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

Daniel Ackerberg — UT - Austin
Jinyong Hahn — UCLA
Qingsong Pan* — UT - Austin

November 1, 2022

Abstract

A recent literature addresses endogeneity utilizing assumptions restricting agents' information sets when they chose endogenous variables. We consider using these identifying assumptions to identify a structural function (e.g. a demand or production function) in a fully nonparametric context. Using Imbens and Newey (2009)'s control function framework we show identification and illustrate how our model's structure permits weaker support conditions than used by Imbens and Newey. We apply our results to production function estimation, finding non-Hicks neutral shocks that generate interesting heterogeneity in output elasticities and biased technological change as defined in Acemoglu (2002) and studied in Doraszelski and Jaumandreu (2018).

1 Introduction

In panel data contexts, one often desires to make inferences about the effects of an endogenously chosen variable x_{it} on an outcome variable y_{it}. Since assuming orthogonality between x_{it} and econometric unobservables seems strong, researchers have looked for weaker assumptions on which to base identification and estimation (we loosely interpret orthogonality here to mean either independence, mean independence, or zero correlation, depending on the situation). One general approach is to, instead of assuming that all unobservables are orthogonal to x_{it}, assume that only a portion of the unobservables are orthogonal to x_{it}. The classic linear

* 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.

fixed effects model is perhaps the best known example of this - the unobservable is divided into two components, a time invariant fixed effect component that can be correlated with the x_{it}'s, and a time varying mean zero component that is assumed uncorrelated with x_{it}. The panel data literature, e.g., Chamberlain (1982), Anderson and Hsiao (1982), Arellano and Bond (1991), Blundell and Bond (1998) and Blundell and Bond (2000), contains a number of generalizations of this assumption. For example, one can estimate models under a sequential exogeneity assumption whereby the time varying component of the unobservable is allowed to be correlated with future x_{it}'s. Another example is Blundell and Bond (2000), who allow the time varying component of the unobservable to contain an AR(1) process, where only the innovation in the AR(1) process is assumed uncorrelated with specific x_{it}'s.

A recent literature focused on estimating production functions in a panel context, i.e., Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, Caves, and Frazer (2015), also use this general strategy to address endogeneity issues, but with a different decomposition of the unobservables. Olley and Pakes (1996) 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, Benkard, et al. (2007) describe these as *timing and information set* assumptions, i.e., as 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 $t-1$).

The 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, Levinsohn, and Pakes (1995), Sweeting (2013), Grennan (2013), Lee (2013), and Sullivan (2017), have used timing and information set assumptions to address the problem of endogenously chosen product characteristics and/or prices. Bajari, Fruehwirth, Timmins, 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 across a variety of literatures.

This literature using these Olley and Pakes timing and information set assumptions has 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, Navarro, and Rivers (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 $y_{it} = f_t(x_{it}, \omega_{it})$.

We use a control function approach to show identification of our model, following, e.g., Heckman (1977), Blundell and Smith (1989), Blundell and Powell (2001), Matzkin (2004), and Imbens and Newey (2009). 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 additional 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 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)).

We then 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, Brambilla, and Sasaki (2016), Kasahara, Schrimpf, and Suzuki (2015), Doraszelski and Jaumandreu (2018), Raval (2019), Zhang (2019), Demirer (2020), Raval (2020), Oberfield and Raval (2021)). 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 differential patterns with which how they interact with capital and labor inputs, and, interestingly, these patterns 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. Other recent papers have also relaxed the assumption of Hicks neutral productivity shocks - including some of the papers mentioned above. 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. Hence, we see our approach as complementary to existing approaches. 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}).[1] We

[1] Kasahara, Schrimpf, and Suzuki (2015) also use deconvolution techniques to study identification of a production function with extensive time invariant unobserved heterogeneity, but where the productivity shock is

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,[2] and ω_{it} is a scalar unobservable term that is also continuously distributed.[3]

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. Note that with auxiliary data, one could potentially add additional unobservables to the model that are identified in a preliminary stage. For example, in a production function context Ackerberg, Caves, and Frazer (2015) show how, with additional assumptions and data m_{it}, one can identify ϵ_{it} in the model $\widetilde{y}_{it} = f_t (x_{it}, \omega_{it}) + \epsilon_{it}$ in a preliminary stage, hence reducing the model to the one above, i.e., $y_{it} = \widetilde{y}_{it} - \epsilon_{it} = f_t (x_{it}, \omega_{it})$.[4] 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 endogenously chosen by an economic agent. We start with our key *timing and information set* assumption:

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

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

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

[4]They actually consider the model $\widetilde{y_{it}} = f_t (x_{it}) + \omega_{it} + \epsilon_{it}$, but the process would be the same with ω_{it} entering non-linearly.

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.*

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).[5] 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.

A good 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, Navarro, and Rivers (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)[6] 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.[7] In the production function context, the unobservable η_{it-1} could represent a multidimensional set of

[5] We will discuss how one might relax the timing and information set assumptions in section (5).

[6] 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).

[7] 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".

factors affecting input and output prices (or those prices themselves if they are competitively set). Typically, such factors will impact optimal choices of x_{it}.[8]

These timing and information set assumptions have also been used for identification in demand models with endogenous product characteristics, e.g., Sweeting (2013), Grennan (2013), Lee (2013), and Sullivan (2017). These papers 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 assume that while period t product characteristics x_{it} can be chosen as a function of prior periods demand shocks ω_{it-1}, they cannot be chosen as a function of the current period demand shock ω_{it}. In this case, η_{it-1} might represent cost shocks that affect firms' choices of product characteristics and prices. Other applications using these types of assumptions include Bajari, Fruehwirth, Timmins, 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}.

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

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

Assumption (2) allows the distribution of ω_{it} 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.[9] We also assume that ω_{it} follows a finite Mth order Markov process.[10] So unlike Arellano and Bond (1991), Blundell and Bond (1998), Blundell and Bond (2000), and Altonji and Matzkin (2005), our assumption does not allow there to be a component of ω_{it} that is fixed over time (e.g. a fixed or random effect). On the other hand, we do not require the 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

[8]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 $\mathcal{I}_{it-1}$.

[9]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.

[10]When $M = 0$, we define $\{\omega_{i\tau}\}_{\tau=t-M}^{t-1} = \emptyset$. Obviously this is not a particularly interesting case, because in this case, our assumptions imply that ω_{it} is independent of x_{it}, and identification of f_t is trivial using Matzkin (2003).

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

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

where 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 and strictly increasing in ξ_{it}.*

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, ω_{it} and ξ_{it} have $U(0,1)$ marginal distributions.*

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 $\varsigma_t = \left(\varsigma_t^1, \ldots, \varsigma_t^J\right)$.

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 $(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1})$. Now 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 $x_{it} = \varphi_t\left(\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right), \varsigma_{it}\right)$. 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 $\omega_{it} = \phi_t\left(\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right), \xi_{it}\right)$. ∎

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:

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

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. $Pr(y_{it} \leq y^0 | x_{it} = x^0, v_{it-1} = v^0)$ 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., 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}.[11]

4 Relaxing Support Conditions

As discussed by Imbens and Newey (2009), the support condition in Assumption (5) might be considered strong. As a result, they investigate bounds on objects of interest when the condition does not hold. The informativeness of these bounds can vary widely depending on the model and the object of interest. In contrast, the additional structure of our specific model allows us to significantly relax this 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 - 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

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

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.

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. For the rest of the paper, we assume ω_{it} follows a first order Markov process for notational convenience, but our results can be generalized to higher order Markov processes.

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.

Assumption 6 *(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 Assumption (6) 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}. Denoting by $f_t^{-1}(x_{it}, y_{it})$ the inverse of $f_t(x_{it}, \omega_{it})$ w.r.t. its second argument, we have

Theorem 2 *If the density of (x_{it}, v_{it-1}) satisfies Assumption (6) 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} and substituting ω_{it-1} with $f_{t-1}^{-1}(x_{it-1}, y_{it-1})$, we have

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

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.[12] 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.

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 $f_t^{-1}(x^A, y^A) = f_t^{-1}(x^B, y^B)$).*

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

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 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).[13] 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

Assumption 7 *There is a x^0 such that for any $v\in\mathcal{S}_t^v$, $x^0\in\mathcal{S}_t^{x|v}$.*

Assumption (7) weakens Imbens and Newey's Assumption (5). While Imbens and Newey require v_{it-1} to have full support conditional on any x, Assumption (7) only requires v_{it-1} to have full support at *one particular* x^0.

Assumption (7) 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. Specifically, it means we can find two points (x^0,y^{0A}) and (x^0,y^{0B}) that are "iso-omegic" to the original points, i.e., such that $f_t^{-1}\left(x^0,y^{0A}\right)=f_t^{-1}(x^A,y^A)$ and $f_t^{-1}\left(x^0,y^{0B}\right)=f_t^{-1}(x^B,y^B)$. It assures we can find these iso-omegic points at x^0 since for any v^A and v^B s.t. (x^A,y^A,v^A) and (x^B,y^B,v^B) are in $\mathcal{S}_t^{xyv}$, Assumption (7) ensures (x^0,v^A) and (x^0,v^B) are

[13] 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 $\succsim$.

in $\mathcal{S}_t^{xv}$. This means that we can use objects identified from the data, the conditional CDF $F_{y_{it}|x_{it},v_{it}}$ and its inverse $F^{-1}_{y_{it}|x_{it},v_{it}}$ to "translate" the implied ξ^A and ξ^B at (x^A, y^A, v^A) and (x^B, y^B, v^B) to (x^0, v^A) and (x^0, v^B) to determine the iso-omegic points $\left(x^0, y^{0A}\right)$ and $\left(x^0, y^{0B}\right)$, i.e.,

$$y^{0A} = F^{-1}_{y_{it}|x^0,v^A}\left(F_{y_{it}|x^A,v^A}\left(y^A\right)\right) \qquad \text{and} \qquad y^{0B} = F^{-1}_{y_{it}|x^0,v^B}\left(F_{y_{it}|x^B,v^B}\left(y^B\right)\right).[14]$$

Then, since f_t^{-1} is strictly monotone in its second argument, whether $f_t^{-1}(x^A, y^A) > f_t^{-1}(x^B, y^B)$ depends on whether $y^{0A} > y^{0B}$. Clearly, the ability to do this again depends crucially on the scalar ω_{it} in our model.

We can relax the support condition further with the following:

Assumption 8 *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 Assumption (7), Assumption (8) 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 examples of $\mathcal{S}_t^{xv}$ where Assumption (8) is satisfied but not Assumption (7). 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

Assumption 9 *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 $(x^{AB}, v^A), (x^{AB}, v^B) \in S^{xv}$*

This 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 Assumption (9) is dependent on S^{xvy}. This differs from Assumptions (5),(7), and (8) which only put restrictions on S^{xv}. However, Condition (9) is implied by the prior conditions and hence weaker.[15]

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:

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

[14] 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) $\xi_{it} \sim U(0, 1)$.

[15] Assumption (9) is implied by Assumption (8), 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 Assumption (8) assures these v^A and v^B have a common x support point.

$\mathcal{W}(x^A, y^A)$ 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, 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

Assumption 10 *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 $N \in \mathbb{N}$).*

Assumption (10) further weakens Assumption (8) and is also sufficient to order any $f_t^{-1}(x^A, y^A)$ and $f_t^{-1}(x^B, y^B)$. Intuitively, Assumption (10) 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 Assumption (10) 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 Assumption (10) 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:

Assumption 11 *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 $(x^0, y^0) = (x^A, y^A), (x^{J+1}, y^{J+1}) = (x^B, y^B)$.*

Condition (11) is weaker than Assumption (10) since Assumption (10) implies that Assumption (11) holds for all (x^A, y^A) and (x^B, y^B) with $J = 0$, i.e., no intermediate points are necessary. Assumption (11) may be helpful in relaxing Assumption (10), 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., Assumption (10) to hold, but Assumption (11) 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 Assumption (11) 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 3 *Under the assumptions of our model and Assumption (11), $f_t(x, \omega)$ is identified for all $t > M$.*

Proof. See Appendix B. ■

Theorem (3) clearly also implies that $f_t(x, \omega)$ is identified under any of the stronger support Assumptions (7), (8), (9), and (10). This is useful since the former conditions, while stronger, may be more economically interpretable. We can also consider other types of support conditions on $\mathcal{S}_t^{xvy}$ that are sufficient for identification. Assumptions (7) and (8) 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 condition on the primitives of our model that helps do this is the following:

Assumption 12 *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 $\{\omega_{i\tau}\}_{\tau=t-M}^{t-1}$.*

With our normalization, Assumption (12) 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 above discussion, this means that any point (x^A, y^A) is consistent with any $v \in \mathcal{S}_t^v$, i.e., $(x^A, v, y^A) \in \mathcal{S}_t^{xvy}$ for all $v \in \mathcal{S}_t^v$. 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. We also utilize the following regularity conditions. Let $Int(\mathcal{S})$ denote the interior of set $\mathcal{S}$.

Assumption 13 *The boundary of $\mathcal{S}_t^{xy}$ has probability measure zero. For any $x_{it} \in Int(\mathcal{S}_t^x)$ there exists a v_{it} such that $(x_{it}, v_{it}) \in Int(\mathcal{S}_t^{xv})$.*

These make the proof of the following theorem easier as we can work with open sets.

Theorem 4 *Under the assumptions of our model, assumption 12, and assumption 13, then if $\mathcal{S}_t^{xv}$ is convex, $f_t(x, \omega)$ is identified for all $t > M$.*

Proof. See Appendix B ■

Theorem (4) 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 (12) 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 (12)), 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 of it additionally requires Assumption (12). We lastly note that under Assumption (12), 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.

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. Assumptions (9), (10), and (11) 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 (9), (10), or (11) hold over some subset $\widetilde{\mathcal{S}}_t^{xy} \subseteq \mathcal{S}_t^{xy}$. We can show

Theorem 5 *Under the assumptions of our model, if Assumption (11) 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 $f_t(x, \omega)$ is identified for all $t > M$ on $\widetilde{\mathcal{S}}_t^{xy}$ under the normalization that $\omega_{it} \sim U(0, 1)$ on $\widetilde{\mathcal{S}}_t^{xy}$.*[16]

[16] To be more precise, maybe we should say $f_t^{-1}(x, \omega)$ is identified on $\widetilde{\mathcal{S}}_t^{xy}$ and the normalization is made on the distribution of ω_{it} that is implied by the distribution of (x_{it}, y_{it}) on $\widetilde{\mathcal{S}}_t^{xy}$.

Proof. Identical to the proof of Theorem 3, with the normalization and identification only on the set $\widetilde{\mathcal{S}}_t^{xy}$. ■

The intuition behind Theorem (5) is that if Assumption (11) (or Assumption (9) or (10)) 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 (5) 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 innoculus, 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 A Nonidentification Result

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, Navarro, and Rivers (2020).

Ideally, one might like to relax this assumption, i.e. allow x_{it} to be chosen as a function of ω_{it}. However, we now show that this is not possible without further assumptions. We briefly illustrate with a simple counterexample based on the nonidentification result from Ackerberg, Frazer, et al. (2020). Suppose for $T = 2$, and we have the following data generating process

(DGP):

$$\begin{aligned} x_{i1} &= \theta_1 + \theta_2\omega_{i1} + \eta_{i1}, \\ y_{i1} &= \theta_3 + \theta_4 x_{i1} + \omega_{i1}, \\ x_{i2} &= \theta_5 + \theta_6\omega_{i2} + \eta_{i2}, \\ y_{i2} &= \theta_7 + \theta_8 x_{i2} + \omega_{i2}, \end{aligned} \tag{5}$$

where

$$\begin{aligned} \eta_{i2} &= \rho_x\eta_{i1} + u_{i2} \qquad \text{with } \eta_{i1} \sim N(0,\sigma_1), u_{i2} \sim N(0,\sigma_2), \\ \omega_{i2} &= \rho_\omega\omega_{i1} + \xi_{i2} \qquad \text{with } \omega_{i1} \sim N(0,\sigma_3), \xi_{i2} \sim N(0,\sigma_4), \end{aligned}$$

and $(\kappa_{i1}, u_{i2}, \omega_{i1}, \xi_{i2})$ are mutually independent. Observe that in the model (5) x_{i1} (x_{i2}) is set as a function of ω_{i1} (ω_{i2}), i.e., x_{it} does not satisfy the timing and information set assumption.

Solving out for the observables in terms of the primitive shocks κ_{i1}, u_{i2}, ω_{i1}, ξ_{i2}, we obtain

$$\begin{bmatrix} x_{i1} \\ y_{i1} \\ x_{i2} \\ y_{i2} \end{bmatrix} = \begin{bmatrix} \theta_1 \\ (\theta_3 + \theta_4\theta_1) \\ \theta_5 \\ (\theta_7 + \theta_8\theta_5) \end{bmatrix} + \begin{bmatrix} \theta_2 & 1 & 0 & 0 \\ (\theta_4\theta_2 + 1) & \theta_4 & 0 & 0 \\ \theta_6\rho_\omega & \rho_x & \theta_6 & 1 \\ (\theta_8\theta_6 + 1)\rho_\omega & \theta_8\rho_x & (\theta_8\theta_6 + 1) & \theta_8 \end{bmatrix} \begin{bmatrix} \omega_{i1} \\ \eta_{i1} \\ \xi_{i2} \\ u_{i2} \end{bmatrix} = \mu + \Sigma \begin{bmatrix} \omega_{i1} \\ \eta_{i1} \\ \xi_{i2} \\ u_{i2} \end{bmatrix}$$

so, for example, if $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, the observed data $(x_{i0}, y_{i0}, x_{i1}, y_{i1}) \sim N(\mu, \Sigma\Sigma')$. Following Ackerberg, Frazer, et al. (2020), one can easily construct examples of nonidentification in this setup. For example, if true DGP is where all the θ's $= 1$, $\rho_\omega = 0.7$, and $\rho_x = 0.5$, then

$$\mu = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0.7 & 0.5 & 1 & 1 \\ 1.4 & 0.5 & 2 & 1 \end{bmatrix}. \tag{6}$$

But at alternative parameter values $\theta_1 = 1, \theta_2 = -1, \theta_3 = 0, \theta_4 = 2, \theta_5 = 1, \theta_6 = -1$, $\theta_7 = 0$, $\theta_8 = 2, \rho_\omega = 0.5, \rho_x = 0.7$,

$$\widetilde{\mu} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \quad \widetilde{\Sigma} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ -0.5 & 0.7 & -1 & 1 \\ -0.5 & 1.4 & -1 & 2 \end{bmatrix}. \tag{7}$$

Since the columns of $\widetilde{\Sigma}$ relative to Σ are just swapped and/or multiplied by -1, one can easily verify that $\Sigma\Sigma' = \widetilde{\Sigma}\widetilde{\Sigma}'$, i.e., these two parameter values generate the exact same distribution of the data. Ackerberg, Frazer, et al. (2020) show that, in a linear model like this, this finite underidentification problem (see also Sentana (2015)) arises generally where ρ_ω and ρ_x are switched, and the signs of θ_2 and θ_6 are negated.

5.2 Additional Restrictions

The above example implies that when the timing and information set is relaxed in this way, our nonparametric model is also not identified without further restrictions.[17] There are a few existing alternatives for such further restrictions. For example, in the parametric linear case, Ackerberg, Frazer, 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, Navarro, and Rivers (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 accomodate a material input that depends on ω_{it}.

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 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.

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

$$\begin{aligned} x_{it}^V &= \psi_t\left(x_{it}^F, \omega_{it}, z_{it}, \eta_{it}\right) \\ &= \psi_t\left(x_{it}^F, g_t\left(\omega_{it-1}, \xi_{it}\right), z_{it}, \eta_{it}\right) \end{aligned} \tag{8}$$

Note how this model of x_{it}^V corresponds to a variable, non-dynamic (see Ackerberg, Benkard, 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

[17] 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., according to $x_{it} = h_t\left(I_{it-2}\right)$. 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).

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}.[18] Appendix (C) shows that given this setup, we can use the framework of Chernozhukov and Hansen (2005) to identify f_t.[19] Intuitively, this works by again considering the following reduced form expression

$$\begin{aligned} y_{it} &= f_t(x_{it}^F, x_{it}^V, \omega_{it}) \\ &= f_t(x_{it}^F, x_{it}^V, g_t(f_{t-1}^{-1}(x_{it-1}, y_{it-1}), \xi_{it})) \\ &= f_t(x_{it}^F, x_{it}^V, g_t(f_{t-1}^{-1}(v_{it-1}), \xi_{it})) \\ &= \bar{f}_t(x_{it}^F, x_{it}^V, v_{it-1}, \xi_{it}). \end{aligned} \tag{9}$$

and, given independence of $(x_{it}^F, v_{it-1}, z_{it})$ 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 (7), (8), (9), (10), and (11)) 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

[18] 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.

[19] 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 (8), e.g. that it is scalar, and additional independence conditions.

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.[20] 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.

We show this for f_1 in the first order Markov case, but our approach can be straightforwardly extended to identify f_t in the first M periods in the Mth order Markov case. The intuition of the approach can be illustrated in a simple linear model. Suppose $T = 2$ and that

$$y_{i1} = \theta_1 + \theta_2 x_{i1} + \omega_{i1} \tag{10}$$

$$y_{i2} = \theta_3 + \theta_4 x_{i2} + \omega_{i2} \tag{11}$$

and

$$\omega_{i2} = \rho\omega_{i1} + \xi_{i2} \tag{12}$$

Our goal is to identify θ_2 (θ_4 can be identified with the control function method). To do this, substitute the inverted (10) into (12), and this into (11) to get

$$y_{i2} = (\theta_3 - \rho\theta_1) + \theta_4 x_{i2} + \rho y_{i1} + \rho\theta_2 x_{i1} + \xi_{i2} \tag{13}$$

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 $\rho > 0$, we can simply run, e.g., OLS on (13) 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 14 *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 15 $\frac{\partial g_t(\omega_{t-1}^0, \xi_t^0)}{\partial \omega_{it-1}}$ *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 $\xi_t^0 \in (0,1)$.*

[20] 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).

The local support Assumption (14) is needed to identify derivatives and is very similar to Assumption (6) - 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 (15) 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 (12).

Theorem 6 *If the model satisfies Assumptions (14) and (15) 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 (6) 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

$$q_{y_{it}|x_{it},v_{it-1}}(\xi_t^0|x_t^0, x_{t-1}^0, y_{t-1}^0) = f_t(x_t^0, g(f_{t-1}^{-1}(x_{t-1}^0, y_{t-1}^0), \xi_t^0)) \tag{14}$$

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} in response to x_{it} to identify derivatives of f_t, while Theorem (6) uses variation in y_{it} in response to 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, Liao, and Ridder (2021) examine other implications of the single index structure of these types of models.

7 Application to Production Function Estimation

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, Brambilla, and Sasaki (2016), Fox et al. (2017), Kasahara, Nishida, and Suzuki (2017), Doraszelski and Jaumandreu (2018), Raval (2019), Zhang (2019), Demirer (2020), Raval (2020), and Oberfield and Raval (2021), 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}.

Since these two approaches to relaxing Hicks neutrality are clearly non-nested, we believe they are complementary to each other, especially to the extent that they find evidence of similar economic phenomena. 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.[21]

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 different aspects of this Hicksian non-neutrality - its implication on production function

[21] 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.

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.

7.1 Data and Model

We use the same Chilean and Colombian data sets as do Levinsohn and Petrin (2003), Gandhi, Navarro, and Rivers (2020) and many others, and focus on 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.

The empirical work in Levinsohn and Petrin (2003) assumes a Cobb-Douglas production function and a Hicks neutral productivity shock, while Gandhi, Navarro, and Rivers (2020) identify a nonparametric 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, Navarro, and Rivers (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).

We estimate our nonparametric model using a sieve maximum likelihood strategy based on polynomial approximations. Specifically, we specify the nonseparable production function f_t as

$$\begin{aligned} y_{it} = \beta_0 + (\beta_k + \sigma_k u_{it})k_{it} + (\beta_l + \sigma_l u_{it})l_{it} + (\beta_{kk} + \sigma_{kk} u_{it})k_{it}^2 + (\beta_{kl} + \sigma_{kl} u_{it})k_{it}l_{it} \\ + (\beta_{ll} + \sigma_{ll} u_{it})l_{it}^2 + u_{it}, \end{aligned} \tag{15}$$

where the productivity term $u_{it} = \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \omega_{it}$ and the unobserved component ω_{it} follows the following first order Markov process g:

$$\omega_{it} = \rho_1 \omega_{it-1} + \rho_2 \omega_{it-1}^2 + \rho_3 \omega_{it-1}^3 + \xi_{it}. \tag{16}$$

We approximate the distribution of the innovation term ξ_{it} with a mixture of two normal distributions.[22] This results in a model with a total of 21 parameters to be estimated.[23] Note that the model does not satisfy our strict monotonicity assumption for all values of the sieve parameters. Strict monotonicity requires that

$$1 + \sigma_k k_{it} + \sigma_l l_{it} + \sigma_{kk} k_{it}^2 + \sigma_{kl} k_{it} l_{it} + \sigma_{ll} l_{it}^2 > 0 \quad \forall i, t.$$

However, in our estimation routines, we did not have problems with our non-linear searches ending up in problematic parts of the parameter space. Hence, we were able to estimate the models without formally enforcing these restrictions on the parameters, and our final estimates are such that the strict monotonicity assumption holds (and is not binding) for all i and t.

For estimation we maximize the following partial log likelihood function:

$$\sum_{i=1}^{I} \sum_{t=2}^{T} \ln \left(P(y_{it} | k_{it}, l_{it}, k_{it-1}, l_{it-1}, y_{it-1}; \theta) \right).$$

Conditioning on $(k_{it-1}, l_{it-1}, y_{it-1})$ in our model is equivalent to conditioning on ω_{it}. Hence, given parameters and $(k_{it}, l_{it}, k_{it-1}, l_{it-1}, y_{it-1})$, the distribution of y_{it} is determined by only ξ_{it}, which is independent of $(k_{it}, l_{it}, k_{it-1}, l_{it-1}, y_{it-1})$. Thus, the conditional density $P(y_{it} | k_{it}, l_{it}, k_{it-1}, l_{it-1}, y_{it-1}; \theta)$ is easy to calculate at each data point - given a guess at parameters, the implied ω_{it}'s can be calculated by inverting (15), and the implied ξ_{it} can be calculated using (16).

The above likelihood is a partial likelihood because it only considers the conditional density of y_{it}. In our model k_{it} and l_{it} are endogenous variables, determined as a potential function of past productivity shocks (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 k_{it} and l_{it}, which is an advantage since these inputs may be chosen as part of compex dynamic, optimization problems that depend on many other factors and parameters (Olley and Pakes (1996) and much of the related literature, which tend to instead use GMM for estimation, share this advantage). Our partial maximum likelihood estimator is consistent, and satisfies the property of "dynamic completeness" discussed by Wooldridge (2010). For inference we follow Newey (1994) and Ackerberg, Chen, and Hahn (2012) and calculate standard errors under the presumption that (15) and (16) constitutes a parametric model. They show that such calculations often produce consistent standard errors

[22] As noted in our theoretical results, if we were completely flexible with f and g we could normalize $\xi \sim U(0,1)$. But since in our finite sample we are not being completely flexible, our modelling of ξ as a mixture of normals as a way of adding flexibility to our model in a way that is fairly easily interpretable.

[23] Note that this formulation make it quite easy to invert u_{it} and construct our likelihood function. If one wanted to allow u_{it} to enter more flexibly, one could either enforce restictions that impose monoticity, or one could directly estimate the inverse function of the production function. In the latter case, the function value is the unobserved productivity shock (up to a monotone transformation), and the output elasticities can be calculated using the implict function theorem.

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.

7.2 Output Elasticities

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.

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. 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 thus overestimate labor market power.[24] A model with a Hicks neutral productivity shock implies that the shock has no impact on output elasticities, i.e. all the heterogeneity in output elasticities is generated by different levels of labor and capital. 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

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

non-Hicks model. 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. If we compare the coefficients of variation of EL and EK in the first two columns, in most industries, the productivity shock generates slightly more heterogeneity in EL than in 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 and coefficient of variation) in this column comes from variation in both the productivity shock *and* in input levels across firms. A first observation is that, while in the first two columns, the coefficients of variation tend to be somewhat higher for EL, in column three it reverses, i.e. the coefficient of variation tends to be higher for EK. In other words, much of the heterogeneity in EK is driven by across firm variation in the observed input levels. In a sense, 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	0.53	0.96	0.93	0.82	0.96	0.92	0.80
	(0.01)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.03)	(0.03)
Capital	0.35	0.35	0.39	0.21	0.21	0.27	0.25	0.26	0.29
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Sum	1.22	1.12	0.92	1.17	1.15	1.09	1.21	1.18	1.10
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)
Mean (capital)/mean (labor)	0.40	0.46	0.74	0.22	0.23	0.32	0.26	0.29	0.37
	(0.01)	(0.02)	(0.04)	(0.02)	(0.03)	(0.04)	(0.01)	(0.02)	(0.03)
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	0.57	0.92	0.87	0.50
	(0.01)	(0.04)	(0.04)	(0.01)	(0.03)	(0.04)	(0.02)	(0.04)	(0.05)
Capital	0.36	0.34	0.40	0.13	0.14	0.22	0.27	0.29	0.22
	(0.01)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)	(0.03)	(0.04)
Sum	1.09	1.06	0.93	1.04	1.00	0.79	1.19	1.16	0.72
	(0.01)	(0.02)	(0.05)	(0.01)	(0.02)	(0.04)	(0.01)	(0.04)	(0.07)
Mean (capital)/mean (labor)	0.49	0.47	0.76	0.15	0.17	0.38	0.30	0.34	0.44
	(0.02)	(0.05)	(0.07)	(0.01)	(0.03)	(0.05)	(0.02)	(0.04)	(0.08)

NOTE.—In the parenthesis are bootstrap standard errors (200 replications). 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 specified above, where we add a "random coefficient" aspect to a translog production function, and they are estimated using the MLE procedure described above. 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)	0.52	0.54	0.53	0.81	0.82	0.82	0.80	0.81	0.80
	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
SD(EL)	0.10	0.10	0.14	0.08	0.09	0.10	0.08	0.11	0.16
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
CV(EL)	0.18	0.18	0.27	0.10	0.11	0.13	0.11	0.13	0.19
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
Mean(EK)	0.39	0.36	0.39	0.27	0.26	0.27	0.30	0.29	0.29
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
SD(EK)	0.05	0.05	0.19	0.03	0.03	0.10	0.02	0.03	0.08
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
CV(EK)	0.13	0.14	0.49	0.12	0.13	0.37	0.07	0.12	0.29
	(0.01)	(0.01)	(0.03)	(0.04)	(0.05)	(0.07)	(0.03)	(0.03)	(0.04)
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)	0.51	0.53	0.53	0.53	0.54	0.57	0.47	0.47	0.50
	(0.04)	(0.04)	(0.01)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.05)
SD(EL)	0.14	0.16	0.20	0.12	0.12	0.17	0.17	0.18	0.21
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
CV(EL)	0.27	0.30	0.37	0.22	0.23	0.30	0.36	0.39	0.41
	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.05)	(0.10)	(0.11)	(0.08)
Mean(EK)	0.41	0.41	0.40	0.23	0.22	0.22	0.20	0.20	0.22
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.04)
SD(EK)	0.08	0.09	0.19	0.04	0.04	0.11	0.06	0.06	0.13
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
CV(EK)	0.18	0.21	0.49	0.18	0.19	0.53	0.30	0.31	0.57
	(0.03)	(0.03)	(0.03)	(0.05)	(0.06)	(0.10)	(0.11)	(0.11)	(0.10)

Note.—In the parenthesis are bootstrap standard errors (200 replications). The numbers are based on our nonseparable model specified above. The numbers in the first two column 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.

7.3 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.[25] So it is interesting to assess the same question with our alternative non-Hicks neutral model. Following Acemoglu (2002), we first formally define the notion of "biased technological change".

Definition 7 *Technological change is biased toward input x_1 over x_2 if*

$$\frac{\partial \frac{\partial F(\mathbf{x},u)/\partial x_1}{\partial F(\mathbf{x},u)/\partial x_2}}{\partial u} \geq 0, \tag{17}$$

i.e., if an increase in the productivity shock u 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 figures (1) and (2). In figure (1), we graph the ratio of the marginal productivity of capital (MPK) to the marginal productivity of labor (MPL) as a function of productivity shock (u), holding capital and labor constant at their median levels.[26] 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 - in most cases, the ratio more than doubles when productivity shock moves from the 10th percentile (the left end) to the 90th percentile (the right end). Figure (2) examines the effect in a slightly different way. Holding capital and labor at the median levels, we graph both MPK and MPL as a function of the productivity shock - normalizing the starting point for both to one to account for different units of measurement. For all the industries, MPK increases above

[25] 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).

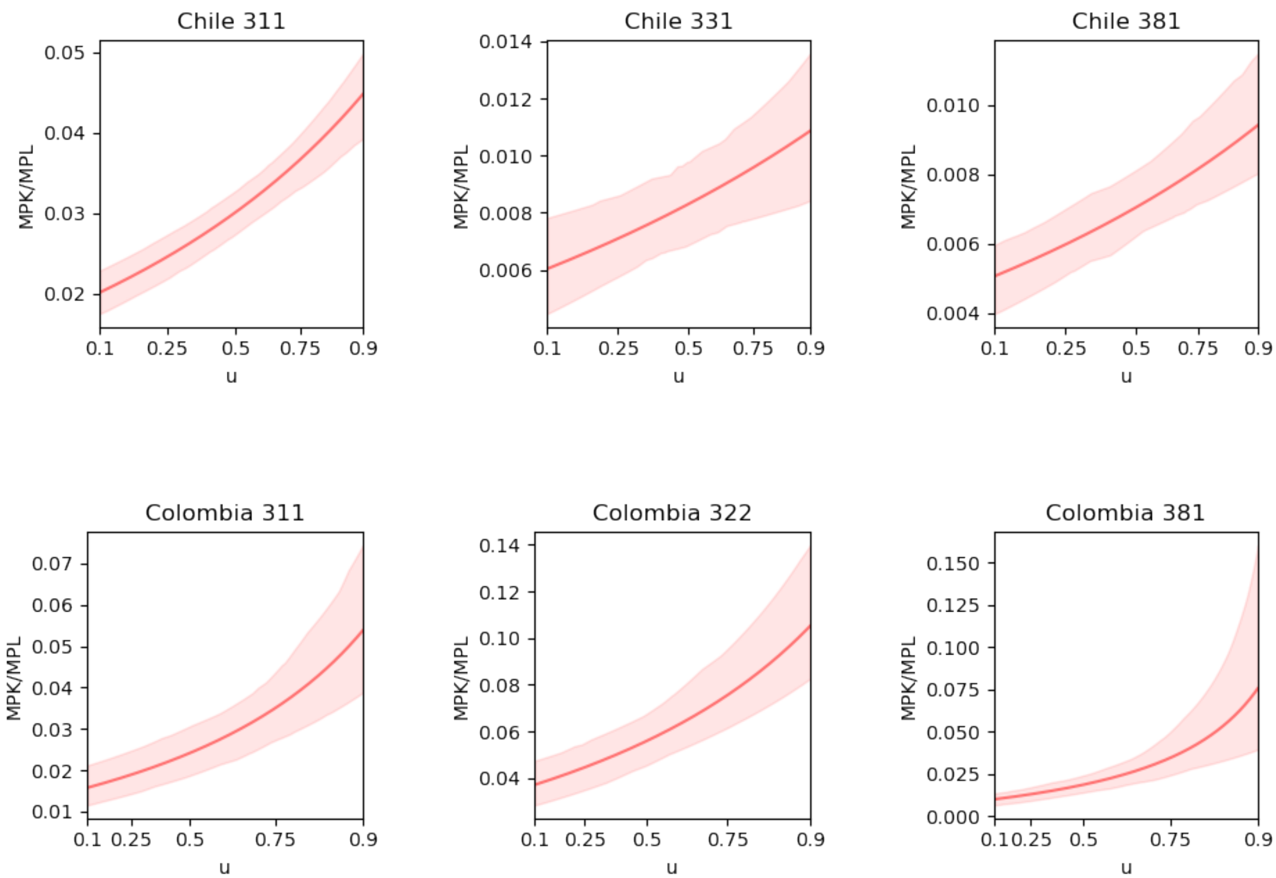
[26] We find similar patterns when holding capital and labor at other representative values, e.g., 25th and 75th percentiles.

MPL as u increases, with confidence intervals at most barely overlapping, suggesting again that technological change in these industries is biased toward capital.

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), labor-saving. Our finding of capital-biased technological change under a nonseparable model 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.

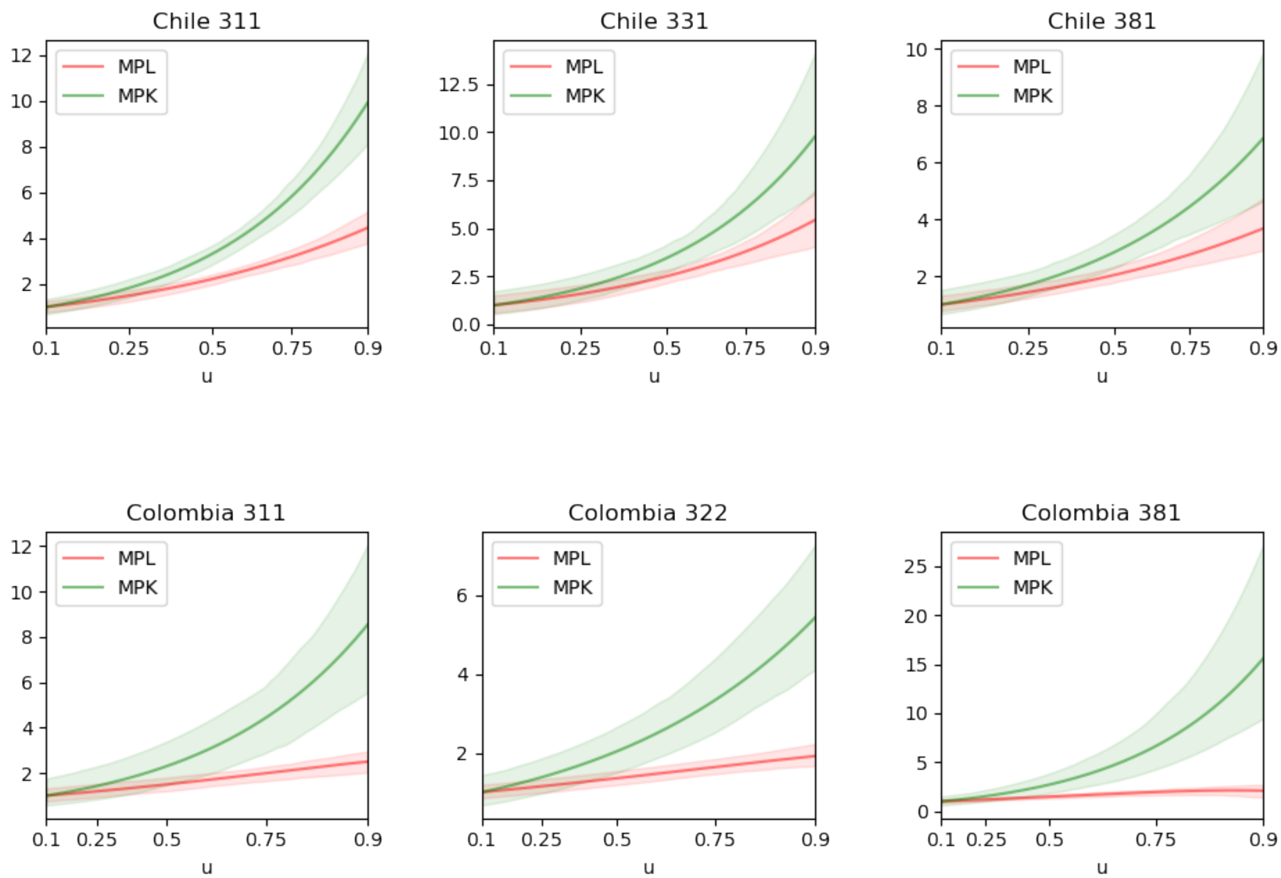
In sum, we believe our results suggesting capital-biased technological change across multiple industries in two countries are interesting in relation to the recent literature concerning factor-augmenting productivity shocks, e.g., Doraszelski and Jaumandreu (2018), Raval (2019), Zhang (2019), and Oberfield and Raval (2021). As described above, one major difference is that they 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 allow labor to have dynamic effects. Unlike these other papers, we also do not need to observe measures of input prices, which can be hard to obtain (in our model, any firm specific input costs are in the unobservables η). In addition, they need to assume labor markets are competitive, but we can allow firms to have monopsony power in labor markets. Given the distinctiveness of the assumptions, we hope the two approaches are complementary - empirical conclusions robust to both approaches and sets of assumptions would seem to be more convincing than those using only one. So our finding of capital-biased technological change is supportive of the findings of the factor-augmenting literature.

Figure 1: Bias of Technological Change



NOTE.—The graphs show the ratios of marginal productivity of capital (MPK) to marginal productivity of labor (MPL), as functions of the productivity shock (u). The shaded areas are 95% confidence bands. The ticks on the horizontal axis are the quantiles of the distributions of u_{it}.

Figure 2: MPK vs MPL



NOTE.—The graphs show marginal productivity of labor (MPL) and capital (MPK) as functions of the productivity shock (u). The shaded areas are 95% confidence bands. The ticks on the horizontal axis are the quantiles of the distributions of u_{it}.

8 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

[1] Daron Acemoglu. "Directed technical change". In: *The review of economic studies* 69.4 (2002), pp. 781–809.

[2] Daniel Ackerberg. *Timing Assumptions and Efficiency: Empirical Evidence in a Production Function Context*. 2020. URL: https://drive.google.com/file/d/1RyDQt_o6nLGkI22DsyxbLLqljLxErsfr/view (visited on 03/21/2022).

[3] Daniel Ackerberg, C Lanier Benkard, et al. "Econometric tools for analyzing market outcomes". In: *Handbook of econometrics* 6 (2007), pp. 4171–4276.

[4] Daniel Ackerberg, Kevin Caves, and Garth Frazer. "Identification properties of recent production function estimators". In: *Econometrica* 83.6 (2015), pp. 2411–2451.

[5] Daniel Ackerberg, Xiaohong Chen, and Jinyong Hahn. "A practical asymptotic variance estimator for two-step semiparametric estimators". In: *Review of Economics and Statistics* 94.2 (2012), pp. 481–498.

[6] Daniel Ackerberg, Garth Frazer, et al. "Under-Identification of Structural Models Based on Timing and Information Set Assumptions". In: (2020).

[7] Daniel Ackerberg and Jinyong Hahn. "Some non-parametric identification results using timing and information set assumptions". In: *Unpublished manuscript* (2015).

[8] Joseph Altonji and Rosa Matzkin. "Cross section and panel data estimators for nonseparable models with endogenous regressors". In: *Econometrica* 73.4 (2005), pp. 1053–1102.

[9] Theodore Wilbur Anderson and Cheng Hsiao. "Formulation and estimation of dynamic models using panel data". In: *Journal of econometrics* 18.1 (1982), pp. 47–82.

[10] Manuel Arellano and Stephen Bond. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations". In: *The review of economic studies* 58.2 (1991), pp. 277–297.

[11] José Azar, Steven Berry, and Ioana Elena Marinescu. "Estimating labor market power". In: *Available at SSRN 3456277* (2019).

[12] Patrick Bajari, Jane Cooley Fruehwirth, Christopher Timmins, et al. "A rational expectations approach to hedonic price regressions with time-varying unobserved product attributes: The price of pollution". In: *American Economic Review* 102.5 (2012), pp. 1898–1926.

[13] Jorge Balat, Irene Brambilla, and Yuya Sasaki. *Heterogeneous firms: Skilled-labor productivity and the destination of exports*. Tech. rep. Working paper, 2016.

[14] Steven Berry, James Levinsohn, and Ariel Pakes. "Automobile prices in market equilibrium". In: *Econometrica: Journal of the Econometric Society* (1995), pp. 841–890.

[15] Richard Blundell and Stephen Bond. "GMM estimation with persistent panel data: an application to production functions". In: *Econometric reviews* 19.3 (2000), pp. 321–340.

[16] Richard Blundell and Stephen Bond. "Initial conditions and moment restrictions in dynamic panel data models". In: *Journal of econometrics* 87.1 (1998), pp. 115–143.

[17] Richard Blundell and James Powell. "Endogeneity in nonparametric and semiparametric regression models". In: (2001).

[18] Richard Blundell and Richard Smith. "Estimation in a class of simultaneous equation limited dependent variable models". In: *The Review of Economic Studies* 56.1 (1989), pp. 37–57.

[19] Gary Chamberlain. "Multivariate regression models for panel data". In: *Journal of econometrics* 18.1 (1982), pp. 5–46.

[20] Victor Chernozhukov and Christian Hansen. "An IV model of quantile treatment effects". In: *Econometrica* 73.1 (2005), pp. 245–261.

[21] Victor Chernozhukov, Guido W Imbens, and Whitney K Newey. "Instrumental variable estimation of nonseparable models". In: *Journal of Econometrics* 139.1 (2007), pp. 4–14.

[22] Nicolas De Roux et al. *Estimating production functions in differentiated-product industries with quantity information and external instruments*. Tech. rep. National Bureau of Economic Research, 2021.

[23] Gerard Debreu. "Representation of a preference ordering by a numerical function". In: *Decision processes* 3 (1954), pp. 159–165.

[24] Mert Demirer. "Production function estimation with factor-augmenting technology: An application to markups". In: *Job Market Paper* (2020).

[25] Sabien Dobbelaere and Jacques Mairesse. "Panel data estimates of the production function and product and labor market imperfections". In: *Journal of Applied Econometrics* 28.1 (2013), pp. 1–46.

[26] Ulrich Doraszelski and Jordi Jaumandreu. "Measuring the bias of technological change". In: *Journal of Political Economy* 126.3 (2018), pp. 1027–1084.

[27] Ulrich Doraszelski and Jordi Jaumandreu. "R&D and productivity: Estimating endogenous productivity". In: *Review of Economic Studies* 80.4 (2013), pp. 1338–1383.

[28] Jeremy T Fox et al. *Heterogenous Production Functions, Panel Data, and Productivity Dispersion*. 2017. URL: https://econ2017.sites.olt.ubc.ca/files/2019/03/pdf_seminar_Productivity-Slides_Fox.pdf (visited on 03/15/2022).

[29] Amit Gandhi, Salvador Navarro, and David A Rivers. "On the identification of gross output production functions". In: *Journal of Political Economy* 128.8 (2020), pp. 2973–3016.

[30] Matthew Grennan. "Price discrimination and bargaining: Empirical evidence from medical devices". In: *American Economic Review* 103.1 (2013), pp. 145–77.

[31] Jinyong Hahn, Zhipeng Liao, and Geert Ridder. *Identification and the Influence Function of Olley and Pakes'(1996) Production Function Estimator*. Tech. rep. Working Paper, UCLA and USC, 2021.

[32] James J Heckman. *Dummy endogenous variables in a simultaneous equation system*. Tech. rep. National Bureau of Economic Research, 1977.

[33] Yingyao Hu and Matthew Shum. "Identifying Dynamic Games with Serially Correlated Unobservables". In: *Structural Econometric Models*. Emerald Group Publishing Limited, 2013.

[34] Yingyao Hu and Matthew Shum. "Nonparametric identification of dynamic models with unobserved state variables". In: *Journal of Econometrics* 171.1 (2012), pp. 32–44.

[35] Guido W Imbens and Whitney K Newey. "Identification and estimation of triangular simultaneous equations models without additivity". In: *Econometrica* 77.5 (2009), pp. 1481–1512.

[36] Jean-Yves Jaffray. "Existence of a continuous utility function: an elementary proof". In: *Econometrica: Journal of the Econometric Society* (1975), pp. 981–983.

[37] Hiroyuki Kasahara, Mitsukuni Nishida, and Michio Suzuki. *Decomposition of Aggregate Productivity Growth with Unobserved Heterogeneity*. RIETI, 2017.

[38] Hiroyuki Kasahara, Paul Schrimpf, and Michio Suzuki. "Identification and estimation of production function with unobserved heterogeneity". In: *University of British Columbia mimeo* (2015).

[39] Ivan Kirov and James Traina. *Labor Market Power and Technological Change in US Manufacturing*. 2021. URL: https://raw.githubusercontent.com/James-Traina/Preprints/main/Traina_JMP.pdf (visited on 03/15/2022).

[40] Robin S Lee. "Vertical integration and exclusivity in platform and two-sided markets". In: *American Economic Review* 103.7 (2013), pp. 2960–3000.

[41] Erich Leo Lehmann, Joseph P Romano, and George Casella. *Testing statistical hypotheses*. Vol. 3. Springer, 2005.

[42] James Levinsohn and Amil Petrin. "Estimating production functions using inputs to control for unobservables". In: *The review of economic studies* 70.2 (2003), pp. 317–341.

[43] Yi Lu, Yoichi Sugita, and Lianming Zhu. *Wage and Markdowns and FDI Liberalization*. Tech. rep. Hitotsubashi Institute for Advanced Study, Hitotsubashi University, 2019.

[44] Rosa Matzkin. "Nonparametric estimation of nonadditive random functions". In: *Econometrica* 71.5 (2003), pp. 1339–1375.

[45] Rosa Matzkin. "Nonparametric identification". In: *Handbook of econometrics* 6 (2007), pp. 5307–5368.

[46] Rosa Matzkin. "Unobservable instruments". In: *Northwestern University Department of Economics Working Paper* (2004).

[47] Salvador Navarro and David A Rivers. *Nonparametric Identification of Productivity in Nonseparable Production Functions*. Tech. rep. Working paper, 2018.

[48] Whitney K Newey. "The asymptotic variance of semiparametric estimators". In: *Econometrica: Journal of the Econometric Society* (1994), pp. 1349–1382.

[49] Ezra Oberfield and Devesh Raval. "Micro data and macro technology". In: *Econometrica* 89.2 (2021), pp. 703–732.

[50] G Steven Olley and Ariel Pakes. "The Dynamics of Productivity in the Telecommunications Equipment Industry". In: *Econometrica* 64.6 (1996), pp. 1263–1297.

[51] Qingsong Pan. "Identification of Gross Output Production Function With Nonseparable Productivity Shock". In: *Unpublished manuscript* (2022).

[52] Qingsong Pan. "Support Conditions in Control Function Approaches". In: *Unpublished manuscript* (2021).

[53] Devesh Raval. "Testing the production approach to markup estimation". In: *Available at SSRN 3324849* (2020).

[54] Devesh Raval. "The micro elasticity of substitution and non-neutral technology". In: *The RAND Journal of Economics* 50.1 (2019), pp. 147–167.

[55] Michael Rubens. "Monopsony power and factor-biased technology adoption". In: *University of Leuven mimeo* (2019).

[56] Ariel Rubinstein. *Lecture notes in microeconomic theory*. Princeton University Press, 2012.

[57] Enrique Sentana. "Finite Underidentification". In: *Documentos de Trabajo (CEMFI)* 8 (2015), p. 1.

[58] Christopher Sullivan. "The ice cream split: Empirically distinguishing price and product space collusion". In: *Available at SSRN 3321948* (2017).

[59] Andrew Sweeting. "Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry". In: *Econometrica* 81.5 (2013), pp. 1763–1803.

[60] Johannes Van Biesebroeck. "Productivity dynamics with technology choice: An application to automobile assembly". In: *The review of economic studies* 70.1 (2003), pp. 167–198.

[61] Jeffrey M Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010.

[62] Hongsong Zhang. "Non-neutral technology, firm heterogeneity, and labor demand". In: *Journal of Development Economics* 140 (2019), pp. 145–168.

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

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

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

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

Proof. since $\varsigma_{it}^2 = F^{-1}_{x_{it}^j|\{y_{i\tau}\}_{\tau=t-M}^{t-1},\{x_{i\tau}\}_{\tau=t-M}^{t-1},\varsigma_t^1}\left(x_{it}^2, \{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}, \varsigma_{it}^1\right)$, 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 $\{\omega_{i\tau}\}_{\tau=t-M}^{t-1}$ 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 $\left((\varsigma_{it}^1, \varsigma_{it}^2), \{\omega_{i\tau}\}_{\tau=t-M}^{t-1}\right)$. By Lemma (4), we have $(\varsigma_{it}^1, \varsigma_{it}^2)$ and $\left(\{y_{i\tau}\}_{\tau=t-M}^{t-1}, \{x_{i\tau}\}_{\tau=t-M}^{t-1}\right)$ independent of each other, from which the conclusion follows. ■

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

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

B Additional Proofs

B.1 Proof of Theorem (3)

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, $y^{0A} \gtreqless y^{0B} \Leftrightarrow f_t^{-1}(x^A, y^A) \gtreqless f_t^{-1}(x^B, y^B)$.

Now we know under Assumption (11), for each consecutive pairs of points $((x^j, y^j), (x^{j+1}, y^{j+1}))$ in the sequence, 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 (4)

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), $(x^B, y^B) \in Int\,(\mathcal{S}_t^{xy})$. Given Assumption (13), 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 (12), 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., $\exists\ \epsilon$ 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 contructive 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 (12), 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 $x^{new} = x^B$.

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

B.3 Proof of Theorem (6)

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 (7), (8), (9), (10), and (11) 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 16 *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 $\tau \in (0,1)$,

$$Pr(y_{it} \leq \bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau) | \widetilde{v}_{it-1}, z_{it}) = \tau. \tag{18}$$

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 (18) 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 (18), 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, Imbens, and Newey (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 $(\widetilde{x}, \widetilde{v}, z) \in \mathcal{S}_t^{\widetilde{x}\widetilde{v}z}$.[27] 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 1 *$E(\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau).w_\tau(\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})|\widetilde{v}_{it-1}, z_{it}) = 0$ a.s. $\Rightarrow \Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau) = 0$ a.s., for $w_\tau(\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it}) = \int_0^1 f_{\epsilon_{it}^\tau}(\delta\Delta(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \tau)|\widetilde{x}_{it}, \widetilde{v}_{it-1}, z_{it})d\delta > 0$.*

Condition 2 *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.*[28] [29]

Condition (1) 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, Romano, and Casella (2005) that Condition (2) suffices for Condition (1). Condition (2) 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 $\bar{f}$.

[27](ii) is another restriction we impose on the relevant parameter space and thus on $\bar{f}(\widetilde{x}, \widetilde{v}, \tau)$. See Theorem (8) below and the discussion that follows.

[28]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 $(\widetilde{v}_{it-1}, z_{it}) = (\widetilde{v}^0, z)$.

[29]Note $\varphi_\tau(\widetilde{x}|\widetilde{v}^0, z)$ depends on $\Delta(\widetilde{x}, \widetilde{v}^0, \tau)$, so Condition (2) (or analogously Condition (1)) 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, Imbens, and Newey (2007).

Theorem 8 *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 (1) (or Condition (2)) 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 (18) 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 (8).

Proof. For each $\tau\in(0,1)$, we know $\bar{f}(.,\tau)$ solves equation (18). 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 (18) in $\mathcal{L}_\tau$. Suppose there is $m(.,\tau)$ that solves equation (18) a.s.. Define $\Delta(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)=\bar{f}(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)-m(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)$, and we have

$$\begin{aligned}
0 &= Pr(y_{it}\le m(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)|\widetilde{v}_{it-1},z_{it})-Pr(y_{it}\le \bar{f}(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)|\widetilde{v}_{it-1},z_{it}) \qquad (19)\\
&= E(Pr(y_{it}\le m(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)|\widetilde{x}_{it},\widetilde{v}_{it-1},z_{it})|\widetilde{v}_{it-1},z_{it})\\
&\quad - E(Pr(y_{it}\le \bar{f}(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)|\widetilde{x}_{it},\widetilde{v}_{it-1},z_{it})|\widetilde{v}_{it-1},z_{it})\\
&= E(Pr(y_{it}-\bar{f}(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)\le m(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)-\bar{f}(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)|\widetilde{x}_{it},\widetilde{v}_{it-1},z_{it})|\widetilde{v}_{it-1},z_{it})\\
&\quad - E(Pr(y_{it}-\bar{f}(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)\le 0|\widetilde{x}_{it},\widetilde{v}_{it-1},z_{it})|\widetilde{v}_{it-1},z_{it})\\
&= E(Pr(\epsilon_{it}^\tau\le \Delta(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)|\widetilde{x}_{it},\widetilde{v}_{it-1},z_{it})|\widetilde{v}_{it-1},z_{it})\\
&\quad - E(Pr(\epsilon_{it}^\tau\le 0|\widetilde{x}_{it},\widetilde{v}_{it-1},z_{it})|\widetilde{v}_{it-1},z_{it})\\
&= E(E(\int_0^1 f_{\epsilon_{it}^\tau}(\delta\Delta(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)|\widetilde{x}_{it},\widetilde{v}_{it-1},z_{it})\Delta(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)d\delta|\widetilde{x}_{it},\widetilde{v}_{it-1},z_{it})|\widetilde{v}_{it-1},z_{it})\\
&= E(\int_0^1 f_{\epsilon_{it}^\tau}(\delta\Delta(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)|\widetilde{x}_{it},\widetilde{v}_{it-1},z_{it})\Delta(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)d\delta|\widetilde{v}_{it-1},z_{it})\\
&= E(\Delta(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)w_\tau(\widetilde{x}_{it},\widetilde{v}_{it-1},z_{it})|\widetilde{v}_{it-1},z_{it}). \quad a.s.
\end{aligned} \qquad (20)$$

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 $(\widetilde{x},\widetilde{v},z)\in\mathcal{S}_t^{\widetilde{x}\widetilde{v}z}$.[30] By continuity and boundedness of $f_{\epsilon_{it}^\tau}(e|\widetilde{x}_{it},\widetilde{v}_{it-1},z_{it})$, the integration after the fifth equality is defined, so the fifth equality holds.

[30]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 (19) does not imply $\Delta(\widetilde{x}_{it},\widetilde{v}_{it-1},\tau)=0$ a.s..

Since Condition (1) 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 9 *If $\bar{f}(\widetilde{x}_{it}, \widetilde{v}_{it-1}, \xi_{it})$ is identified, under Assumption (11), $f(x_{it}, \omega_{it})$ 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.

$$\mathcal{W}(\mathcal{S}) = \left\{ \begin{array}{c} (x,y): \text{ for some } (x^A, y^A) \in \mathcal{S} \ \exists v^0 \text{ s.t. } (x^A, y^A, v^0) \in \mathcal{S}_t^{xyv}, x \in \mathcal{S}_t^{x|v^0}, \\ y = \bar{f}(\widetilde{x}, \widetilde{v}^0, \bar{f}^{-1}(\widetilde{x}^A, \widetilde{v}^0, y^A)) \end{array} \right\}.$$[31]

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 (3). ■

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