
Sum	1.09	1.06	0.94	1.04	1.00	0.86	1.19	1.16	1.00
	(0.01)	(0.02)	(0.03)	(0.01)	(0.02)	(0.04)	(0.01)	(0.04)	(0.04)
Mean (capital)/mean (labor)	0.49	0.47	0.79	0.15	0.17	0.37	0.30	0.34	0.57
	(0.02)	(0.05)	(0.09)	(0.01)	(0.03)	(0.06)	(0.02)	(0.04)	(0.07)

NOTE.—The number of observations for each dataset are 14,904, 3,601, 4,456, 5,040, 4,454, and 2,958, respectively. In the parenthesis are bootstrapped standard errors (200 replications) where resampling is done at the firm (plant) level. The numbers in the first column are based on a translog production function with Hicks neutral productivity and are estimated using OLS. The numbers in the second column are based on a translog production function with Hicks neutral productivity and are estimated under our timing and information set assumptions, with the endogeneity problem addressed. The numbers in the third column are based on our nonseparable model, which is estimated using the sieve MLE procedure described in the main text. The "Labor" and "Capital" row report average output elasticities of labor and capital respectively, the "Sum" row reports the sum of the average labor and capital elasticities, and the "Mean (capital)/mean (labor)" row reports the ratio of the average capital elasticity to the average labor elasticity.

Table 2: Heterogeneity in Output Elasticities

	Industry (ISIC Code)								
Chile:	Food Products (311)			Wood Products (331)			Fabricated Metals (381)		
	Evaluated at Mean k&l	Evaluated at Median k&l	Evaluated at Actual k&l	Evaluated at Mean k&l	Evaluated at Median k&l	Evaluated at Actual k&l	Evaluated at Mean k&l	Evaluated at Median k&l	Evaluated at Actual k&l
Mean(EL)	0.64	0.67	0.64	0.86	0.86	0.86	0.77	0.79	0.77
	(0.02)	(0.03)	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
SD(EL)	0.08	0.08	0.14	0.06	0.06	0.08	0.07	0.06	0.13
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
CV(EL)	0.13	0.12	0.22	0.08	0.07	0.09	0.08	0.08	0.17
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)
Mean(EK)	0.39	0.36	0.39	0.25	0.24	0.25	0.30	0.30	0.30
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
SD(EK)	0.05	0.05	0.19	0.03	0.03	0.11	0.01	0.02	0.05
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
CV(EK)	0.12	0.13	0.48	0.13	0.13	0.42	0.05	0.05	0.18
	(0.01)	(0.02)	(0.03)	(0.05)	(0.05)	(0.08)	(0.03)	(0.03)	(0.06)
Colombia:	Food Products (311)			Apparel (322)			Fabricated Metals (381)		
	Evaluated at Mean k&l	Evaluated at Median k&l	Evaluated at Actual k&l	Evaluated at Mean k&l	Evaluated at Median k&l	Evaluated at Actual k&l	Evaluated at Mean k&l	Evaluated at Median k&l	Evaluated at Actual k&l
Mean(EL)	0.54	0.54	0.54	0.64	0.63	0.64	0.65	0.65	0.65
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
SD(EL)	0.09	0.09	0.12	0.09	0.09	0.11	0.12	0.12	0.19
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
CV(EL)	0.17	0.17	0.23	0.14	0.14	0.17	0.18	0.18	0.30
	(0.04)	(0.04)	(0.05)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
Mean(EK)	0.41	0.41	0.41	0.23	0.23	0.23	0.36	0.37	0.36
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
SD(EK)	0.07	0.07	0.21	0.03	0.03	0.09	0.05	0.05	0.19
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.03)
CV(EK)	0.16	0.16	0.51	0.12	0.12	0.40	0.13	0.13	0.53
	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)	(0.08)	(0.04)	(0.04)	(0.07)

NOTE.— In the parenthesis are bootstrapped standard errors (200 replications) where resampling is done at the firm (plant) level. The numbers in the first two columns are counterfactual quantities that are estimated holding labor and capital at the mean and median levels, respectively. The numbers in the third column are evaluated at the actual observed values of labor and capital. "EL" and "EK" stand for output elasticities of labor and capital. "Mean()", "SD()", and "CV()" stand for mean, standard deviation and coefficient of variation of the distributions.

8.5 Biased Technological Change

By definition, Hicks neutrality implies that an increase in the productivity shock increases the marginal productivity of capital and labor by the same proportion. Since Tables (1) and (2) have provided strong evidence against Hicks neutrality of the productivity shock, the next question that we examine is whether this non-neutrality favors one factor of production over another. This determines whether technological change in the form of increased productivity shocks is "biased" toward capital or toward labor. As argued by Acemoglu (2002), for many problems in macroeconomics, development economics, labor economics, and international trade, whether technological change is biased toward particular factors is of central importance. In their work that utilizes a CES production function with a labor augmenting shock to "break" Hicks neutrality, Doraszelski and Jaumandreu (2018) find that technological change is capital biased.[35] So it is interesting to assess the same question with our alternative non-Hicks neutral model based on different assumptions. Following Acemoglu (2002), we first formally define the notion of "biased technological change".[36]

Definition 8 *Technological change is biased toward input x_1 over x_2 if*

$$\frac{\partial \frac{\partial F(\mathbf{x},\omega)/\partial x_1}{\partial F(\mathbf{x},\omega)/\partial x_2}}{\partial \omega} \geq 0, \tag{15}$$

i.e., if an increase in the productivity shock ω increases the marginal productivity of x_1 relatively more than it increases the marginal productivity of x_2.

We show the bias in our estimated non-Hicks neutral production functions in figure (3). We graph the ratio of the marginal productivity of capital (MPK) to the marginal productivity of labor (MPL) (i.e. the marginal rate of technical substitution (MRTS)), as a function of productivity shock (ω), holding capital and labor constant at their median levels.[37] If technological change is Hicks neutral, then the graph will be a horizontal line. However, as one can see, the graphs are upward-sloping for all industries, meaning technological change is capital biased. This is not a small effect - for example, for the industry of Chile 311 the ratio nearly doubles when productivity shock moves from the 10th percentile (the left side of the graph in each panel) to the 90th percentile (the right side). While the bootstrapped confidence intervals at different percentiles (e.g. the 10th vs 90th) tend to overlap on the figures, the estimation error

[35] More precisely, they estimate a CES production function, and their estimated elasticity of substitution is less than one, so the labor-augmenting productivity shock is biased toward capital (and intermediate inputs).

[36] In our datasets, we generally observe that productivity levels across most industries increase over time. This upward trend suggests technological advancements or changes.

[37] We find similar patterns when holding capital and labor at other representative values, e.g., 25th and 75th percentiles.

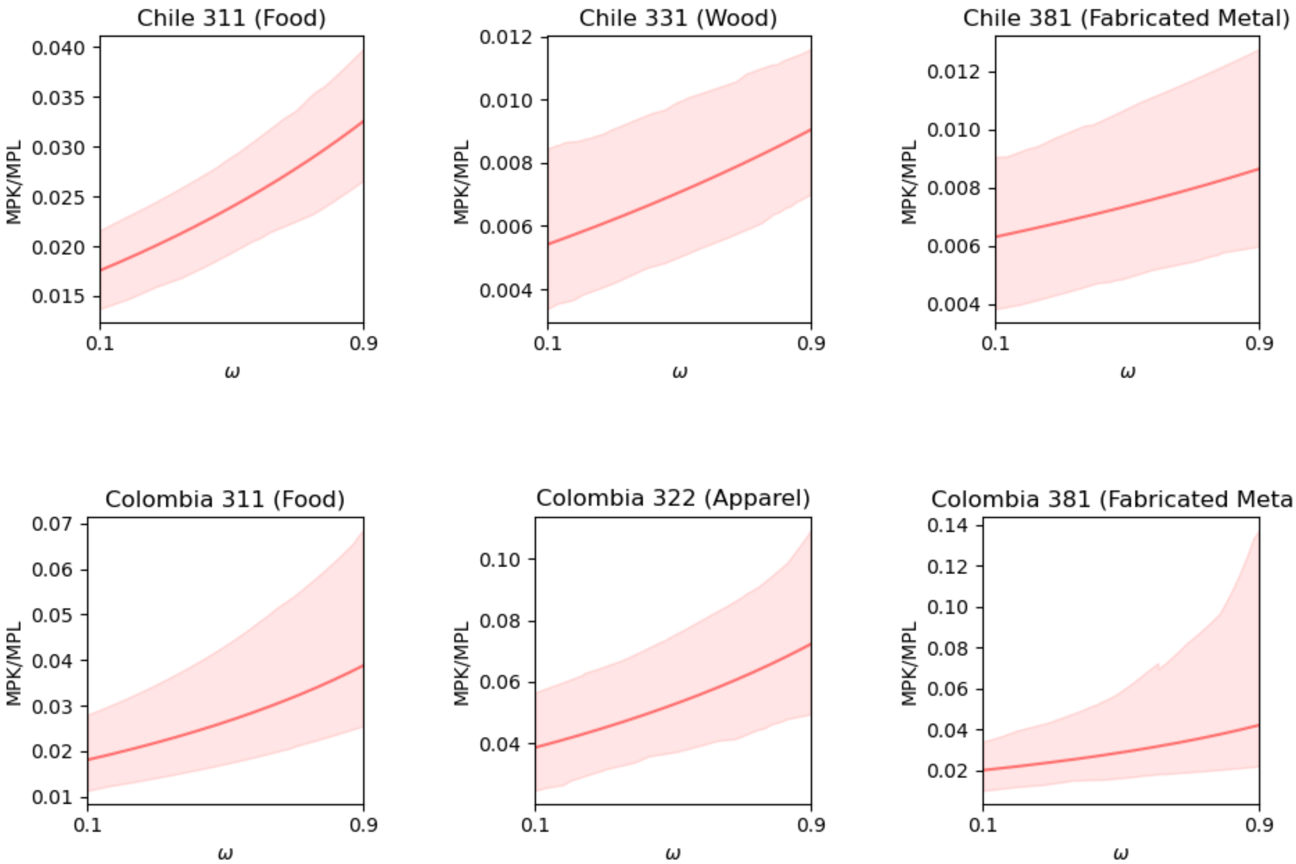
is correlated across different percentiles (since they depend on the same underlying parameters of our sieves). Therefore, to formally test whether the capital bias is statistically significant, figure (4) reports the *difference* in the MRTS at each percentile relative to the MRTS at the 10th percentile. The fact that the confidence intervals do not cover 0 confirms that the effects are statistically significant in all 6 industries.

Capital-biased technological change has important economic implications. A series of papers in the recent literature, including Doraszelski and Jaumandreu (2018), Zhang (2019), and Oberfield and Raval (2021), argue that biased technological change is one of the primary driving forces behind the recent secular trend of declining labor share in national income. The logic goes as follows: if relative prices of inputs remain constant, capital-biased technological change will tend to increase firms' demand for capital relative to labor, i.e., capital-biased technological change is, as also pointed out by Van Biesebroeck (2003), tends to be labor-saving.[38] Our finding of capital-biased technological change in the context of our alternative assumptions is supportive of their conclusions. Second, capital-biased technological change implies that high-productivity firms have a "comparative advantage" in using capital compared to low-productivity firms, i.e., high-productivity firms are relatively more efficient in using capital than low-productivity firms. This can have important implications on allocative efficiency.

In sum, we believe our results suggesting capital-biased technological change across multiple industries in two countries are interesting in relation to the recent literature concerning factor-augmenting productivity shocks, e.g., Doraszelski and Jaumandreu (2018), Raval (2019), Zhang (2019), and Oberfield and Raval (2021). As described above, one major difference is that these papers typically assume a CES production function with an additional labor-augmenting productivity shock, while we allow a more flexible production function with a scalar productivity shock. But there are other differences. Those papers typically assume that labor is fully flexible but has no dynamic implications. On the other hand, we make a stronger timing assumption that labor is predetermined, but because we do not rely on a first order condition w.r.t. the labor input, our approach allows labor to have dynamic effects, allows firms to potentially have monopsony power in labor markets, and does not require observed measures of input prices. Given the distinctiveness of the different assumptions, we hope the two approaches are viewed as complementary - empirical conclusions robust to both approaches and sets of assumptions would seem to be more convincing than those using only one. And in this case, we believe our finding of capital-biased technological change is highly supportive of the findings of the existing factor-augmenting literature.

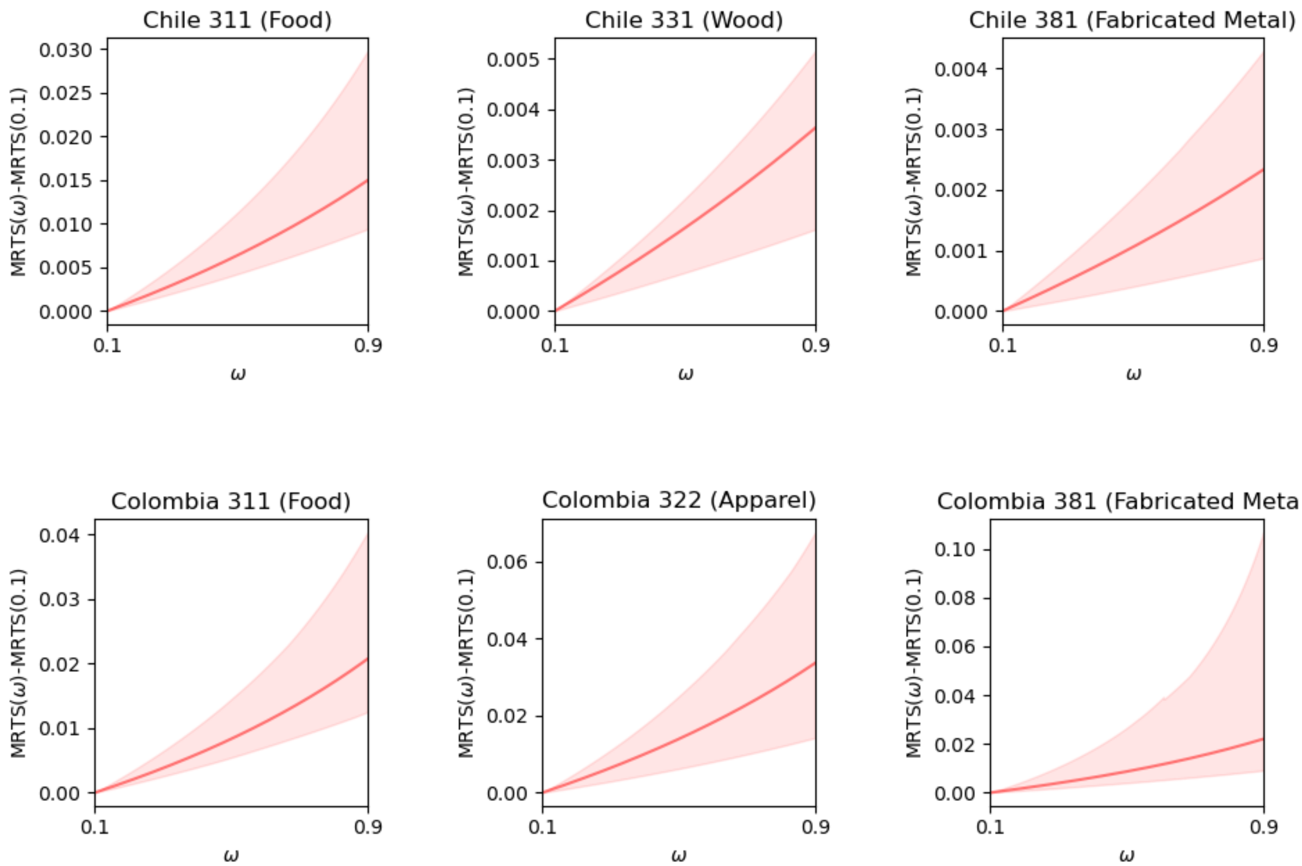
[38] Note that in our non-parametric production function, verifying this (and even calculating the related elasticity of substitution) is considerably more complicated than in a CES production function.

Figure 3: Bias of Technological Change (Ratio)



NOTE.—The graphs show how, for each industry, MRTS (the ratio of marginal productivity of capital (MPK) to the marginal productivity of labor (MPL)), changes with the productivity shock (ω). On the horizontal axis, the left represents the 10th percentile of the distribution of ω_{it} and the right end represents the 90th percentile. The shaded areas are 90% confidence bands.

Figure 4: Bias of Technological Change (Difference of Ratio)



NOTE.— The graphs show the difference in the MRTS at different percentiles of the ω_{it} distributions versus the MRTS at the 10th percentile of the ω_{it} distribution. The shaded areas are 90% confidence bands.

9 Conclusion

We have illustrated that the "timing and information set assumption" approach to solving endogeneity problems has identification power in a fully nonparametric model with a nonseparable error term. This means that empirical researchers can be quite flexible in these contexts, and perhaps be more comfortable that results are not driven by functional form assumptions. We apply this result to a variety of production datasets using a sieve (partial) maximum likelihood estimator, finding evidence of non-Hicks neutral technology shocks. These results are supportive of other recent empirical papers examining these phenomena.

References

Acemoglu, D. (2002). Directed technical change. *The Review of Economic Studies*, *69*(4), 781–809.

Ackerberg, D. (2020). *Timing assumptions and efficiency: Empirical evidence in a production function context*. Retrieved March 21, 2022, from https://drive.google.com/file/d/1RyDQt_o6nLGkI22DsyxbLLqljLxErsfr/view

Ackerberg, D., Benkard, C. L., Berry, S., & Pakes, A. (2007). Econometric tools for analyzing market outcomes. *Handbook of econometrics*, *6*, 4171–4276.

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

Ackerberg, D., Chen, X., & Hahn, J. (2012). A practical asymptotic variance estimator for two-step semiparametric estimators. *Review of Economics and Statistics*, *94*(2), 481–498.

Ackerberg, D., & De Loecker, J. (2024). Production function identiöcation uunder imperfect competition. *Working Paper*.

Ackerberg, D., Frazer, G., il Kim, K., & Su, Y. (2020). Under-identification of structural models based on timing and information set assumptions.

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

Altonji, J., & Matzkin, R. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, *73*(4), 1053–1102.

Anderson, T. W., & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of econometrics*, *18*(1), 47–82.

Arellano, M., & Bond, S. (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.

Azar, J., Berry, S., & Marinescu, I. E. (2019). Estimating labor market power. *Available at SSRN 3456277*.

Bajari, P., Fruehwirth, J. C., Timmins, C., et al. (2012). A rational expectations approach to hedonic price regressions with time-varying unobserved product attributes: The price of pollution. *American Economic Review*, *102*(5), 1898–1926.

Balat, J., Brambilla, I., & Sasaki, Y. (2016). *Heterogeneous firms: Skilled-labor productivity and the destination of exports* (tech. rep.). Working paper.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, *87*(1), 115–143.

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

Blundell, R., & Powell, J. (2001). Endogeneity in nonparametric and semiparametric regression models.

Blundell, R., & Smith, R. (1989). Estimation in a class of simultaneous equation limited dependent variable models. *The Review of Economic Studies*, *56*(1), 37–57.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of econometrics*, *18*(1), 5–46.
Chernozhukov, V., & Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, *73*(1), 245–261.
Chernozhukov, V., Imbens, G. W., & Newey, W. K. (2007). Instrumental variable estimation of nonseparable models. *Journal of Econometrics*, *139*(1), 4–14.
Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, *71*(5), 1405–1441.
Chesher, A. (2007). Identification of nonadditive structural functions. *ECONOMETRIC SOCIETY MONOGRAPHS*, *43*, 1.
Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, *78*(3), 883–931.
De Roux, N., Eslava, M., Franco, S., & Verhoogen, E. (2021). *Estimating production functions in differentiated-product industries with quantity information and external instruments* (tech. rep.). National Bureau of Economic Research.
Dearing, A. (2025). Identification and estimation of demand using timing and information set assumptions. *mimeo, Cornell University*.
Debreu, G. (1954). Representation of a preference ordering by a numerical function. *Decision processes*, *3*, 159–165.
Demirer, M. (2020). Production function estimation with factor-augmenting technology: An application to markups. *Job Market Paper*.
D'Haultfœuille, X., & Février, P. (2015). Identification of nonseparable triangular models with discrete instruments. *Econometrica*, *83*(3), 1199–1210.
Dobbelaere, S., & Mairesse, J. (2013). Panel data estimates of the production function and product and labor market imperfections. *Journal of Applied Econometrics*, *28*(1), 1–46.
Doraszelski, U., & Jaumandreu, J. (2013). R&d and productivity: Estimating endogenous productivity. *Review of Economic Studies*, *80*(4), 1338–1383.
Doraszelski, U., & Jaumandreu, J. (2018). Measuring the bias of technological change. *Journal of Political Economy*, *126*(3), 1027–1084.
Florens, J.-P., Heckman, J. J., Meghir, C., & Vytlacil, E. (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica*, *76*(5), 1191–1206.
Fox, J. T., Haddad, V., Hoderlein, S., Petrin, A. K., & Sherman, R. P. (2017). *Heterogenous production functions, panel data, and productivity dispersion*. Retrieved March 15, 2022, from https://econ2017.sites.olt.ubc.ca/files/2019/03/pdf_seminar_Productivity-Slides_Fox.pdf
Gandhi, A., Navarro, S., & Rivers, D. A. (2020). On the identification of gross output production functions. *Journal of Political Economy*, *128*(8), 2973–3016.
Grennan, M. (2013). Price discrimination and bargaining: Empirical evidence from medical devices. *American Economic Review*, *103*(1), 145–77.
Hahn, J., Liao, Z., & Ridder, G. (2021). *Identification and the influence function of olley and pakes'(1996) production function estimator* (tech. rep.). Working Paper, UCLA and USC.
Hansen & Singleton, K. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society*, 1269–1286.

Heckman, J. J. (1977). *Dummy endogenous variables in a simultaneous equation system* (tech. rep.). National Bureau of Economic Research.

Hu, Y., & Shum, M. (2012). Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics*, *171*(1), 32–44.

Hu, Y., & Shum, M. (2013). Identifying dynamic games with serially correlated unobservables. In *Structural econometric models*. Emerald Group Publishing Limited.

Imbens, G. W., & Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, *77*(5), 1481–1512.

Jaffray, J.-Y. (1975). Existence of a continuous utility function: An elementary proof. *Econometrica: Journal of the Econometric Society*, 981–983.

Kasahara, H., Nishida, M., & Suzuki, M. (2017). *Decomposition of aggregate productivity growth with unobserved heterogeneity*. RIETI.

Kasahara, H., Schrimpf, P., & Suzuki, M. (2015). Identification and estimation of production function with unobserved heterogeneity. *University of British Columbia mimeo*.

Kirov, I., & Traina, J. (2021). *Labor market power and technological change in us manufacturing*. Retrieved March 15, 2022, from https://raw.githubusercontent.com/James-Traina/Preprints/main/Traina_JMP.pdf

Lee, R. S. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review*, *103*(7), 2960–3000.

Lehmann, E. L., Romano, J. P., & Casella, G. (2005). *Testing statistical hypotheses* (Vol. 3). Springer.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies*, *70*(2), 317–341.

Li, T., & Sasaki, Y. (2024). Identification of heterogeneous elasticities in gross-output production functions. *Journal of Econometrics*, *238*(2), 105637.

Lu, Y., Sugita, Y., & Zhu, L. (2019). *Wage and markdowns and fdi liberalization* (tech. rep.). Hitotsubashi Institute for Advanced Study, Hitotsubashi University.

Matzkin, R. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, *71*(5), 1339–1375.

Matzkin, R. (2004). Unobservable instruments. *Northwestern University Department of Economics Working Paper*.

Matzkin, R. (2007). Nonparametric identification. *Handbook of econometrics*, *6*, 5307–5368.

Navarro, S., & Rivers, D. A. (2018). *Nonparametric identification of productivity in nonseparable production functions* (tech. rep.). Working paper.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society*, 1349–1382.

Oberfield, E., & Raval, D. (2021). Micro data and macro technology. *Econometrica*, *89*(2), 703–732.

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

Pan, Q. (2021). Support conditions in control function approaches. *Unpublished manuscript*.

Pan, Q. (2022). Identification of gross output production function with nonseparable productivity shock. *Unpublished manuscript*.

Raval, D. (2019). The micro elasticity of substitution and non-neutral technology. *The RAND Journal of Economics*, *50*(1), 147–167.

Raval, D. (2020). Testing the production approach to markup estimation. *Available at SSRN 3324849*.
Rubens, M. (2019). Monopsony power and factor-biased technology adoption. *University of Leuven mimeo*.
Rubinstein, A. (2012). *Lecture notes in microeconomic theory*. Princeton University Press.
Sullivan, C. (2017). The ice cream split: Empirically distinguishing price and product space collusion. *Available at SSRN 3321948*.
Sweeting, A. (2013). Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry. *Econometrica*, *81*(5), 1763–1803.
Torgovitsky, A. (2015). Identification of nonseparable models using instruments with small support. *Econometrica*, *83*(3), 1185–1197.
Van Biesebroeck, J. (2003). Productivity dynamics with technology choice: An application to automobile assembly. *The review of economic studies*, *70*(1), 167–198.
Vuong, Q., & Xu, H. (2017). Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity. *Quantitative Economics*, *8*(2), 589–610.
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
Zhang, H. (2019). Non-neutral technology, firm heterogeneity, and labor demand. *Journal of Development Economics*, *140*, 145–168.

Appendices

A Lemmas

Lemma 2 *ς_{it}^1 is independent of $\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$.*

Proof. By construction,

$$p\left(\varsigma_{it}^1 \middle| \{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right) \sim U(0,1)$$

regardless of the realization of $\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}$. ■

Lemma 3 *ξ_{it}, ς_{it}^1, and $\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$ are independent of each other.*

Proof. Since $\varsigma_{it}^1 = F_{x_{it}^1|\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}}\left(x_{it}^1, \{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$, and since $x_{it} = h_t\left(\mathcal{I}_{it-1}\right)$ by Assumption (1), we can conclude that the ς_{it}^1 is a function of $\mathcal{I}_{it-1}$. Therefore, both ς_{it}^1 and $\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$ are functions of $\mathcal{I}_{it-1}$. Because ξ_{it} is independent of $\mathcal{I}_{it-1}$ by construction, we can conclude that ξ_{it} is independent of $\left(\varsigma_{it}^1, \left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)\right)$. By Lemma (2), we have ς_{it}^1 and $\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$ independent of each other. We therefore conclude that ξ_{it}, ς_{it}^1, and $\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$ are independent of each other. ■

Lemma 4 *$(\varsigma_{it}^1, \varsigma_{it}^2)$ is independent of $\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$, and ς_{it}^1 and ς_{it}^2 are independent of each other.*

Proof. By construction,

$$p\left(\varsigma_{it}^{2}\middle|\left\{y_{i\tau}\right\}_{\tau=t-M}^{t-1},\left\{x_{i\tau}\right\}_{\tau=t-M}^{t-1},\varsigma_{it}^{1}\right)\sim U(0,1)$$

regardless of the realization of $\left(\left\{y_{i\tau}\right\}_{\tau=t-M}^{t-1},\left\{x_{i\tau}\right\}_{\tau=t-M}^{t-1},\varsigma_{it}^{1}\right)$. By Lemma (2), we know that ς_{it}^{1} is independent of $\left(\left\{y_{i\tau}\right\}_{\tau=t-M}^{t-1},\left\{x_{i\tau}\right\}_{\tau=t-M}^{t-1}\right)$. The conclusion follows from these two observations. ■

Lemma 5 *ξ_{it}, $(\varsigma_{it}^{1},\varsigma_{it}^{2})$, and $\left(\left\{y_{i\tau}\right\}_{\tau=t-M}^{t-1},\left\{x_{i\tau}\right\}_{\tau=t-M}^{t-1}\right)$ are independent of each other.*

Proof. since $\varsigma_{it}^{2}=F_{x_{it}^{2}|\{y_{i\tau}\}_{\tau=t-M}^{t-1},\{x_{i\tau}\}_{\tau=t-M}^{t-1},\varsigma_{t}^{1}}\left(x_{it}^{2},\left\{y_{i\tau}\right\}_{\tau=t-M}^{t-1},\left\{x_{i\tau}\right\}_{\tau=t-M}^{t-1},\varsigma_{it}^{1}\right)$, and since $x_{it}=h_{t}\left(\mathcal{I}_{it-1}\right)$ by Condition 1, we can conclude that the ς_{it}^{2} is a function of $\mathcal{I}_{it-1}$. Therefore, both $(\varsigma_{it}^{1},\varsigma_{it}^{2})$ and $\left\{\omega_{i\tau}\right\}_{\tau=t-M}^{t-1}$ are functions of $\mathcal{I}_{it-1}$. Because ξ_{it} is independent of $\mathcal{I}_{it-1}$, we can conclude that ξ_{it} is independent of $\left((\varsigma_{it}^{1},\varsigma_{it}^{2}),\left\{\omega_{i\tau}\right\}_{\tau=t-M}^{t-1}\right)$. By Lemma (4), we have $(\varsigma_{it}^{1},\varsigma_{it}^{2})$ and $\left(\left\{y_{i\tau}\right\}_{\tau=t-M}^{t-1},\left\{x_{i\tau}\right\}_{\tau=t-M}^{t-1}\right)$ independent of each other, from which the conclusion follows. ■

Lemma 6 *ξ_{it} and ς_{it} are independent of each other given $\left(\left\{y_{i\tau}\right\}_{\tau=t-M}^{t-1},\left\{x_{i\tau}\right\}_{\tau=t-M}^{t-1}\right)$.*

Proof. By iterating Lemmas (2) - (5), we obtain ξ_{it}, ς_{it}, and $\left(\left\{y_{i\tau}\right\}_{\tau=t-M}^{t-1},\left\{x_{i\tau}\right\}_{\tau=t-M}^{t-1}\right)$ are independent of each other, from which the conclusion follows. ■

B Additional Proofs

B.1 Proof of Theorem (4)

Proof.

Given Lemma (1), it is sufficient to prove that for any two points (x^A,y^A), $(x^B,y^B)\in\mathcal{S}_t^{xy}$, we can order $f_t^{-1}(x^A,y^A)$ vs $f_t^{-1}(x^B,y^B)$.

To do that, first we prove the statement: given two points (x^A,y^A), $(x^B,y^B)\in\mathcal{S}_t^{xy}$, if $\mathcal{W}^N(x^A,y^A)$ and $\mathcal{W}^N(x^B,y^B)$ have a common support point x^0, then we can order $f_t^{-1}(x^A,y^A)$ vs $f_t^{-1}(x^B,y^B)$. By definition of $\mathcal{W}^N(x^A,y^A)$ and $\mathcal{W}^N(x^B,y^B)$, we can find some y^{0A} and y^{0B} such that $(x^0,y^{0A})\in\mathcal{W}^N(x^A,y^A)$, $(x^0,y^{0B})\in\mathcal{W}^N(x^B,y^B)$, and $f_t^{-1}(x^0,y^{0A})=f_t^{-1}(x^A,y^A)$, $f_t^{-1}(x^0,y^{0B})=f_t^{-1}(x^B,y^B)$. Since we assume $f_t(x,\omega)$ is strictly monotone in ω, $y^{0A}\gtreqless y^{0B}\Leftrightarrow f_t^{-1}(x^A,y^A)\gtreqless f_t^{-1}(x^B,y^B)$.

Given this, for each consecutive pairs of points $((x^j,y^j),(x^{j+1},y^{j+1}))$ in the omegically monotone sequence given in support condition (6), we can order $f_t^{-1}(x^j,y^j)$ vs $f_t^{-1}(x^{j+1},y^{j+1})$. Since either $f_t^{-1}(x^0,y^0)\geq...\geq f_t^{-1}(x^{J+1},y^{J+1})$ or $f_t^{-1}(x^0,y^0)\leq...\leq f_t^{-1}(x^{J+1},y^{J+1})$ is true, we can order $f_t^{-1}(x^A,y^A)$ vs $f_t^{-1}(x^B,y^B)$. ■

B.2 Proof of Theorem (5)

Proof. Given Lemma (1), we only need to order any $f_t^{-1}(x^A,y^A)$ vs $f_t^{-1}(x^B,y^B)$, and because the boundary of $\mathcal{S}_t^{xy}$ has probability measure zero, we only need to consider (x^A,y^A), $(x^B,y^B)\in Int\left(\mathcal{S}_t^{xy}\right)$. Given Assumption (6) (iii), we can find some v^A and v^B such that (x^A,v^A), $(x^B,v^B)\in Int(\mathcal{S}_t^{xv})$. And under Assumption (6) (i) , we also know (x^A,v^A,y^A), $(x^B,v^B,y^B)\in\mathcal{S}_t^{xvy}$. Now consider a straight line connecting (x^A, v^A) and (x^B, v^B) defined by $p(z) = (x(z), v(z))$ and indexed by $z \in [0,1]$ s.t. $p(0) = (x^A, v^A)$ and $p(1) = (x^B, v^B)$. Because $Int(\mathcal{S}_t^{xv})$ is open and convex, every point on the line $p(z)$ is in $Int(\mathcal{S}_t^{xv})$. In addition we can find an ϵ s.t. every point within distance ϵ of the line $p(z)$ is also in $Int(\mathcal{S}_t^{xv})$, i.e., $\exists$ ϵ s.t. if $\|p - p(z)\| \leq \epsilon$ for some $z \in [0,1]$, then $p \in Int(\mathcal{S}_t^{xv})$.

Now consider the following constructive algorithm that orders $f_t^{-1}(x^A, y^A)$ and $f_t^{-1}(x^B, y^B)$:

1) Start at (x^A, v^A).

2) Travel distance ϵ along $p(z)$. Denote the new point (x^{new}, v^{new}). We know $(x^{new}, v^{new}) \in Int(\mathcal{S}_t^{xv})$. Also consider the point (x^A, v^{new}). Since $\left\|(x^A, v^{new}) - (x^{new}, v^{new})\right\| \leq \epsilon$, it must also be the case that $(x^A, v^{new}) \in Int(\mathcal{S}_t^{xv})$.

3) By Assumption (6) (i), since $(x^A, v^A, y^A) \in \mathcal{S}_t^{xvy}$ it must also be the case that $(x^A, v^{new}, y^A) \in \mathcal{S}_t^{xvy}$.

4) Using the identified $\bar{f}_t(x, v, \xi)$, determine the ξ^A corresponding to (x^A, v^{new}, y^A), i.e., $\xi^A = \bar{f}_t^{-1}(x^A, v^{new}, y^A)$.

5) Determine y^{new} corresponding to (x^{new}, v^{new}) and ξ^A, i.e., $y^{new} = \bar{f}_t(x^{new}, v^{new}, \xi^A)$.

6) By construction, $f_t^{-1}(x^{new}, y^{new}) = f_t^{-1}(x^A, y^A)$, i.e. the points have the same ω

7) Go to step 2. Continue moving along path $p(z)$ distance ϵ each step until get to $x^{new} = x^B$ (the iterate at which one reaches $x^{new} = x^B$ may require moving distance less than ϵ).

8) Compare the resulting y^{new} to y^B. $y^{new} \gtreqless y^B \to f_t^{-1}(x^A, y^A) \gtreqless f_t^{-1}(x^B, y^B)$. ■

B.3 Proof of Theorem (7)

Proof. Plugging in g_t and substituting ω_{it} with f_t^{-1}, we can write the observed ξ_t^0th quantile of y_{it} conditional on $(x_{it}, x_{it-1}, y_{it-1}) = \left(x_t^0, x_{t-1}^0, y_{t-1}^0\right)$ as

$$q_{y_{it}|x_{it},v_{it-1}}(\xi_t^0|x_t^0, x_{t-1}^0, y_{t-1}^0) = f_t(x_t^0, g_t(f_{t-1}^{-1}(x_{t-1}^0, y_{t-1}^0), \xi_t^0)).$$

We know this equality holds because f_t is strictly monotone in ω_{it}, g_t is strictly monotone in ξ_{it}, and ξ_{it} is independent from $(x_{it}, x_{it-1}, y_{it-1})$. We can see from the above equation that (x_{it-1}, y_{it-1}) affects the conditional quantile of y_{it} only through f_{t-1}^{-1}, so we can rely on this structural variation to identify aspects of f_{t-1}. Taking the negative ratios of derivatives of the conditional quantile w.r.t. x_{it-1} and y_{it-1} at $\left(x_t^0, x_{t-1}^0, y_{t-1}^0\right)$ gives

$$\begin{aligned}
&- \frac{\partial q_{y_{it}|x_{it},v_{it-1}}(\xi_t^0|x_t^0, x_{t-1}^0, y_{t-1}^0)}{\partial x_{it-1}} \Big/ \frac{\partial q_{y_{it}|x_{it},v_{it-1}}(\xi_t^0|x_t^0, x_{t-1}^0, y_{t-1}^0)}{\partial y_{it-1}} \\
&= -\left(\frac{\partial f_t(x_t^0, \omega_t^0)}{\partial \omega_{it}} \frac{\partial g_t(\omega_{t-1}^0, \xi_t^0)}{\partial \omega_{it-1}} \frac{\partial f_{t-1}^{-1}(x_{t-1}^0, y_{t-1}^0)}{\partial x_{it-1}}\right) \Big/ \left(\frac{\partial f_t(x_t^0, \omega_t^0)}{\partial \omega_{it}} \frac{\partial g_t(\omega_{t-1}^0, \xi_t^0)}{\partial \omega_{it-1}} \frac{\partial f_{t-1}^{-1}(x_{t-1}^0, y_{t-1}^0)}{\partial y_{it-1}}\right) \\
&= -\frac{\partial f_{t-1}^{-1}(x_{t-1}^0, y_{t-1}^0)}{\partial x_{it-1}} \Big/ \frac{\partial f_{t-1}^{-1}(x_{t-1}^0, y_{t-1}^0)}{\partial y_{it-1}} \\
&= \frac{\partial f_{t-1}(x_{t-1}^0, \omega_{t-1}^0)}{\partial x_{it-1}},
\end{aligned}$$

where $\omega_{t-1}^0 = f_{t-1}^{-1}(x_{t-1}^0, y_{t-1}^0)$, $\omega_t^0 = g_t(\omega_{t-1}^0, \xi_t^0)$, and the last equality follows from the implicit function theorem. ■

C IV Approach to Relaxing the Timing Assumption

In section (5.1), we have shown that when the timing and information set assumption of x_{it} is relaxed, our model is not identified without additional restrictions. In this section we build on Chernozhukov and Hansen (2005) and use an IV approach to establish identification while allowing x_{it} and ξ_{it} to be correlated. Without causing confusion, we suppress the subscript t of functions in this section.

Recall the reduced form function $y = \bar{f}(x_{it}, v_{it-1}, \xi_{it})$ from equation (4). Our IV strategy has two steps: first, we rely on a conditional quantile restrictions to identify the reduced form function $\bar{f}$; second, we make use of one of the support conditions (2), (3), (4), (5), and (6) to identify the structural function f. In relation to the discussion in the main text regarding x_{it}^F, x_{it}^V, and $y = \bar{f}(x_{it}^F, x_{it}^V, v_{it-1}, \xi_{it})$, note that for the purpose of identification of $\bar{f}$, the x_{it}^F (the subset of x_{it}'s that satisfy our timing assumption) can be treated the same as v_{it-1}. So we define $\widetilde{v}_{it-1} = \left(v_{it-1}, x_{it}^F\right)$ and define $\widetilde{x}_{it} = x_{it}^V$ to be the elements of x_{it} that are correlated with ξ_{it}. To make use of instrument variables, we make the following assumption.

Assumption 9 *We observe a vector of instrument variables z_{it} such that $(z_{it}, \widetilde{v}_{it-1})$ are jointly independent from ξ_{it}.*

Following Chernozhukov and Hansen (2005), because $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \xi_{it})$ is strictly monotone in $\xi_{it} \sim U(0,1)$, independence of ξ_{it} and $(z_{it}, \widetilde{v}_{it-1})$ implies that for each $\tau \in (0,1)$,

$$Pr(y_{it} \leq \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau) | \widetilde{v}_{it-1}, z_{it}) = \tau. \tag{16}$$

This is because $Pr(y_{it} \leq \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{v}_{it-1}, z_{it}) = Pr(\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \xi_{it}) \leq \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{v}_{it-1}, z_{it})$ $= Pr(\xi_{it} \leq \tau|\widetilde{v}_{it-1}, z_{it}) = Pr(\xi_{it} \leq \tau) = \tau$. Equation (16) is the conditional quantile restriction that we rely on to identify $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \xi_{it})$. Identification requires showing that if there is some function $m(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$ that solves equation (16), it must be that $m(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau) = \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$ with probability one, i.e., almost surely (a.s.). Note that if we can identify the quantile response function $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$ for each $\tau \in (0,1)$, then we can identify the function $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \xi_{it})$. It is also worth noting that, as pointed out by Chernozhukov et al. (2007), the conditional quantile restriction is not the only restriction that is implied by our model. For example, our model imposes that $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$ is strictly monotone in τ, which is not implied by the conditional quantile restriction.

Again following Chernozhukov and Hansen (2005), for each $\tau \in (0,1)$, fix some small constant $\delta_\tau > 0$, define the relevant parameter space $\mathcal{L}_\tau$ as the convex hull of functions $m(.,\tau)$ that satisfy (i) for each $(\widetilde{v}, z) \in \mathcal{S}_t^{\widetilde{v}z}$, $Pr(y_{it} \leq m(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{v}, z) \in [\tau - \delta_\tau, \tau + \delta_\tau]$ and (ii) for each $(\widetilde{x}, \widetilde{v}) \in \mathcal{S}_t^{\widetilde{x}\widetilde{v}}$, $m(\widetilde{x}, \widetilde{v}, \tau) \in \mathcal{S}_t^{y|\widetilde{x}\widetilde{v}z}$ for all z such that $(\widetilde{x}, \widetilde{v}, z) \in \mathcal{S}_t^{\widetilde{x}\widetilde{v}z}$.[39] For any bounded $\Delta(\widetilde{x}, \widetilde{v}, \tau) = m(\widetilde{x}, \widetilde{v}, \tau) - \bar{f}(\widetilde{x}, \widetilde{v}, \tau)$ with $m(.,\tau) \in \mathcal{L}_\tau$ and $\epsilon_{it}^\tau = y_{it} - \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it}, \tau)$, consider two conditions:

Condition 7 $E(\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau).w_\tau(\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})|\widetilde{v}_{it-1}, z_{it}) = 0$ *a.s.* $\Rightarrow \Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau) = 0$ *a.s., for* $w_\tau(\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it}) = \int_0^1 f_{\epsilon_{it}^\tau}(\delta\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})d\delta > 0$.

[39](ii) is another restriction we impose on the relevant parameter space and thus on $\bar{f}(\widetilde{x}, \widetilde{v}, \tau)$. See Theorem (9) below and the discussion that follows.

Condition 8 *For each $\widetilde{v}^0 \in \mathcal{S}_t^{\widetilde{v}}$, conditional on $\widetilde{v} = \widetilde{v}^0$, $\varphi_\tau(\widetilde{x}|\widetilde{v}^0, z) \equiv c_\tau(\widetilde{v}^0, z) w_\tau(\widetilde{x}, \widetilde{v}^0, z) f_{\widetilde{x}_{it}}(\widetilde{x}|\widetilde{v}^0, z)$ is a full rank exponential or other boundedly-complete family.*[40] [41]

Condition (7) is a bounded completeness condition, which is sufficient for global identification of $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$. By definition of $\varphi_\tau(\widetilde{x}|\widetilde{v}^0, z)$, for each $\widetilde{v}^0 \in \mathcal{S}_t^v$, $E(\Delta(\widetilde{x}_{it}, \widetilde{v}^0, \tau).w_\tau(\widetilde{x}_{it}, \widetilde{v}^0, z_{it}) |\widetilde{v}^0, z_{it}) \propto E_{\varphi_\tau(.|\widetilde{v}^0, z)}(\Delta(\widetilde{x}_{it}, \widetilde{v}^0, \tau))$. Here $E_{\varphi_\tau(.|\widetilde{v}^0, z)}$ denotes the expectation with $\varphi_\tau(\widetilde{x}|\widetilde{v}^0, z)$ used as a density. It follows by Lehmann et al. (2005) that Condition (8) suffices for Condition (7). Condition (8) might be reasonable because the exponential families includes a broad variety of distributions. The "full rank" restriction requires that the impact of instrument z_{it} on the distribution of $\widetilde{x}_{it}$ is sufficiently rich. Corresponding to Theorem 4 of Chernozhukov and Hansen (2005), the following theorem establishes the identification of our reduced form function $\bar{f}$.

Theorem 9 *Under the assumptions of our model, suppose supports $\mathcal{S}_t^y$ and $\mathcal{S}_t^{\widetilde{x}\widetilde{v}}$ are bounded, and for all $(\widetilde{x}, \widetilde{v}, z) \in \mathcal{S}_t^{\widetilde{x}\widetilde{v}z}$, $\mathcal{S}_t^{\xi|\widetilde{x}\widetilde{v}z} = \mathcal{S}_t^{\xi}$. For each $\tau \in (0, 1)$, assume that the density of $f_{\epsilon_{it}^\tau}(e|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})$ is continuous and bounded in e over $\mathcal{R}$ a.s.. Then the function $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \epsilon_{it})$ is identified if Condition (7) (or Condition (8)) holds for each $\tau \in (0, 1)$.*

In our case, we can transform y_{it} and $(\widetilde{x}_{it}, \widetilde{v}_{it-1})$ to have bounded supports, without loss of generality. Boundedness of $\mathcal{S}_t^y$ implies, under our conditions, that $m(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$ and $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$ are bounded, which in turn implies $\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$ is bounded. The condition $\mathcal{S}_t^{\xi|\widetilde{x}\widetilde{v}z} = \mathcal{S}_t^{\xi}$ is not innocuous. It requires the other determinants of $\widetilde{x}_{it}$ to generate sufficient variation in $\widetilde{x}_{it}$ conditional on $(\widetilde{v}_{it-1}, z_{it}, \xi_{it})$. This condition guarantees that for each $\tau \in (0, 1)$ and for each $(\widetilde{x}, \widetilde{v}) \in \mathcal{S}_t^{\widetilde{x}\widetilde{v}}$, $\bar{f}(\widetilde{x}, \widetilde{v}, \tau) \in \mathcal{S}_t^{y|\widetilde{x}\widetilde{v}z}$ for all z such that $(\widetilde{x}, \widetilde{v}, z) \in \mathcal{S}_t^{\widetilde{x}\widetilde{v}z}$, which implies that for each $\tau \in (0, 1)$, $\bar{f}(., \tau) \in \mathcal{L}_\tau$. Hence, to show identification of $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$, we only need to show that $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$ is the only solution to equation (16) in $\mathcal{L}_\tau$. Continuity and boundedness of $f_{\epsilon_{it}^\tau}(e|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})$ ensures that the integration in the definition of $w_\tau(\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})$ is well defined. Below is the formal proof of Theorem (9).

Proof. For each $\tau \in (0, 1)$, we know $\bar{f}(., \tau)$ solves equation (16). This condition guarantees that for each $\tau \in (0, 1)$ and for each $(\widetilde{x}, \widetilde{v}) \in \mathcal{S}_t^{\widetilde{x}\widetilde{v}}$, $\bar{f}(\widetilde{x}, \widetilde{v}, \tau) \in \mathcal{S}_t^{y|\widetilde{x}\widetilde{v}z}$, for all z such that $(\widetilde{x}, \widetilde{v}, z) \in \mathcal{S}_t^{\widetilde{x}\widetilde{v}z}$, which implies that for each $\tau \in (0, 1)$, $\bar{f}(., \tau) \in \mathcal{L}_\tau$. Hence, for each $\tau \in (0, 1)$, to show identification of $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$, we only need to show $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$ is the only solution to equation (16) in $\mathcal{L}_\tau$. Suppose there is $m(., \tau)$ that solves equation (16) a.s.. Define

[40]The constant $c_\tau(\widetilde{v}^0, z) > 0$ is chosen so that $\varphi_\tau(\widetilde{x}|\widetilde{v}^0, z)$ integrates to one over the support of $\widetilde{x}_{it}$ given $(\widetilde{v}_{it-1}, z_{it}) = (\widetilde{v}^0, z)$.

[41]Note $\varphi_\tau(\widetilde{x}|\widetilde{v}^0, z)$ depends on $\Delta(\widetilde{x}, \widetilde{v}^0, \tau)$, so Condition (8) (or analogously Condition (7)) puts a restriction on all the candidate parameters, i.e., $m(\widetilde{x}, \widetilde{v}, \tau)$'s in the parameter space $\mathcal{L}_\tau$. This sufficient condition for global identification is stronger than that for local identification, which only puts a restriction on the true parameter $\bar{f}(\widetilde{x}, \widetilde{v}, \tau)$. See Chernozhukov et al. (2007).

$\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau) = \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau) - m(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$, and we have

$$\begin{aligned} 0 &= Pr(y_{it} \leq m(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{v}_{it-1}, z_{it}) - Pr(y_{it} \leq \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{v}_{it-1}, z_{it}) \quad (17) \\ &= E(Pr(y_{it} \leq m(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})|\widetilde{v}_{it-1}, z_{it}) \\ &\quad - E(Pr(y_{it} \leq \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})|\widetilde{v}_{it-1}, z_{it}) \\ &= E(Pr(y_{it} - \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau) \leq m(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau) - \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})|\widetilde{v}_{it-1}, z_{it}) \\ &\quad - E(Pr(y_{it} - \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau) \leq 0|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})|\widetilde{v}_{it-1}, z_{it}) \\ &= E(Pr(\epsilon^{\tau}_{it} \leq \Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})|\widetilde{v}_{it-1}, z_{it}) \\ &\quad - E(Pr(\epsilon^{\tau}_{it} \leq 0|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})|\widetilde{v}_{it-1}, z_{it}) \\ &= E(E(\int_0^1 f_{\epsilon^{\tau}_{it}}(\delta\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)d\delta|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})|\widetilde{v}_{it-1}, z_{it}) \\ &= E(\int_0^1 f_{\epsilon^{\tau}_{it}}(\delta\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)d\delta|\widetilde{v}_{it-1}, z_{it}) \\ &= E(\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)w_{\tau}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})|\widetilde{v}_{it-1}, z_{it}). \quad a.s. \end{aligned} \tag{18}$$

The second equality holds by the law of iterated expectation. Note our conditions guarantee for each $(\widetilde{x}, \widetilde{v}) \in \mathcal{S}_t^{\widetilde{x}\widetilde{v}}$, $\bar{f}(\widetilde{x}, \widetilde{v}, \tau)$ and $m(\widetilde{x}, \widetilde{v}, \tau)$ are within $\mathcal{S}_t^{y|\widetilde{x}\widetilde{v}z}$ and $\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$ is within $\mathcal{S}_t^{\epsilon^{\tau}|\widetilde{x}\widetilde{v}z}$, for all z such that $(\widetilde{x}, \widetilde{v}, z) \in \mathcal{S}_t^{\widetilde{x}\widetilde{v}z}$.[42] By continuity and boundedness of $f_{\epsilon^{\tau}_{it}}(e|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})$, the integration after the fifth equality is defined, so the fifth equality holds.

Since Condition (7) holds for each $\tau \in (0, 1)$, for each $\tau \in (0, 1)$ $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)$ is identified. Thus, function $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \xi_{it})$ is identified. ■

With $\bar{f}$ identified, we can establish identification of f under one of our support conditions.

Theorem 10 *If $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \xi_{it})$ is identified, under Assumption (6), $f(x_{it}, \omega_{it})$ is identified.*

Proof. Rewrite the $\mathcal{W}$ operator by replacing $F^{-1}_{y_{it}|\widetilde{v}^0, \widetilde{x}}\left(F_{y_{it}|\widetilde{v}^0, \widetilde{x}^A}\left(y^A\right)\right)$ with $\bar{f}(\widetilde{x}, \widetilde{v}^0, \bar{f}^{-1}(\widetilde{x}^A, \widetilde{v}^0, y^A))$, i.e.

$$\mathcal{W}(\mathcal{S}) = \left\{ \begin{array}{c} (x, y): \text{ for some } (x^A, y^A) \in \mathcal{S} \ \exists v^0 \text{ s.t. } (x^A, y^A, v^0) \in \mathcal{S}_t^{xyv}, x \in \mathcal{S}_t^{x|v^0}, \\ y = \bar{f}(\widetilde{x}, \widetilde{v}^0, \bar{f}^{-1}(\widetilde{x}^A, \widetilde{v}^0, y^A)) \end{array} \right\}.$$ [43]

Note that with our notation, $y = \bar{f}(\widetilde{x}, \widetilde{v}^0, \bar{f}^{-1}(\widetilde{x}^A, \widetilde{v}^0, y^A))$ is equivalent to $y = \bar{f}(x, v^0, \bar{f}^{-1}(x^A, v^0, y^A))$. The Theorem then follows from the proof of Theorem (4). ■

D Monte Carlo

This appendix reports a Monte Carlo exercise designed to evaluate the first-step sieve estimator in a setting where the identifying assumptions hold. This exercise is motivated by the concern,

[42] This ensures the conditional quantile restriction is "binding" in a sense, and rules out the case where both $\bar{f}(\widetilde{x}, \widetilde{v}, \tau)$ and $m(\widetilde{x}, \widetilde{v}, \tau)$ are out of $\mathcal{S}_t^{y|\widetilde{x}\widetilde{v}z}$ for all $(\widetilde{x}, \widetilde{v}, z)$. In that case, it is easy to see equation (17) does not imply $\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau) = 0$ a.s..

[43] This is equivalent to the definition of $\mathcal{W}$ in section (4.2) when x_{it} is independent of ξ_{it}.

emphasized for example by Leon-Ledesma, McAdam, and Willman (2010), that production functions with non-neutral productivity can be difficult to estimate from the levels equation alone. Since the empirical analysis in the paper uses the first-step inverse production function directly, the Monte Carlo focuses on recovery of the first-step objects, rather than on the second-step CES refit.

Data-generating process. The simulated panel has $N = 500$ firms and $T = 10$ observed periods, giving 5,000 firm-year observations in each replication. The production function is a log translog value-added production function with non-Hicks productivity:

$$\begin{aligned} y_{it} = a_k k_{it} + a_\ell \ell_{it} + \frac{1}{2} b_{kk} k_{it}^2 + b_{k\ell} k_{it} \ell_{it} + \frac{1}{2} b_{\ell\ell} \ell_{it}^2 \\ + u_{it} + g_k k_{it} u_{it} + g_\ell \ell_{it} u_{it}. \end{aligned} \tag{19}$$

The coefficient on the standalone productivity term is normalized to one. This normalization makes the simulated inverse production function satisfy the same normalization imposed in the empirical first step, $\Phi(0, 0, y) = y$. The parameter values are

$$a_k = 0.36, \quad a_\ell = 0.64, \quad b_{kk} = 0.020, \quad b_{\ell\ell} = 0.020, \quad b_{k\ell} = -0.120, \quad g_k = 0.020, \quad g_\ell = -0.090.$$

Thus $f_{ku} = g_k > 0$, $f_{\ell u} = g_\ell < 0$, and $RTS_u = g_k + g_\ell < 0$, matching the broad non-Hicks patterns documented in the empirical application. The implied true elasticity of substitution is centered near 1.4.

Productivity and input prices follow persistent AR(1) processes:

$$u_{it} = \rho_u u_{i,t-1} + \xi_{it}, \tag{20}$$

$$v_{it}^K = \rho_v v_{i,t-1}^K + \eta_{it}^K, \tag{21}$$

$$v_{it}^L = \rho_v v_{i,t-1}^L + \eta_{it}^L. \tag{22}$$

The input equations are

$$k_{it} = \rho_k k_{i,t-1} + \lambda_k E[u_{it} \mid u_{i,t-1}] - \chi_k v_{it}^K, \tag{23}$$

$$\ell_{it} = \rho_\ell \ell_{i,t-1} + \lambda_\ell E[u_{it} \mid u_{i,t-1}] + \lambda_{\ell k} k_{it} - \chi_\ell v_{it}^L. \tag{24}$$

The input process deliberately excludes transitory input-choice shocks. Independent variation in inputs comes from persistent input prices, with unconditional standard deviation 0.90. The remaining input-process parameters are

$$\rho_u = 0.70, \quad \rho_v = 0.75, \quad \rho_k = 0.55, \quad \rho_\ell = 0.50, \quad \lambda_k = \lambda_\ell = 0.20, \quad \lambda_{\ell k} = 0.12, \quad \chi_k = \chi_\ell = 0.45.$$

Estimator and evaluation. In each replication, we estimate the same inverse-function first step used in the paper:

$$u_{it} = \Phi(k_{it}, \ell_{it}, y_{it}),$$

with the normalization $\Phi(0,0,y) = y$. The Monte Carlo uses a cubic polynomial for Φ and a quadratic polynomial for the productivity law of motion $h(u_{it}, u_{i,t-1})$. This $\Phi3{+}h2$ specification was chosen because it is flexible enough to recover the relevant first-step objects while avoiding unnecessary second-derivative noise. We then compute from the estimated inverse function

$$\hat{f}_k = -\frac{\Phi_k}{\Phi_y}, \qquad \hat{f}_\ell = -\frac{\Phi_\ell}{\Phi_y}, \qquad \hat{f}_u = \frac{1}{\Phi_y}, \qquad \widehat{RTS} = \hat{f}_k + \hat{f}_\ell.$$

The Hicks elasticity of substitution is computed from the first and second derivatives of the implied log production function:

$$\hat{\sigma} = \frac{\hat{f}_k \hat{f}_\ell (\hat{f}_k + \hat{f}_\ell)}{\hat{f}_k \hat{f}_\ell (\hat{f}_k + \hat{f}_\ell) + 2\hat{f}_k \hat{f}_\ell \hat{f}_{k\ell} - \hat{f}_k^2 \hat{f}_{\ell\ell} - \hat{f}_\ell^2 \hat{f}_{kk}}. \tag{25}$$

The main table uses the full regular sample: observations are not trimmed based on (k, ℓ), output, or productivity. We only require the derived estimated objects to be finite and economically admissible, $\hat{f}_k > 0$, $\hat{f}_\ell > 0$, $\hat{f}_u > 0$, and $\hat{\sigma} > 0$. This keeps 4,991 observations on average out of 5,000. All 100 Monte Carlo replications converge successfully.

Table 3: First-step Monte Carlo recovery: full regular sample

Object	Corr.	RMSE	Mean est.	Mean true	Median est.	Median true	P10 est.	P10 true	P90 est.	P90 true
u	0.998	0.019	0.001	0.001	0.002	0.001	-0.384	-0.384	0.385	0.384
f_k	0.966	0.025	0.359	0.360	0.359	0.360	0.247	0.253	0.469	0.467
f_ℓ	0.974	0.025	0.641	0.640	0.641	0.640	0.519	0.520	0.764	0.760
f_u	0.925	0.027	0.997	1.000	0.998	1.000	0.914	0.920	1.077	1.080
RTS	0.979	0.031	1.000	1.000	1.002	1.000	0.841	0.847	1.155	1.153
σ	0.242	2.015	1.512	1.422	1.418	1.406	1.238	1.331	1.713	1.530

Notes: The table reports averages across 100 Monte Carlo replications. "Corr." is the within-replication correlation between the estimated object and the true object, averaged across replications. The main sample is the full regular sample; no trimming is imposed on (k, ℓ). The regularity restriction only removes observations for which the derived estimated object is non-finite or economically inadmissible.

The first-step estimator recovers productivity, marginal products, and returns to scale very accurately. The distribution of $\hat{f}_k$, $\hat{f}_\ell$, $\hat{f}_u$, and $\widehat{RTS}$ closely tracks the true distribution in both the center and the tails. The elasticity of substitution is more delicate because it is a ratio involving second derivatives. Its median is close to the truth, but the raw mean and RMSE are affected by a small number of observations where the Hicks denominator is close to zero. For this reason, Table 4 reports robust summaries of σ after trimming the most extreme estimated $\hat{\sigma}$ observations within each replication.

Table 4: Robust recovery of the elasticity of substitution

Estimated-σ trim	Mean est.	Mean true	Median est.	Median true	SD est.	SD true	P10 est.	P10 true	P90 est.	P90 true
None	1.512	1.422	1.418	1.406	1.999	0.094	1.238	1.331	1.713	1.530
1% each tail	1.454	1.420	1.418	1.405	0.209	0.085	1.243	1.331	1.697	1.525
2.5% each tail	1.446	1.418	1.418	1.405	0.174	0.081	1.252	1.331	1.676	1.520
5% each tail	1.439	1.416	1.418	1.404	0.147	0.078	1.264	1.332	1.646	1.514

Notes: The table trims the indicated fraction of the estimated $\hat{\sigma}$ distribution within each replication and then averages the resulting replication-level summaries. The true values are recomputed on the same retained observations.

Overall, the Monte Carlo shows that, under a non-Hicks data-generating process with persistent productivity and persistent input-price variation, the first-step sieve estimator recovers the economically relevant objects well from the levels equation. The results are strongest for productivity, marginal products, and returns to scale. The Hicks elasticity of substitution, which depends on second derivatives, is naturally more sensitive to local curvature and near-zero denominators, but its center remains close to the truth under robust summaries.