Identification and Estimation of Production Function and Consumer Demand Function under Monopolistic Competition from Revenue Data*

Chun Pang Chow† Hiroyuki Kasahara‡ Yoichi Sugita§

November 18, 2025
Incomplete Draft: Please Do Not Quote

Abstract

Commonly used methods for estimating production functions assume that firms' output quantities are observable, whereas typical datasets contain only revenue. We study the nonparametric identification of production and consumer demand functions from revenue data under monopolistic competition with a general nonparametric demand function. Under standard assumptions, we provide a constructive, nonparametric identification of several firm-level objects, including the gross production function, total factor productivity (TFP), price markups over marginal costs, output prices, output quantities, and the demand function. Considering the homothetic single-aggregator (HSA) demand system of **?**, we further identify the representative consumer's utility function, enabling counterfactual analysis of market outcomes and welfare. We propose a semiparametric estimator feasible for standard firm-level datasets and show in simulations that it performs well, whereas treating revenue as output generates substantial bias. Applying the estimator to Chilean manufacturing data, we reject the CES specification in favor of HSA, and counterfactual analysis indicates that market power reduces welfare by about 3%–6% of industry revenue in the three largest manufacturing industries in 1996.

*Acknowledgement: We are grateful to comments made at seminars and conferences with regard to an earlier version of this work. We thank Zheng Han, Kohei Hashinokuchi, Yoko Sakamoto and Makoto Tanaka for excellent research assistance. Sugita acknowledges financial supports from JSPS KAKENHI (grant numbers: 17H00986, 19H01477, and 19H00594).

†Department of Economics, University of British Columbia, Canada. (Email: alexccp@student.ubc.ca)

‡Department of Economics, University of British Columbia, Canada. (Email: hkasahar@mail.ubc.ca)

§Faculty of Business and Commerce, Keio University, Japan. (E-mail: ysugita@fbc.keio.ac.jp)

1 Introduction

The estimation of production functions and markups is a central tool in empirical analyses of market outcomes.[1] The residual from an estimated production function, commonly interpreted as total factor productivity (TFP), serves as a key measure of firm-level technological efficiency (see Bartelsman and Doms (2000); Syverson (2011)) and its contribution to aggregate productivity (e.g., Olley and Pakes, 1996). Researchers also use estimated elasticities of production functions to study technological change (e.g., Van Biesebroeck, 2003; Doraszelski and Jaumandreu, 2018) and to infer price markups over marginal cost (e.g., Hall, 1988; De Loecker and Warzynski, 2012). Markup estimation based on production functions has been widely applied across diverse fields and complements the demand-based approach (e.g., Berry, Levinsohn, and Pakes, 1995) in empirical analyses of firms' market power.

A common assumption underlying many production function and markup estimation methods is that firms' output quantities are observable. In practice, however, most firm-level datasets contain only revenue information rather than physical quantities. As a result, many empirical studies approximate output by deflating firm-level revenue using an industry-level price deflator.[2] For production function estimation, this practice of using revenue in place of output quantity can be justified under perfect competition, where output prices are exogenous and identical across firms. Following Marschak and Andrews (1944)'s pioneering critique, a large body of research has cautioned against this approach under imperfect competition. Numerous studies have demonstrated that replacing output quantity with revenue can severely bias estimates of production function parameters (e.g., Klette and Griliches, 1996; De Loecker, 2011) and TFP (e.g., Foster, Haltiwanger, and Syverson, 2008; Katayama, Lu, and Tybout, 2009; De Loecker, 2011). More recently, Bond, Hashemi, Kaplan, and Zoch (2020) show that using revenue in place of output quantity may lead to serious biases in estimation of firm's markups. Despite these concerns, the practice persists due to the scarcity of firm-level quantity data.[3]

This paper contributes to the literature on production function and markup estimation by establishing nonparametric identification of the production function, TFP, markup, and the consumer demand function from firm-level revenue data under monopolistic competition. Our identification proof is constructive, relying on standard assumptions from the production

[1] Griliches and Mairesse (1999) and Ackerberg, Benkard, Berry, and Pakes (2007) provide excellent surveys of production function estimation methods.

[2] A few studies employ datasets with firm-level quantity information (e.g., Foster, Haltiwanger, and Syverson, 2008; De Loecker, Goldberg, Khandelwal, and Pavcnik, 2016; Lu and Yu, 2015; Nishioka and Tanaka, 2019), but such data are typically limited to specific countries, industries, and time periods and remain inaccessible to most researchers.

[3] Researchers also rely on revenue when products differ in quality, since physical output alone may not reflect true production, though such practices often lack theoretical foundations.

function literature alongside additional non-parametric restrictions on firms' demand functions. This contribution is novel because the existing literature has not established whether it is possible to identify production functions and consumer demand function from firm-level revenue data without imposing parametric functional forms. Building on the constructive identification result, we further develop a semiparametric estimator and demonstrate through simulations that it performs well in finite samples.

Following Marschak and Andrews (1944), Klette and Griliches (1996) and De Loecker (2011), we explicitly model a demand function that an individual firm faces as a function of its output, observable characteristics, and an unobserved transitory demand shock.[4] While each of these earlier studies examines a demand function with a constant and identical demand elasticity which implies identical markups across firms, we consider a nonparametric demand function that generates heterogeneity in markups. Apart from this extension, our approach relies on standard assumptions commonly adopted in the literature and can be implemented using typical firm-level data found in empirical applications.

We develop a three-step identification approach that combines the control function approach developed by Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves, and Frazer (2015) and the first-order condition approach recently developed by Gandhi, Navarro, and Rivers (2020).[5] Following Levinsohn and Petrin (2003) and Ackerberg et al. (2015), the inverse function of a material demand function serves as a control function for TFP. In the first step, we identify an unobserved demand shock that nonlinearly affects revenue by using the control function as in Ackerberg et al. (2015) and the instrument variable quantile regression by Chernozhukov and Hansen (2005). Our novel second step identifies the control function for TFP by applying the nonparametric identification of transformation models (e.g., Horowitz, 1996) examined by Ekeland, Heckman, and Nesheim (2004) and Chiappori, Komunjer, and Kristensen (2015). By identifying the control function, TFP is identified (up to normalization) from the dynamics of inputs, without output data. In the third step, we identify a production function, markup, and a demand function, using the first-order condition for the material and the control function identified in the second step.

Our method identifies several key objects from revenue data. In our main setting, markups and output elasticities are identified up to scale, while the output price, output quantity, total factor productivity (TFP), gross production function, and consumer demand function are identified up to scale and location, without imposing any parametric functional form. Identification

[4]De Loecker, Eeckhout, and Unger (2020) study an alternative approach using an exogenous variable to remove output price variation from revenue data.

[5]These approaches assumed quantity data or perfect competition. Gandhi et al. (2020) also examined an imperfect competition with a constant elastic demand as in Klette and Griliches (1996) and De Loecker (2011) where markups must be constant and identical across firms.

is cross-sectional, allowing these objects to vary over time. With the additional assumption of *local* constant returns to scale, we can identify the levels of markups and output elasticities, and identify the output price, output quantity, TFP, production function, and consumer demand function up to location.[6] Finally, by considering the homothetic single-aggregator (HSA) demand system of **?**, we further identify the demand system and the representative consumer's utility function, and establish identification of counterfactual welfare effects.

We also develop a semiparametric estimator that assumes a Cobb-Douglas production function but imposes no parametric restrictions on the demand system. The estimation proceeds in three main steps, with an additional step devoted to demand-system estimation. In the first step, we nonparametrically estimate the transitory demand shock using the smooth GMM IV quantile regression proposed by Firpo, Galvao, Pinto, Poirier, and Sanroman (2022), which ensures quantile monotonicity in instrumental-variable quantile regression. In the second step, we estimate the control function using the profile likelihood estimator of Linton, Sperlich, and Van Keilegom (2008). In the third step, we recover the production function, markups, and TFP. This three-step procedure provides a standalone estimation of the production function that does not rely on any parametric assumptions about demand. In the fourth step, for the purpose of counterfactual welfare analyses as well as testing the CES demand assumption, we estimate the the CoPaTh-HSA demand system of Matsuyama and Ushchev (2020).[7]

Simulation results show that our estimator performs well in recovering structural parameters, markups, and TFPs. Applying the estimator to Chilean plant-level data from the three largest manufacturing industries (SIC 31, 32 and 38), we find evidence of misspecification under the CES demand system. Finally, our counterfactual welfare analysis reveals that the market power results in welfare losses of approximately 3%–6% of industry revenue in the three largest Chilean manufacturing industries in 1996.

The remainder of this paper is organized as follows. Section 2 summarizes previous studies on how using revenue as output could bias the identification of production function, TFP, and markup; readers familiar with the literature can skip this section and proceed to Section 3. Section 3 presents our nonparametric identification results. Subsection 3.1 explains our setting, and subsection 3.2 demonstrates our three-step approach by offering a parametric example. Subsection 3.3 presents our nonparametric identification results, and subsection 3.4 discusses additional assumptions for fixing scale and location normalization. Subsection 3.5 examines the

[6]Flynn, Gandhi, and Traina (2019) impose *global* constant returns to scale to identify a production function. In Subsection 3.4.2, we clarify the distinction between *local* and *global* constant returns to scale.

[7]One frequently sees within the literature an assumption of market structure for the identification of demand and supply side objects. For example, Berry, Levinsohn, and Pakes (1995) identify firm-level marginal costs by specifying oligopolistic competition; meanwhile, Ekeland, Heckman, and Nesheim (2004) and Heckman, Matzkin, and Nesheim (2010) identify various demand and supply side objects of a hedonic model by exploiting the properties of perfect competition.

identification of a demand system and a representative consumer's utility function. Subsection 3.6 and Appendix B presents identification results in alternative settings, including endogenous labor input. Section 4 present our semiparametric estimator while Section 5 presents the simulation results, comparing the performance of ACF method and our proposed method. Section 6 presents an empirical application of our estimator with counterfactual welfare analysis to the Chilean manufacturing plant data. Section 7 provides concluding remarks.

2 Biases from Using Revenue as Output Quantity

This section summarizes potential biases in the identification of the production function, total factor productivity (TFP), and markups when revenue is used in place of output quantity. Let p_{it}, y_{it}, and $r_{it} := p_{it} + y_{it}$ denote the logarithms of price, output, and revenue for firm i at time t, respectively. The corresponding uppercase letters represent their levels, e.g., $P_{it} = \exp(p_{it})$. Suppose that these variables are related via the inverse demand function $p_{it} = \psi_{it}(y_{it})$ and the revenue function $r_{it} = \varphi_{it}(y_{it}) := y_{it} + \psi_{it}(y_{it})$. Let $y_{it} = f_t(m_{it}, k_{it}, l_{it}) + \omega_{it}$ be firm i's production function where ω_{it} is TFP and $x_{it} := (m_{it}, k_{it}, l_{it})$ is a vector of the logarithms of material, capital, and labor, respectively. To highlight the sources of biases from using revenue as output, assume that TFP is identical across firms within time t, with $\omega_{it} = \omega_t$ for all i. This simplification eliminates an additional and well-known source of bias, correlations between inputs and TFP.

From the first-order condition for profit maximization, $P_{it}\left(1 + \psi'_{it}(y_{it})\right) = MC_{it}$, where MC_{it} denotes the marginal cost of producing one additional unit of output, the elasticity of revenue with respect to output equals the inverse of the markup:

$$\frac{d\varphi_{it}(y_{it})}{dy} = \frac{MC_{it}}{P_{it}}. \tag{1}$$

Under perfect competition, where $P_{it} = MC_{it}$, variation in revenue across firms coincides with that in output. However, they are generally different when markups vary across firms.

Suppose that, using revenue as output, a researcher identifies a true relationship between revenue and inputs, $\tilde{\varphi}_{it}(x_{it}) := \varphi_{it}(f_t(x_{it}) + \omega_t)$ to use $\tilde{\varphi}_{it}(x_{it})$ as a proxy for $f_t(x_{it})$. Prior studies show that the use of revenue as output could cause biases in three forms. First, Marschak and Andrews (1944) and Klette and Griliches (1996) establish that, from (1), the elasticity of $\tilde{\varphi}_{it}(x_{it})$ relates to the true elasticity of $f_t(x_{it})$ via markup:

$$\frac{\partial \tilde{\varphi}_{it}(x_{it})}{\partial v_{it}} = \frac{MC_{it}}{P_{it}} \frac{\partial f_t(x_{it})}{\partial v_{it}} \text{ for } v_{it} \in \{m_{it}, k_{it}, l_{it}\}. \tag{2}$$

Thus, output elasticities would be underestimated by the extent of markup.

Second, Katayama et al. (2009) and De Loecker (2011) demonstrated a bias in TFP estimates. Let $\Delta\omega_t$ be a TFP change. Suppose that a TFP change for firm i is estimated as a change in revenue with inputs being fixed, $\Delta\tilde{\omega}_{it} = \Delta\tilde{\varphi}_{it}(x_{it})|_{\Delta x_{it}=0}$. From (1), we see that this TFP estimate relates to the true TFP change via markup:

$$\Delta\tilde{\omega}_{it} = \frac{MC_{it}}{P_{it}}\Delta\omega_t. \tag{3}$$

Therefore, TFP would be underestimated by the extent of markup.

Finally, Bond et al. (2020) show that markup estimates using the method of Hall (1988) and De Loecker and Warzynski (2012) are generally biased when revenue elasticity is used in place of output elasticity. Suppose a firm is a price-taker of flexible input v. Hall (1988) and De Loecker and Warzynski (2012) developed the following equation relating to markup and output elasticity with respect to v as:

$$\frac{P_{it}}{MC_{it}} = \frac{\partial f_t(x_{it})/\partial v_{it}}{\alpha_{it}^v} \tag{4}$$

where α_{it}^v is the ratio of expenditure on input v to revenue. If a researcher uses $\partial\tilde{\varphi}_{it}(x_{it})/\partial m_{it}$ instead of $\partial f_t(x_{it})/\partial m_{it}$ in markup equation (4), then from (2), the estimated markup is 1:

$$\frac{\partial\tilde{\varphi}_{it}(x_{it})/\partial v_{it}}{\alpha_{it}^v} = \frac{\frac{MC_{it}}{P_{it}}\frac{\partial f_t(x_{it})}{\partial v_{it}}}{\alpha_{it}^v} = 1. \tag{5}$$

In such a case, the markup would be underestimated.[8]

Klette and Griliches (1996) and De Loecker (2011) developed methods by which to identify production functions from revenue data, by assuming a constant elastic demand function with an identical elasticity.[9] However, with this specific demand function, markups must be constant and identical across firms. Studies estimating markups from quantity data report substantial heterogeneity in markups across firms (e.g., De Loecker, Goldberg, Khandelwal, and Pavcnik, 2016; Lu and Yu, 2015; Nishioka and Tanaka, 2019).

[8]Result (5) by Bond et al. (2020) relies on the assumption that a researcher can correctly identify $\tilde{\varphi}_{it}(x_{it})$. In practice, misspecification of $\tilde{\varphi}_{it}(x_{it})$ could derive markup estimates (5) that contain some information on true markups. For instance, De Loecker and Warzynski (2012, Section VI) show that when f is Cobb–Douglas, it is possible to identify the effect of firm-level variables (e.g., export) on markups.

[9]Katayama et al. (2009) also developed a method by which to identify production functions from revenue data. Their method allows for markup heterogeneity but requires the ability to estimate firm's marginal costs from total costs.

3 Identification

3.1 Setting

We denote the logarithm of physical output, material, capital, and labor as y_{it}, m_{it}, k_{it}, and l_{it}, respectively, with their respective supports denoted as $\mathscr{Y}$, $\mathscr{M}$, $\mathscr{K}$, and $\mathscr{L}$. We collect the three inputs (material, capital, and labor) into a vector as $x_{it} := (m_{it}, k_{it}, l_{it})' \in \mathscr{X} := \mathscr{M} \times \mathscr{K} \times \mathscr{L}$.

At time t, output y_{it} is related to inputs $x_{it} = (m_{it}, k_{it}, l_{it})'$ through the production function

$$y_{it} = f_t\big(x_{it}, z^s_{it}\big) + \omega_{it}, \tag{6}$$

where z^s_{it} is a vector of exogenous characteristics with support $\mathscr{Z}_S$ that may affect either the functional form of $f_t(\cdot)$ or the level of total factor productivity (TFP) (e.g., ownership status). Firm-level productivity ω_{it} follows a first-order stationary Markov process given by

$$\omega_{it} = h\big(\omega_{it-1}, z^h_{it-1}\big) + \eta_{it}, \qquad \eta_{it} \overset{iid}{\sim} F_\eta, \tag{7}$$

where η_{it} is an innovation to productivity that is serially uncorrelated, and z^h_{it-1} is a vector of lagged characteristics with support $\mathscr{Z}_h$ that may affect the productivity process (e.g., previous import status as in Kasahara and Rodrigue, 2008). We assume that both the function $h(\cdot)$ and the marginal distribution of η_{it} are time-invariant.

The demand function for a firm's product is strictly decreasing in its price, and its inverse demand function is given by

$$p_{it} = \tilde{\psi}_t(y_{it}, z^d_{it}, \epsilon_{it}),$$

where z^d_{it} is an observable firm characteristic with support $\mathscr{Z}_D$ that affects firm's demand (e.g., firm's export status in De Loecker and Warzynski (2012)) while ϵ_{it} represents an unobserved demand shock.

The unobserved demand shock ϵ_{it} is generated by the process

$$\epsilon_{it} = \Upsilon(\zeta_{it}, \zeta_{it-1}, ..., \zeta_{it-\upsilon}), \qquad \zeta_{it-s} \overset{iid}{\sim} F_\zeta \quad \text{for } s = 0, 1, .., \upsilon. \tag{8}$$

Therefore, conditional on z^d_{it}, the underlying innovation ζ_{it} has a transitory effect on the demand shock ϵ_{it}. Consequently, ϵ_{it} is serially correlated over υ periods but its persistence is limited in that ϵ_{it} is independent of $\epsilon_{i,t-s}$ for $s \geq \upsilon + 1$. In contrast, an innovation to productivity η_{it} has a permanent impact on future productivity in (7). The difference between demand and supply shock specifications between (7) and (8) captures the idea that demand shocks are temporary while supply shocks are permanent (e.g., Nelson and Plosser, 1982).[10]

As shown in Matzkin (2003), the identification of a non-additive unobservable ϵ_{it} has to be up to its monotonic transformation. Let F_ϵ be the c.d.f. of ϵ_{it}. Without loss of generality, we transform ϵ_{it} to a uniform variable, using $u_{it} := F_\epsilon(\epsilon_{it})$,

$$\begin{aligned} p_{it} &= \tilde{\psi}_t(y_{it}, z^d_{it}, F^{-1}_\epsilon(u_{it})) \\ &= \psi_t(y_{it}, z^d_{it}, u_{it}),\ \ u_{it} \sim \text{Unif}(0,1). \end{aligned} \tag{9}$$

Given t, u_{it} cross-sectionally follows an independent and identical uniform distribution.

The inverse demand function (9) is non-parametrically specified and generalizes the constant-elasticity demand function examined by Marschak and Andrews (1944), Klette and Griliches (1996), and De Loecker (2011). Equation (9) implicitly imposes two key assumptions. First, $\psi_t(\cdot, z^d_{it}, u_{it})$ is common across firms once we control for observed demand characteristics z^d_{it} and a transitory scalar unobserved demand shock u_{it}. Second, $\psi_t(\cdot, z^d_{it}, u_{it})$ represents the demand curve that each individual firm takes as given. This assumption is satisfied under monopolistic competition (without free entry), where ψ_t can be expressed as $\psi_t(y_{it}, z^d_{it}, u_{it}, a_t)$, with a_t denoting a vector of aggregate price and quantity indices which each firm treats as exogenous.

Let r_{it} and $\mathscr{R}$ be the logarithm of revenue and its support, respectively. Then, from (6), the observed revenue relates to output and input as follows:

$$\begin{aligned} r_{it} &= \varphi_t(y_{it}, z^d_{it}, u_{it}) \\ &= \varphi_t\left(f_t(m_{it}, k_{it}, l_{it}, z^s_{it}) + \omega_{it}, z^d_{it}, u_{it}\right) \end{aligned} \tag{10}$$

where $\varphi_t(y_{it}, z^d_{it}, u_{it}) := \psi_t(y_{it}, z^d_{it}, u_{it}) + y_{it}$.

We make the following timing assumption.

Assumption 1. *(a) (l_{it}, k_{it}) is determined at the end of period $t-1$ and is independent of η_{is} and ζ_{is} for $s \geq t$. (b) m_{it} is determined after firm's observing $(\omega_{it}, u_{it}, z^s_{it}, z^d_{it})$ but is independent of η_{is} and ζ_{is} for $s \geq t+1$. (c) $\left(z^s_{it}, z^d_{it}, z^h_{it-1}\right)$ is continuous and independent of u_{is} and η_{is} for $s \geq t$. (d) each firm is a price-taker for material input.*

Assumptions 1(a)(b) specify the timing structure, which is similar to that in Gandhi et al. (2020). In Subsection 3.6, we present identification results when l_{it} is also endogenous. The continuity requirement in Assumption 1(c) can be relaxed, but the exogeneity of $\left(z^s_{it}, z^d_{it}, z^h_{it-1}\right)$ re-

[10]Here, the productivity ω_{it} may capture factors such as persistent differences in product quality across firms. From this perspective, the firm's output y_{it} can be interpreted as as a quality-adjusted measure of output quantity.

mains an important—albeit potentially strong—assumption, though it is commonly maintained in the empirical literature.[11] Assumption 1(d) is also standard in most empirical applications.

Under Assumption 1, the firm chooses $m_{it} = \mathbb{M}_t\left(\omega_{it}, k_{it}, l_{it}, z^s_{it}, z^d_{it}, u_{it}\right)$ at time t to maximize the profit:

$$\mathbb{M}_t\left(\omega_{it}, k_{it}, l_{it}, z^s_{it}, z^d_{it}, u_{it}\right) \in \arg\max_m \exp(\varphi_t(f_t(m, k_{it}, l_{it}, z^s_{it}) + \omega_{it}, z^d_{it}, u_{it})) - \exp(p^m_t + m), \quad (11)$$

where p^m_t denotes the logarithm of the common material price at time t.

Equation (10) highlights two identification issues, originally raised by Marschak and Andrews (1944). First, m_{it} correlates with two unobservables ω_{it} and u_{it}. Second, r_{it} relates to $x_{it} = (m_{it}, k_{it}, l_{it})$ via two unknown nonlinear functions $\varphi_t(\cdot, z^d_{it}, u_{it})$ and $f_t(\cdot)$.

For identification, we make the following assumptions.

Assumption 2. *(a) $f_t(\cdot)$ is continuously differentiable with respect to (m, k, l, z^s) on $\mathscr{M} \times \mathscr{K} \times \mathscr{L} \times \mathscr{Z}_s$ and strictly increasing in m. (b) For every $\left(z^d, u\right) \in \mathscr{Z}_d \times [0, 1]$, $\varphi_t(\cdot, z^d, u)$ is strictly increasing and invertible with its inverse $\varphi^{-1}_t(r, z^d, u)$, which is continuously differentiable with respect to (r, z^d, u) on $\mathscr{R} \times \mathscr{Z}_d \times [0, 1]$. (c) For every $(k, l, z^s, z^d, u) \in \mathscr{K} \times \mathscr{L} \times \mathscr{Z}_s \times \mathscr{Z}_d \times [0, 1]$, $\mathbb{M}_t(\cdot, k, l, z^s, z^d, u)$ is strictly increasing and invertible with its inverse $\mathbb{M}^{-1}_t(m, k, l, z^s, z^d, u)$, which is continuously differentiable with respect to (m, k, l, z^s, z^d, u) on $\mathscr{M} \times \mathscr{K} \times \mathscr{L} \times \mathscr{Z}_s \times \mathscr{Z}_d \times [0, 1]$. (d) $(\zeta_{it}, ..., \zeta_{it-v})$ are independent from η_{it}.*

Assumptions 2(a)(b) are standard assumptions about smooth production and demand functions. In Assumption 2(b), the condition $\partial \varphi_t(y_{it}, z^d_{it}, u_{it})/\partial y_{it} > 0$ is equivalent to that the elasticity of demand with respect to price, $-\left(\partial \psi_t(y_{it}, z^d_{it}, u_{it})/\partial y_{it}\right)^{-1}$, being greater than 1; this necessarily holds under profit maximization. Assumption 2(c) is a standard assumption in the control function approach that uses material as a control function for TFP (Levinsohn and Petrin, 2003; Ackerberg et al., 2015). Assumption 2(d) requires the demand shock and the productivity shock are independent.

Let $w_{it} := (k_{it}, l_{it}, z^s_{it}, z^d_{it})$ be observable exogenous variables at t. The inverse function of the material demand function with respect to TFP

$$\omega_{it} = \mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it})$$

is used as a control function for ω_{it}. Since $\partial \varphi_t(y_{it}, z^d_{it}, u_{it})/\partial y_t > 0$, there exists the inverse function $\varphi^{-1}_t(\cdot, z^d_{it}, u_{it})$ so that the revenue function $r_{it} = \varphi_t(f_t(x_{it}, z^s_{it}) + \omega_{it}, z^d_{it}, u_{it})$ can be

[11] In Appendix B, we further discuss identification when these variables are discrete and endogenous, under the availability of suitable instruments.

written as:

$$\varphi_t^{-1}\left(r_{it}, z_{it}^d, u_{it}\right) = f_t(x_{it}, z_{it}^s) + \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}). \tag{12}$$

Let

$$v_{it} := (w_{it}, u_{it}, m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h)' \in \mathscr{V} := \mathscr{W} \times [0,1] \times \mathscr{M} \times \mathscr{W} \times [0,1] \times \mathscr{Z}_h$$

where $\mathscr{W} := \mathscr{K} \times \mathscr{L} \times \mathscr{Z}_s \times \mathscr{Z}_d$. We assume that the data constitute a random sample of N firms observed over multiple periods,

$$\{\{r_{is}, m_{is}, v_{is}\}_{s=t-\upsilon-2}^t\}_{i=1}^N,$$

drawn from the population. Given a sufficiently large N, the econometrician can consistently recover the corresponding population joint distributions.

Assumption 3. *An econometrician is assumed to know the following objects: (a) the population joint distribution of $\{r_{is}, m_{is}, v_{is}\}_{s=t-\upsilon-2}^t$; and (b) the material input cost for each firm,* $\exp(p_t^m + m_{it})$.

Our objective is to identify $\varphi_t^{-1}(\cdot), f_t(\cdot), \mathbb{M}t^{-1}(\cdot)$ from the population joint distribution of $\{r_{is}, m_{is}, v_{is}\}_{s=t-\upsilon-2}^t\}$. Let $\{\varphi_t^{*-1}(\cdot), f_t^*(\cdot), \mathbb{M}_t^{*-1}(\cdot)\}$ be the true model structure that satisfies (12). Then, for any $(a_{1t}, a_{2t}, b_t) \in \mathbb{R}^2 \times \mathbb{R}_{++}$,

$$\varphi_t^{-1}\left(r_{it}, z_{it}^d, u_{it}\right) = (a_{1t} + a_{2t}) + b_t\varphi_t^{*-1}\left(r_{it}, z_{it}^d, u_{it}\right),\ f_t(x_{it}, z_{it}^s) = a_{1t} + b_t f_t^*(x_{it}, z_{it}^s),$$
$$\text{and } \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}) = a_{2t} + b_t\mathbb{M}_t^{*-1}(m_{it}, w_{it}, u_{it}) \tag{13}$$

also satisfy (12), and the true structure $\{\varphi_t^{*-1}(\cdot), f_t^*(\cdot), \mathbb{M}_t^{*-1}(\cdot)\}$ is observationally equivalent to the structure (13). That is, the structure $\{\varphi_t^{-1}(\cdot), f_t(\cdot), \mathbb{M}_t^{-1}(\cdot)\}$ is identified only up to location and scale normalization (a_{1t}, a_{2t}, b_t) from restriction (12).

Therefore, identification requires location and scale normalization. We fix (a_{1t}, a_{2t}, b_t) in (13) by fixing the values of $\{\varphi_t^{-1}(\cdot), f_t(\cdot), \mathbb{M}_t^{-1}(\cdot)\}$ at some points. Specifically, choosing two points (m_{t0}^*, w_t^*, u_t^*) and (m_{t0}^*, w_t^*, u_t^*) on the support $\mathscr{X} \times \mathscr{Z}$ where $m_{t0}^* < m_{t1}^*$, we denote

$$c_{1t} := f_t(m_{t0}^*, k_t^*, l_t^*, z_t^{s*}),\ c_{2t} = \mathbb{M}_t^{-1}(m_{t0}^*, w_t^*, u_t^*),\ \text{and } c_{3t} := \mathbb{M}_t^{-1}(m_{t1}^*, w_t^*, u_t^*). \tag{14}$$

Note that $\partial\mathbb{M}_t^{-1}/\partial m_t > 0$ implies that $c_{2t} < c_{3t}$. Then, there exists a unique one-to-one mapping between (c_{1t}, c_{2t}, c_{3t}) in (14) and (a_{1t}, a_{2t}, b_t) in (13) such that $b_t = (c_{3t} - c_{2t}) / \left(\mathbb{M}_t^{*-1}(m_{t1}^*, w_t^*, u_t^*) - \mathbb{M}_t^{*-1}(m_{t0}^*, w_t^*, u_t^*)\right)$, $a_{1t} = c_{1t} - b_{1t} f_t^*(m_{t0}^*, k_t^*, l_t^*, z_t^{s*})$ and $a_{2t} =$

$c_{2t} - b_{1t}\mathbb{M}_t^{*-1}(m_{t0}^*, w_t^*, u_t^*)$. Thus, we can fix the value of (a_{1t}, a_{2t}, b_t) by choosing arbitrary values $(c_{1t}, c_{2t}, c_{3t}) \in \mathbb{R}^3$ that satisfies $c_{2t} < c_{3t}$. In particular, we impose the following normalization that corresponds to (N2) in Chiappori et al. (2015).

Assumption 4. *(Normalization) The support $\mathscr{M} \times \mathscr{W} \times [0,1]$ includes two points (m_{t0}^*, w_t^*, u_t^*) and (m_{t0}^*, w_t^*, u_t^*) such that $c_{1t} = c_{2t} = 0$ and $c_{3t} = 1$ in (14).*

As Chiappori et al. (2015) demonstrates, this choice of normalization makes the identification proofs transparent. In Section 3.4, we discuss how we can use additional restrictions and data to identify the normalization parameters (a_{1t}, a_{2t}, b_t).

3.2 Identification in a Parametric Example

Before presenting the nonparametric identification results, we demonstrate our identification approach by applying it to a simple parametric example without exogenous covariates, i.e., where $(z_{it}^d, z_{it}^s, z_{it}^h)$ is empty. Consider a monopolistically competitive market where each firm i faces the following constant elastic inverse demand function:

$$p_{it} = \alpha_t + (\rho(u_{it}) - 1)y_{it}, \tag{15}$$

where α_t is an unknown parameter and $\rho(\cdot)$ is an unknown function satisfying $0 < \rho(\cdot) \leq 1$.[12] We assume that $\rho'(u) < 0$, which implies that the markup $1/\rho(u)$ is increasing in u.

Firm i has a Cobb–Douglas production function with the TFP ω_{it} follows a first-order autoregressive (AR(1)) process:

$$\begin{aligned} y_{it} &= \theta_0 + \theta_m m_{it} + \theta_k k_{it} + \theta_l l_{it} + \omega_{it}, \\ \omega_{it} &= h_1 \omega_{it-1} + \eta_{it}, \end{aligned} \tag{16}$$

where $\{\theta_0, \theta_m, \theta_k, \theta_l, h_1\}$ are unknown parameters. The firm's revenue function is expressed as:

$$r_{it} = \alpha_t + \rho(u_{it})\theta_0 + \rho(u_{it})\theta_m m_{it} + \rho(u_{it})\theta_k k_{it} + \rho(u_{it})\theta_l l_{it} + \rho(u_{it})\omega_{it}. \tag{17}$$

Denote the ratio of material cost to revenue as

$$s_{it} := \frac{\exp(p_t^m + m_{it})}{\exp(r_{it})}.$$

[12] The demand function (15) can be derived from a constant elasticity of substitution (CES) utility function, where the elasticity of substitution parameter depends on u. The term α_t implicitly captures aggregate expenditure and an aggregate price index.

Then, the first-order condition for (11) can be written as

$$\rho(u_{it})\theta_m = s_{it}, \tag{18}$$

which, in turn, determines the control function for ω_{it} as

$$\omega_{it} = \mathbb{M}_t^{-1}(m_{it}, k_{it}, l_{it}, u_{it}) = \beta_t(u_{it}) + \beta_m(u_{it})m_{it} + \beta_k k_{it} + \beta_l l_{it} \tag{19}$$

where $\beta_t(u_{it}) = \left(p_t^m - \alpha_t(u_{it}) - \theta_0 - \ln\rho(u_{it})\theta_m\right)/\rho(u_{it})$, $\beta_m(u_{it}) = (1-\rho(u_{it})\theta_m)/\rho(u_{it}) > 0$, $\beta_k = -\theta_k$ and $\beta_l = -\theta_l$.

For notational brevity, assume that the support $\mathscr{X}$ includes two points $(m_{t1}^*, k_t^*, l_t^*) = (0,0,0)$ and $(m_{t0}^*, k_t^*, l_t^*) = (1,0,0)$. Following Assumption 4, we fix the location and scale of $f_t(\cdot)$ and $\mathbb{M}_t^{-1}(\cdot)$ by imposing the following normalization:

$$\begin{aligned} &0 = f_t(0,0,0) = \theta_0,\ 0 = \mathbb{M}_t^{-1}(0,0,0,0.5) = \beta_t(0.5), \\ &1 = \mathbb{M}_t^{-1}(1,0,0,0.5) = \beta_t(0.5) + \beta_m(0.5) \end{aligned} \tag{20}$$

which implies $\theta_0 = 0$, $\beta_t(0.5) = 0$, and $\beta_m(0.5) = 1$.

Our identification approach follows three steps.

3.2.1 Step 1: Identification of the Demand Shocks

The first step identifies the demand shock u_{it}. Substituting $\omega_{it} = \mathbb{M}_t^{-1}(m_{it}, k_{it}, l_{it}, u_{it})$ and using $\theta_0 = 0$, we obtain

$$\begin{aligned} r_{it} =& (\alpha_t + \rho(u_{it})\beta_t(u_{it})) + \rho(u_{it})(\theta_m + \beta_m(u_{it}))m_{it} \\ &+ \rho(u_{it})(\theta_k + \beta_k)k_{it} + \rho(u_{it})(\theta_l + \beta_l)l_{it} \qquad (21) \\ =& \tilde{\phi}(u_{it}) + m_{it}, \qquad (22) \end{aligned}$$

where $\tilde{\phi}(u_{it}) := \alpha_t + \rho(u_{it})\beta_t(u_{it})$ with $\tilde{\phi}'(u) = -\theta_m\rho'(u)/\rho(u) > 0$ for all u.

From (22), we have

$$\Pr[r_{it} - m_{it} \le \tilde{\phi}_t(u)] = u \quad \text{for all } u \in [0,1]$$

because $\Pr[r_{it} - m_{it} \le \tilde{\phi}_t(u)] = \Pr[\tilde{\phi}_t(u_{it}) \le \tilde{\phi}_t(u)] = u$ by the monotonicity of $\tilde{\phi}_t(\cdot)$. Therefore, the quantile of $r_{it} - m_{it}$ identifies u_{it} while the moment condition $E\left[1\left\{r_{it} - m_{it} \le \tilde{\phi}_t(u)\right\} - u\right] = 0$ for $u \in [0,1]$ identifies $\tilde{\phi}_t(\cdot)$.

Alternatively, from the first-order condition (18) and the monotonicity of $\rho(\cdot)$ with $\rho'(\cdot) < 0$, the demand shock u_{it} is identified as the quantile of $1/s_{it}$. This equivalence arises because the quantile of $r_{it} - m_{it}$ coincides with that of $1/s_{it} = \exp(r_{it} - m_{it} - p_t^m)$.

Step 2: Identification of Control Function and TFP The second step identifies the control function $\mathbb{M}_t^{-1}(\cdot)$. Substituting (19) into the AR(1) process (16) leads to

$$\mathbb{M}_t^{-1}(m_{it}, k_{it}, l_{it}, u_{it}) = h_1 \mathbb{M}_{t-1}^{-1}(m_{it-1}, k_{it-1}, l_{it-1}, u_{it-1}) + \eta_{it}. \tag{23}$$

Since $\mathbb{M}_t^{-1}(m_{it}, k_{it}, l_{it}, u_{it})$ is linear in m_{it} from (19), we can rearrange (23) as:

$$\begin{aligned} m_{it} = {} & \gamma(u_{it}, u_{t-1}) + \gamma_k(u_{it})k_{it} + \gamma_l(u_{it})l_{it} + \delta_m(u_{it}, u_{it-1})m_{it-1} \\ & + \delta_k(u_{it})k_{it-1} + \delta_l(u_{it})l_{it-1} + \tilde{\eta}_{it}, \end{aligned} \tag{24}$$

where

$$\gamma_k(u_{it}) = -\frac{\beta_k}{\beta_m(u_{it})},\ \gamma_l(u_{it}) = -\frac{\beta_l}{\beta_m(u_{it})},\ \delta_k(u_{it}) = \frac{h_1\beta_k}{\beta_m(u_{it})},\ \delta_l(u_{it}) = \frac{h_1\beta_l}{\beta_m(u_{it})}, \tag{25}$$

$$\gamma(u_{it}, u_{it-1}) = \frac{-\beta_t(u_{it}) + h_1\beta_{t-1}(u_{it-1})}{\beta_m(u_{it})}, \tag{26}$$

$\tilde{\eta}_{it} = \eta_{it}/\beta_m(u_{it})$, and $\delta_m(u_{it}, u_{it-1}) = h_1\beta_m(u_{it-1})/\beta_m(u_{it})$. For a given (u_{it}, u_{it-1}), (24) is a linear model. Since $E[\,\tilde{\eta}_{it}|\, v_{it}] = E[\,\eta_{it}|\, v_{it}]\,/\beta_m(u_{it}) = 0$, where $v_{it} := (k_{it}, l_{it}, x_{it-1}, u_{it}, u_{it-1})$, we can identify $\{\gamma(\cdot,\cdot),\ \gamma_k(\cdot),\ \gamma_l(\cdot),\ \delta_m(\cdot,\cdot),\ \delta_k(\cdot),\ \delta_l(\cdot)\}$ in (24) from the conditional moment restriction $E[\,\tilde{\eta}_{it}|\, v_{it}] = 0$.

Then, because $\beta_m(0.5) = 1$, we can recover $(\theta_k, \theta_l, h_1)$ from (25) as

$$\theta_k = -\beta_k = \gamma_k(0.5),\ \ \theta_l = -\beta_l = \gamma_l(0.5),\ \text{and } h_1 = -\frac{\delta_k(0.5)}{\gamma_k(0.5)} = -\frac{\delta_l(0.5)}{\gamma_l(0.5)}.$$

Also, applying the normalization (20) to (26), we have $\gamma(0.5, u_{t-1}) = h_1\beta_{t-1}(u_{t-1})$. Then, $\beta_m(u)$ and $\beta_t(u)$ are identified from (25)-(26) as

$$\beta_m(u) = \frac{\gamma_k(0.5)}{\gamma_k(u)} = \frac{\gamma_l(0.5)}{\gamma_l(u)} \text{ and } \beta_t(u) = \gamma(0.5, u_{t-1}) - \frac{\gamma(u, u_{t-1})\gamma_k(0.5)}{\gamma_k(u)}.$$

Given the identification of $(\beta_k, \beta_l, \beta_m(\cdot), \beta_t(\cdot))$, we can identify ω_{it} from (19).

Step 3: Identification of Production Function, Markup, and Demand Function The identification of $\rho(u)$ follows from substituting (18) into $\beta_m(u) = (1 - \rho(u)\theta_m)\,/\rho(u)$, and rearranging the terms, which yields

$$\rho(u_{it}) = \frac{1-s_{it}}{\beta_m(u_{it})} = \frac{1-s_{it}}{\gamma_k(0.5)/\gamma_k(u_{it})}.$$

Therefore, the markup $1/\rho(u_{it})$ is identified.

The first order condition (18) implies that the revenue share of material expenditure is a function of u_{it}, which we denote by $s(u)$, such that $s_{it} = s(u_{it})$. In particular, $s(0.5)$ represents the median revenue share of material expenditure. Then, the identification of θ_m follows from the identification of $\rho(u)$ and the first order condition (18) as

$$\theta_m = \frac{s(0.5)}{\rho(0.5)} = \frac{s(0.5)}{1-s(0.5)}. \tag{27}$$

The Identification under Normalization In view of the first-order condition $\rho(u_{it})\theta_m = s_{it}$, it is clear from the argument above that the markup level cannot be separately identified from the material input coefficient θ_m without imposing the normalization restriction $\beta_m(0.5) = 1$.

More generally, the parameters are identified under the scale and location normalization of $f_t(\cdot)$ and $\mathbb{M}_t^{-1}(\cdot)$ in (20). Let θ_i $(i = 0, m, k, l)$ and $\beta_j(u_t)$ $(j = t, m, k, l)$ be those parameters identified above and let θ_j^* and $\beta_i^*(u_t)$ be the true parameters. Then, there exist unknown normalization parameters $(a, b) \in \mathbb{R} \times \mathbb{R}_+$ such that

$$\theta_0 = a + b\theta_0^*,\ \beta_t = a + b\beta_t^*,\ \theta_i = b\theta_i^*,\ \beta_j(u_t) = b\beta_j^*(u_t).$$

We can fix the normalization by imposing further restrictions. For instance, if constant returns to scale $\theta_m^* + \theta_k^* + \theta_l^* = 1$ holds, then the scale parameter b can be identified as

$$b = b\left(\theta_m^* + \theta_k^* + \theta_l^*\right) = \theta_m + \theta_k + \theta_l = \frac{s(0.5)}{1-s(0.5)} - \beta_k - \beta_l.$$

We discuss in subsection 3.4 additional assumptions for fixing normalization.

The above identification argument is illustrative but relies on the linearity of $\mathbb{M}_t^{-1}(m_{it}, k_{it}, l_{it}, u_{it})$ under restrictive parametric assumptions. The next subsection establishes nonparametric identification in a more general framework presented in Section 3.1.

3.3 Nonparametric Identification

3.3.1 Step 1: Identification of the Demand Shocks

Substituting $\omega_{it} = \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})$ into the revenue function (10), we can rewrite it as

$$\begin{aligned} r_{it} &= \varphi_t\left(f(m_{it}, k_{it}, l_{it}, z_{it}^s) + \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}), z_{it}^d, u_{it}\right) \\ &=: \phi_t(m_{it}, w_{it}, u_{it}), \;\; u_{it} \sim \text{Unif}(0,1). \end{aligned} \tag{28}$$

We impose the following assumptions.

Assumption 5. *(a) (Monotonicity)* $\partial \phi_t(m, w, u)/\partial u > 0$ *for all* $(m, w, u) \in \mathscr{M} \times \mathscr{W} \times [0, 1]$.
(b) (Completeness) The conditional distribution of (m_{it}, w_{it}) *given* (m_{it-v-1}, w_{it-v}) *is complete in the sense of Chernozhukov and Hansen (2005); that is, for any measurable function* $g(m, w)$*,*

$$E[g(m_{it}, w_{it}) \mid m_{it-v-1}, w_{it-v}] = 0 \;\textit{a.s.} \;\; \Rightarrow g(m_{it}, w_{it}) = 0 \;\textit{a.s.}$$

Assumption 5(a) implicitly imposes restrictions on the shape of the demand function. The Appendix shows that Assumption 5(a) holds if and only if $\frac{\partial \varphi_t}{\partial u}\frac{\partial \sigma_t}{\partial y} > \frac{\partial \varphi_t}{\partial y}\frac{\partial \sigma_t}{\partial u}$, where $\sigma_t(y, z^d, u) := -1/\left(\frac{\partial \psi_t(y,z^d,u)}{\partial y}\right) > 0$ denote the demand elasticity. Since $\frac{\partial \varphi_t}{\partial u} > 0$ and $\frac{\partial \varphi_t}{\partial y} > 0$, a sufficient condition for Assumption 5(a) is that an increase in the demand shock ϵ_{it} makes demand less elastic (i.e., increases the markup), while an increase in consumption makes demand more elastic (i.e., decreases the markup).

Under Assumption 5(a), given values of (m, w), $\phi_t(m, w, \cdot)$ in (28) can be interpreted as the quantile function of revenue r. Although m_{it} is endogenous and correlated with u_{it}, Assumption 1(ii) and equation (8) imply that u_{it} is independent of (m_{it-v-1}, w_{it-v}) while u_{it} is serially correlated with u_{is} for $s = 1, ..., v$. Then, we have[13]

$$\Pr[r_{it} \le \phi_t(m_{it}, w_{it}, u) | m_{it-v-1}, w_{it-v}] = u \quad \text{for all } u \in [0, 1].$$

Assumption 5(b), referred to as the completeness condition, implies the following uniqueness property: for any two candidate functions ϕ_t^1 and ϕ_t^2 and any fixed $u \in [0, 1]$, $E\left[1\left\{r_{it} \le \phi_t^1(m_{it}, w_{it}, u)\right\}\middle| m_{it-v-1}, w_{it-v}\right] = E\left[1\left\{r_{it} \le \phi_t^2(m_{it}, w_{it}, u)\right\}\middle| m_{it-v-1}, w_{it-v}\right]$ a.s.

[13]This follows because

$$\begin{aligned} \Pr[r_{it} \le \phi_t(m_{it}, w_{it}, u) | m_{it-v-1}, w_{it-v}] &= \Pr[\phi_t(m_{it}, w_{it}, u_{it}) \le \phi_t(m_{it}, w_{it}, u) | m_{it-v-1}, w_{it-v}] \\ &= \Pr[u_{it} \le u | m_{it-v-1}, w_{it-v}] \\ &= u, \end{aligned}$$

where the second equality follows from the monotonicity of $\phi_t(m, w, \cdot)$ while the last equality holds because $u_{it} \perp (m_{it-v-1}, w_{it-v})$.

implies that $\phi_t^1(\cdot,\cdot,u) = \phi_t^2(\cdot,\cdot,u)$ almost surely. Then, following Chernozhukov and Hansen (2005), the moment condition

$$E\left[1\left\{r_{it} \leq \phi_t\left(m_{it}, w_{it}, u\right)\right\} - u \middle| m_{it-v-1}, w_{it-v}\right] = 0 \text{ for } u \in [0,1] \tag{29}$$

identifies $\phi_t(\cdot)$, and the demand shock u_{it} is identified as $u_{it} = \phi_t^{-1}(r_{it}, m_{it}, w_{it})$ under Assumption 5.

Proposition 1. *Under Assumptions 1, 2, 3, and 5 hold, $\phi_t(\cdot)$ and u_{it} are identified.*

Hereafter, $\phi_t(\cdot)$ and u_{it} are assumed to be known.

3.3.2 Step 2: Identification of Control Function and TFP

From (7), the control function $\omega_{it} = \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})$ satisfies

$$\mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}) = \bar{h}_t\left(m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h\right) + \eta_{it}, \tag{30}$$

where $\bar{h}_t\left(m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h\right) := h\left(\mathbb{M}_{t-1}^{-1}(m_{it-1}, w_{it-1}, u_{it-1}), z_{it-1}^h\right)$. As $\partial\mathbb{M}_t^{-1}/\partial m_{it} > 0$, given the values of (w_{it}, u_{it}), the dependent variable in (30) is a monotonic transformation of m_{it}. Therefore, the model (30) belongs to a class of transformation models, the identification of which Chiappori et al. (2015) analyze.

We make the following assumption, which corresponds to Assumptions A1–A3, A5, and A6 in Chiappori et al. (2015).[14]

Assumption 6. *(a) The distribution $G_\eta(\cdot)$ of η_{it} is absolutely continuous with a density function $g_\eta(\cdot)$ that is continuous on its support. (b) η_{it} is independent of $v_{it} := (w_{it}, u_{it}, m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h)' \in \mathscr{V}$ with $E[\eta_{it}|v_{it}] = 0$. (c) v_{it} is continuously distributed on $\mathscr{V}$. (d) Support Ω of ω_{it} is an interval $[\underline{\omega}, \bar{\omega}] \subset \mathbb{R}$, where $\underline{\omega} < 0$ and $1 < \bar{\omega}$. (e) $h(\cdot)$ is continuously differentiable with respect to (ω, z_h) on $\Omega \times \mathscr{Z}_h$. (f) The set $\mathscr{A}_{q_{t-1}} := \{(m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h) \in \mathscr{M} \times \mathscr{W} \times [0,1] \times \mathscr{Z}_h : \partial G_{m_t|v_t}(m_{it}|v_{it})/\partial q_{it-1} \neq 0$ for all $(m_{it}, w_{it}, u_t) \in \mathscr{M} \times \mathscr{W} \times [0,1]\}$ is nonempty for some $q_{it-1} \in \{m_{it-1}, k_{it-1}, l_{it-1}, z_{it-1}^s, z_{it-1}^d, u_{it-1}, z_{it-1}^h\}$.*

We can relax Assumption 6(b) by allowing z_{it}^h and l_{it} to correlate with η_{it} as discussed in subsection B. The sign restriction in Assumption 6(d) holds without loss of generality because we can choose any two points in place of $\{0,1\}$ on the support of ω_{it} without changing the essence of our argument.

[14] Assumption 2 (c) corresponds to Assumption A4 of Chiappori et al. (2015).

Assumption 6(f) can be interpreted as a generalized rank condition. Suppose $g_\eta(\eta_{it}) > 0$ for all $\eta_{it} \in \mathbb{R}$. Then, as will be shown below in (32), Assumption 6(f) holds if either

$$\frac{\partial \bar{h}\left(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)}{\partial \tilde{z}^h_{it-1}} \neq 0 \text{ or}$$

$$\frac{\partial h\left(\mathbb{M}^{-1}_{t-1}(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}), \tilde{z}^h_{it-1}\right)}{\partial \omega_{it-1}} \frac{\partial \mathbb{M}^{-1}_{t-1}(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1})}{\partial q_{it-1}} \neq 0$$

holds for some $\left(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)$ and $q_{it-1} \in \{m_{it-1}, k_{it-1}, l_{it-1}, z^s_{it-1}, z^d_{it-1}, u_{it-1}, z^h_{it-1}\}$. The latter condition is equivalent to (1) ω_{it-1} has a causal impact on ω_{it} ($\partial h / \partial \omega_{it-1} \neq 0$) and (2) q_{it-1} has a causal impact on ω_{it-1}, ($\partial \mathbb{M}^{-1}_{it-1} / \partial q_{it-1} \neq 0$). These conditions must be satisfied for at least one exogenous variable q_{it-1} and some point $(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1})$.

Proposition 2 shows that the control function is identified from the distribution of (m_{it}, v_{it}).

Proposition 2. *Suppose that Assumptions 1–6 hold. Then, we can identify $\mathbb{M}^{-1}_t(\cdot)$ up to scale and location and $G_\eta(\cdot)$ up to the scale normalization of η_{it}.*

Proof. The proof follows the proof of Theorem 1 in Chiappori et al. (2015). In view of equation (30), the conditional distribution of m_{it} given v_{it} satisfies

$$\begin{aligned} G_{m_t|v_t}(m_{it}|v_{it}) &= G_{\eta_t|v_t}\left(\mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it}) - \bar{h}_t\left(m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}\right)|v_{it}\right) \\ &= G_\eta\left(\mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it}) - \bar{h}_t\left(m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}\right)\right), \end{aligned}$$

where the second equality follows from $\eta_{it} \perp v_{it}$ in Assumption 6(b). Let $q_{it} \in \{m_{it}, k_{it}, l_{it}, z^s_{it}, z^d_{it}, u_{it}\}$ and $q_{it-1} \in \{m_{it-1}, k_{it-1}, l_{it-1}, z^s_{it-1}, z^d_{it-1}, z^h_{it-1}, u_{it-1}\}$. The derivatives of $G_{m_t|v_t}(m_{it}|v_{it})$ are

$$\frac{\partial G_{m_t|v_t}(m_{it}|v_{it})}{\partial q_{it}} = \frac{\partial \mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it})}{\partial q_{it}} g_\eta(\eta_{it}), \tag{31}$$

$$\frac{\partial G_{m_t|v_t}(m_{it}|v_{it})}{\partial q_{it-1}} = -\frac{\partial \bar{h}\left(m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}\right)}{\partial q_{it-1}} g_\eta(\eta_{it}), \tag{32}$$

where $\eta_{it} = \mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it}) - \bar{h}_t\left(m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}\right)$. Using Assumption 6(f), we can choose $q_{it-1} \in \{m_{it-1}, k_{it-1}, l_{it-1}, z^s_{it-1}, z^d_{it-1}, z^h_{it-1}, u_{it-1}\}$ and $(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}) \in \mathscr{A}_{q_{t-1}}$ such that $\partial G_{m_t|v_t}\left(m_t|k_t, l_t, z_t, u_t, \tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right) / \partial q_{t-1} \neq 0$ for all $(m_t, k_t, l_t, z_t, u_t) \in \mathscr{M} \times \mathscr{K} \times \mathscr{L} \times \mathscr{Z} \times [0, 1]$.

Dividing (31) by (32), we derive

$$\frac{\partial \mathbb{M}_t^{-1}(m_{it},w_{it},u_{it})}{\partial q_{it}} = -\frac{\partial \bar{h}\left(\tilde{m}_{it-1},\tilde{w}_{it-1},\tilde{u}_{it-1},\tilde{z}_{it-1}^h\right)}{\partial q_{it-1}} \times \frac{\partial G_{m_t|v_t}\left(m_{it}|w_{it},u_{it},\tilde{m}_{it-1},\tilde{w}_{it-1},\tilde{u}_{it-1},\tilde{z}_{it-1}^h\right)/\partial q_t}{\partial G_{m_t|v_t}\left(m_{it}|w_{it},u_{it},\tilde{m}_{it-1},\tilde{w}_{it-1},\tilde{u}_{it-1},\tilde{z}_{it-1}^h\right)/\partial q_{t-1}}. \tag{33}$$

Then, from (33) for $q_{it}=m_{it}$ and the normalization in Assumption 4, we obtain

$$\begin{aligned}1 &= \mathbb{M}_t^{-1}(m_{t1}^*,k_t^*,l_t^*,z_t^*,u_t^*) - \mathbb{M}_t^{-1}(m_{t0}^*,k_t^*,l_t^*,z_t^*,u_t^*)\\ &= -\frac{1}{S_{q_{t-1}}}\frac{\partial \bar{h}\left(\tilde{m}_{it-1},\tilde{w}_{it-1},\tilde{u}_{it-1},\tilde{z}_{it-1}^h\right)}{\partial q_{t-1}},\end{aligned} \tag{34}$$

where

$$S_{q_{t-1}} := \left(\int_{m_{t0}^*}^{m_{t1}^*}\frac{\partial G_{m_t|v_t}\left(m|w_t^*,u_t^*,\tilde{m}_{it-1},\tilde{w}_{it-1},\tilde{u}_{it-1},\tilde{z}_{it-1}^h\right)/\partial m_t}{\partial G_{m_t|v_t}\left(m|w_t^*,u_t^*,\tilde{m}_{it-1},\tilde{w}_{it-1},\tilde{u}_{it-1},\tilde{z}_{it-1}^h\right)/\partial q_{t-1}}dm\right)^{-1}.$$

Then, we identify $\partial \bar{h}\left(\tilde{m}_{it-1},\tilde{w}_{it-1},\tilde{u}_{it-1},\tilde{z}_{it-1}^h\right)/\partial q_{it-1} = -S_{q_{t-1}}$. Substituting this into (33), $\partial \mathbb{M}_t^{-1}(m_{it},w_{it},u_{it})/\partial q_{it}$ for $q_t \in \{m_{it},k_{it},l_{it},z_{it}^s,z_{it}^d,u_{it}\}$ are identified as follows:

$$\frac{\partial \mathbb{M}_t^{-1}(m_{it},w_{it},u_{it})}{\partial q_{it}} = S_{q_{t-1}}\frac{\partial G_{m_t|v_t}\left(m_{it}|w_{it},u_{it},\tilde{m}_{it-1},\tilde{w}_{it-1},\tilde{u}_{it-1},\tilde{z}_{it-1}^h\right)/\partial q_t}{\partial G_{m_t|v_t}\left(m_{it}|w_{it},u_{it},\tilde{m}_{it-1},\tilde{w}_{it-1},\tilde{u}_{it-1},\tilde{z}_{it-1}^h\right)/\partial q_{t-1}}. \tag{35}$$

Integrating (35) with respective to $q_{it} \in \{m_{it},k_{it},l_{it},z_{it}\}$ obtains for any $(m_t,k_t,l_t,z_t^s,z_t^d,u_t)$,

$$\begin{aligned}&\mathbb{M}_t^{-1}(m_t,k_t,l_t,z_t^s,z_t^d,u_t)\\ &= \int_{m_{t0}^*}^{m_t}\frac{\partial \mathbb{M}_t^{-1}(s,k_t,l_t,z_t^s,z_t^d,u_t)}{\partial m_{it}}ds + \int_{k_t^*}^{k_t}\frac{\partial \mathbb{M}_t^{-1}(m_{t0}^*,s,l_t,z_t^s,z_t^d,u_t)}{\partial k_{it}}\\ &+ \int_{l_t^*}^{l_t}\frac{\partial \mathbb{M}_t^{-1}(m_{t0}^*,k_t^*,s,z_t^s,z_t^d,u_t)}{\partial l_{it}}ds + \int_{z_t^{s*}}^{z_t^s}\frac{\partial \mathbb{M}_t^{-1}(m_{t0}^*,k_t^*,l_t^*,s,z_t^d,u_t)}{\partial z_{it}^s}ds\\ &+ \int_{z_t^{d*}}^{z_t^d}\frac{\partial \mathbb{M}_t^{-1}(m_{t0}^*,k_t^*,l_t^*,z_t^{s*},s,u_t)}{\partial z_{it}^d}ds + \int_{u_t^*}^{u_t}\frac{\partial \mathbb{M}_t^{-1}(m_{t0}^*,k_t^*,l_t^*,z_t^{s*},z_t^{d*},s)}{\partial u_{it}}ds\end{aligned} \tag{36}$$

where the equality follows from $\mathbb{M}_t^{-1}(m_{t0}^*,k_t^*,l_t^*,z_t^*,u_t^*)=0$ in Assumption 4. Substituting the identified derivatives of $\mathbb{M}_t^{-1}(\cdot)$ in (35) into (36), we can identify $\mathbb{M}_t^{-1}(m_t,k_t,l_t,z_t^s,z_t^d,u_t)$ for all $\left(m_t,k_t,l_t,z_t^s,z_t^d,u_t\right)$ on their supports.

Finally, from $\omega_{it} = \mathbb{M}_t^{-1}(m_{it},w_{it},u_{it})$, we can identify $\bar{h}_t(m_{it-1},w_{it-1},u_{it-1},z_{it-1}^h) = E\left[\omega_{it}|m_{it-1},w_{it-1},u_{it-1},z_{it-1}^h\right]$ and $\eta_{it} = \omega_{it} - \bar{h}_t(m_{it-1},w_{it-1},u_{it-1},z_{it-1}^h)$. Thus, we can identify the distribution of η_{it}, $G_{\eta_t}(\cdot)$. $\square$

3.3.3 Step 3: Identification of Production Function, Markup, and Demand Function

The final step identifies production function, markup, and demand function. From $\phi_t(m_{it},w_{it},u_{it}) = \varphi_t\left(f_t(x_{it},z^s_{it}) + \mathbb{M}_t^{-1}(m_{it},w_{it},u_{it}), z^d_{it}, u_{it}\right)$ and the monotonicity of φ_t, differentiating $\varphi_t^{-1}(\phi_t(m_{it},w_{it},u_{it}), z^d_{it}, u_{it}) = f_t(x_{it},z^s_{it}) + \mathbb{M}_t^{-1}(m_{it},w_{it},u_{it})$ with respect to $q^s_{it} \in \{m_{it},k_{it},l_{it},z^s_{it}\}$ and $q^d_{it} \in \{z^d_{it},u_{it}\}$ gives:

$$\frac{\partial \varphi_t^{-1}(r_{it},z_{it},u_{it})}{\partial r_{it}}\frac{\partial \phi_t(m_{it},w_{it},u_{it})}{\partial q^s_{it}} = \frac{\partial f_t(x_{it},z^s_{it})}{\partial q^s_{it}} + \frac{\partial \mathbb{M}_t^{-1}(m_{it},w_{it},u_{it})}{\partial q^s_t}, \tag{37}$$

$$\frac{\partial \varphi_t^{-1}(r_{it},z^d_{it},u_{it})}{\partial r_{it}}\frac{\partial \phi_t(m_{it},w_{it},u_{it})}{\partial q^d_{it}} = \frac{\partial \mathbb{M}_t^{-1}(m_{it},w_{it},u_{it})}{\partial q^d_{it}} - \frac{\partial \varphi_t^{-1}(r_{it},z^d_{it},u_{it})}{\partial q^d_t}. \tag{38}$$

Note that $\partial \varphi_t^{-1}(r_{it},z^d_{it},u_{it})/\partial r_t = \left(\partial \varphi_t(y_{it},z^d_{it},u_{it})/\partial y_t\right)^{-1}$ represents the markup from (1). If the markup $\partial \varphi_t^{-1}(r_{it},z^d_{it},u_{it})/\partial r_t$ were known, then equations (37) and (38) could identify $\partial f_t(x_t,z^s_{it})/\partial q^s_t$ and $\partial \varphi_t^{-1}(r_{it},z^d_{it},u_{it})/\partial q^d_t$ given that $\mathbb{M}_t^{-1}(m_{it},w_{it},u_{it})$ is identified. However, since the markup is unknown, identification requires further restriction. Following Gandhi et al. (2020), we use the first-order condition with respect to the material as an additional restriction.

Assumption 7. *The first-order condition with respect to material for the profit maximization problem (11)*

$$\frac{\partial f_t(x_{it},z^s_{it})}{\partial m_{it}} = \frac{\partial \varphi_t^{-1}(r_{it},z^d_{it},u_{it})}{\partial r_{it}}\frac{\exp(p^m_t + m_{it})}{\exp(r_{it})} \tag{39}$$

holds for all firms.

Rearranging the first-order condition, we obtain the markup equation used by De Loecker and Warzynski (2012):

$$\frac{\partial \varphi_t^{-1}(r_{it},z^d_{it},u_{it})}{\partial r_t} = \frac{\partial f_t(x_{it},z^s_{it})/\partial m_{it}}{\exp(p^m_t + m_{it})/\exp(r_{it})}. \tag{40}$$

We establish the following proposition.

Proposition 3. *Suppose that Assumptions 1–7 hold. Then, we can identify $\varphi_t^{-1}(\cdot)$, $f_t(\cdot)$, and $\psi_t(\cdot)$ up to scale and location and each firm's markup $\partial \varphi_t^{-1}(r_{it},z^d_{it},u_{it})/\partial r_t$ up to scale.*

Proof. From (37) and (39), the markup $\partial \varphi_t^{-1}(r_{it},z^d_{it},u_{it})/\partial r_t$ is identified as

$$\frac{\partial \varphi_t^{-1}(r_{it},z^d_{it},u_{it})}{\partial r_t} = \frac{\partial \mathbb{M}_t^{-1}(m_{it},w_{it},u_{it})}{\partial m_t}\left(\frac{\partial \phi_t(m_{it},w_{it},u_{it})}{\partial m_t} - \frac{\exp(p^m_t + m_{it})}{\exp(r_{it})}\right)^{-1}. \tag{41}$$

From ϕ_t and (41), the markup is also identified as a function of (m_t, w_t, u_t) as

$$\mu_t(m_t, w_t, u_t) := \frac{\partial \varphi_t^{-1}(\phi_t(m_t, w_t, u_t), z_t)}{\partial r_t}$$
$$= \frac{\partial \mathbb{M}_t^{-1}(m_t, w_t, u_t)}{\partial m_t} \left(\frac{\partial \phi_t(m_t, w_t, u_t)}{\partial m_t} - \frac{\exp(p_t^m + m_t)}{\exp(\phi_t(m_t, w_t, u_t))} \right)^{-1} \tag{42}$$

Substituting (42) into (37), we identify $\partial f_t(x_t, z_t^s)/\partial q_t$ for $q_t^s \in \{m_t, k_t, l_t, z_t^s\}$ as follows:

$$\frac{\partial f_t(x_t, z_t^s)}{\partial q_t} = \mu_t(m_t, w_t, u_t) \frac{\partial \phi_t(m_t, w_t, u_t)}{\partial q_t} - \frac{\partial \mathbb{M}_t^{-1}(m_t, w_t, u_t)}{\partial q_t}. \tag{43}$$

Using $f_t(m_{t0}^*, k_t^*, l_t^*, z_t^{s*}) = 0$ in Assumption 4, we identify $f_t(x_t)$ by integration:

$$f_t(m_t, k_t, l_t, z_t^s) = \int_{m_{t0}^*}^{m_t} \frac{\partial f_t(s, k_t, l_t, z_t^s)}{\partial m_t} ds + \int_{k_t^*}^{k_t} \frac{\partial f_t(m_{t0}^*, s, l_t, z_t^s)}{\partial k_t} ds$$
$$+ \int_{l_t^*}^{l_t} \frac{\partial f_t(m_{t0}^*, k_t^*, s, z_t^s)}{\partial l_t} ds + \int_{z_t^{s*}}^{z_t^s} \frac{\partial f_t(m_{t0}^*, k_t^*, l_t^*, s)}{\partial z_t^s} ds. \tag{44}$$

Let $\mathscr{R} := \{r_t : r_t = \phi_t(m_t, w_t, u_t) \text{ for some } (m_t, w_t, u_t) \in \mathscr{X} \times \mathscr{Z} \times [0,1]\}$ be the support of r_t. For given $(r_t, z_t^d) \in \mathscr{R} \times \mathscr{Z}_d$, $B_t(r_t, z_t^d, u_t) := \{(x_t, z_t^s) \in \mathscr{X} \times \mathscr{Z}_s : \phi_t(x_t, z_t^s, z_t^d, u_t) = r_t\}$ is non-empty by the construction of $\mathscr{R}$. Then, because $f_t(x_t, z_t^s)$ and $\mathbb{M}_t^{-1}(m_t, w_t, u_t)$ are identified, the output quantity $\varphi_t^{-1}(r_t, z_t, u_t)$ for any $(r_t, z_t, u_t) \in \mathscr{R} \times \mathscr{Z} \times [0,1]$ is identified by

$$\varphi_t^{-1}(r_t, z_t^d, u_t) = f_t(x_t, z_t^s) + \mathbb{M}_t^{-1}(m_t, w_t, u_t) \text{ for } (x_t, z_t^s) \in B_t(r_t, z_t^d, u_t).$$

By monotonicity, $\varphi_t(y_t, z_t^d, u_t)$ is identified from $\varphi_t^{-1}(r_t, z_t^d, u_t)$. Then, we can identify $\psi_t(y_t, z_t^d, u_t)$ as $\psi_t(y_t, z_t^d, u_t) = \varphi_t(y_t, z_t^d, u_t) - y_t$. □

The output quantity and price for individual firms are identified as $y_{it} = \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})$ and $p_{it} = \psi_t(y_{it}, z_{it}^d, u_{it}) = r_{it} - \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})$, respectively.

Corollary 1. *Suppose that Assumptions 1–7 hold. Then, the production function, the demand function, output quantities, output prices, and TFP are identified up to scale and location; markups and output elasticities are identified up to scale.*

Remark 1. Examination of the proofs reveals that we have over-identifying restrictions. In particular, the proof of Proposition 2 goes through with any choice of $q_{it-1} \in \{k_{it-1}, l_{it-1}, m_{it-1}, z_{it-1}^s, z_{it-1}^d, u_{it-1}, z_{it-1}^h\}$ in (35). Furthermore, the proof of Proposition 3 does not rely on the restriction in (38) for identifying $\varphi_t^{-1}(\cdot)$. These over-identifying restrictions can be exploited to construct specification tests for the model and to obtain more efficient estimation.

3.3.4 Comparison to Existing Identification Approaches

Our setup extends existing identification analyses of production functions by allowing prices to depend on output through an inverse demand function and by incorporating transitory unobserved demand shocks as a source of heterogeneous markups. While our approach builds on existing identification methods, our use of control functions and the IVQR framework differs from conventional formulations.

First, because the model includes both productivity and demand shocks, the standard control function approach cannot account for both sources of unobserved heterogeneity. We therefore assume that demand shocks are transitory while productivity shocks are persistent and use the IVQR approach to identify demand shocks in Step 1.

Second, Step 2 identifies the control function from the dynamics of input choices without relying on output measures, distinguishing our approach from the standard control function framework (e.g., Ackerberg et al., 2015).

Third, Ackerberg et al. (2015) identify a structural value-added function, $y_{it} = \tilde{f}_t(k_{it}, l_{it}) + \omega_{it}$, derived under perfect competition from a Leontief production function $y_{it} = \min\{\tilde{f}_t(k_{it}, l_{it}) + \omega_{it}, a + m_{it}\}$. This formulation is difficult to apply under imperfect competition because $y_{it} < \tilde{f}_t(k_{it}, l_{it}) + \omega_{it}$ can occur. The maximum output capacity $y^*_{it} := \tilde{f}_t(k_{it}, l_{it}) + \omega_{it}$ is determined before a firm chooses m_{it} and y_{it}, so when y^*_{it} is large—e.g., due to a high productivity shock—a profit-maximizing firm may produce $y_{it} < y^*_{it}$.[15] Intuitively, when TFP doubles, a firm may avoid a large price decline by expanding output less than proportionally.

Fourth, our approach differs from Gandhi et al. (2020) in the use of the first-order condition for materials. Their method identifies the material elasticity $\partial f_t(x_{it}, z^s_{it})/\partial m_{it}$ from the first-order condition (39): $\ln \frac{\partial f_t(x_{it}, z^s_{it})}{\partial m_{it}} = \ln \frac{\exp(p^m_t + m_{it})}{\exp(r_{it})} + \ln \frac{\partial \varphi_t^{-1}(r_{it}, z^d_{it}, u_{it})}{\partial r_t}$, assuming perfect competition where $\partial \varphi_t^{-1}(r_{it}, z^d_{it}, u_{it})/\partial r_{it} = 1$ for all i. Under imperfect competition, when the markup depends on revenue r_{it}, $\partial f_t(x_{it}, z^s_{it})/\partial m_{it}$ cannot be identified solely from this condition.

3.4 Fixing Normalization across Periods

Let $(\varphi_t^{-1}(\cdot), f_t(\cdot), \mathbb{M}_t^{-1}(\cdot))$ be a model structure for period t identified by using Propositions 2 and 3 under the normalization in Assumption 4. Let $(\varphi_t^{*-1}(\cdot), f_t^*(\cdot), \mathbb{M}_t^{*-1}(\cdot))$ denote the true model structure. Since the structure is identified up to scale and location normalization, there

[15] As noted by Ackerberg et al. (2015), under perfect competition $y_{it} < y^*_{it}$ implies zero output, so only firms with $y_{it} = y^*_{it}$ are observed. Under imperfect competition, however, positive output with $y_{it} < y^*_{it}$ is possible.

exist period-specific location and scale parameters $(a_{1t}, a_{2t}, b_t) \in \mathbb{R}^2 \times \mathbb{R}_+$ such as

$$\varphi_t^{-1}(r_{it}, z_{it}^d, u_{it}) = a_{1t} + a_{2t} + b_t \varphi_t^{*-1}(r_{it}, z_{it}^d, u_{it}),\ f_t(x_{it}, z_{it}^s) = a_{1t} + b_t f_t^*(x_{it}, z_{it}^s),$$
$$\mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}) = a_{2t} + b_t \mathbb{M}_t^{*-1}(m_{it}, w_{it}, u_{it}). \tag{45}$$

Generally speaking, the location and scale normalization differ across periods—that is, $(a_{1t}, a_{2t}, b_t) \neq (a_{1t+1}, a_{2t+1}, b_{t+1})$. For the identified objects to be comparable across periods, we need to fix normalization across periods by assuming that some object in the model is time-invariant. The subsection discusses these additional assumptions.[16]

3.4.1 Scale Normalization

From (45), the ratio of identified markups across two periods relates to the ratio of true markups as

$$\frac{\partial \varphi_{t+1}^{-1}(r_{it+1}, z_{it+1}^d, u_{it+1})/\partial r}{\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})/\partial r} = \frac{b_{t+1}}{b_t} \frac{\partial \varphi_{t+1}^{*-1}(r_{it+1}, z_{it+1}^d, u_{it+1})/\partial r}{\partial \varphi_t^{*-1}(r_{it}, z_{it}^d, u_{it})/\partial r}.$$

Therefore, the ability to identify how true markups change over two periods requires identification of the ratio of scale parameters, b_{t+1}/b_t. Similarly, the ratio of identified output elasticities across periods and that of identified TFP deviation from the mean are related to their true values via the ratio of scale parameters:

$$\frac{\partial f_{t+1}(x_{it+1}, z_{it+1}^s)/\partial q}{\partial f_t(x_{it}, z_{it}^s)/\partial q} = \frac{b_{t+1}}{b_t} \frac{\partial f_{t+1}^*(x_{it+1}, z_{it+1}^s)/\partial q}{\partial f_t^*(x_{it}, z_{it}^s)/\partial q}$$
$$\frac{\omega_{it+1} - E[\omega_{it+1}]}{\omega_{it} - E[\omega_{it}]} = \frac{b_{t+1}}{b_t} \left(\frac{\omega_{it+1}^* - E\left[\omega_{it+1}^*\right]}{\omega_{it}^* - E\left[\omega_{it}^*\right]} \right)$$

for $q \in \{m, k, l, z^s\}$.

To identify b_{t+1}/b_t, we consider the following assumptions.

Assumption 8. *At least one of the following conditions (a)–(c) holds. (a) The unconditional variance of η_{it} does not change over time. (b) For some known interval $\mathscr{B}$ of $\mathscr{X}$ and some known point $z^s \in \mathscr{Z}_s$, the output elasticity of one of the inputs evaluated at $z_{it}^s = z^s$does not change over time for all $x \in \mathscr{B}$. (c) For some known interval $\mathscr{B}$ of $\mathscr{X}$ and some known point $z^s \in \mathscr{Z}_s$, the*

[16] Klette and Griliches (1996) and De Loecker (2011) identify the levels of markups and output elasticities from revenue data by using a functional form property of a demand function. They consider a constant elastic demand function leading to $\varphi_t(y_{it}, z_{it}) = \alpha y_{it} - (\alpha - 1) z_{it}$ where z_{it} is an aggregate demand shifter, which is an weighted average of revenue across firms, and α is an unknown parameter. This formulation implies $\varphi_t^{-1}(r_{it}, z_{it}) = (1/\alpha) r_{it} + (1 - 1/\alpha) z_{it}$ and imposes a linear restriction $\partial \varphi_t^{-1}(r_{it}, z_{it})/\partial r_{it} + \partial \varphi_t^{-1}(r_{it}, z_{it})/\partial z_{it} = 1$, which fixes the scale parameter b_t.

sum of output elasticities of the three inputs evaluated at $z_{it}^s = z^s$ *does not change over time for all* $x \in \mathscr{B}$.

Assumption 8(a) holds, for example, if the productivity shock ω_{it} follows a stationary process because stationarity requires that the distribution of η_{it} does not change over time. Assumption 8(b) assumes that the elasticity of output with respect to one input does not change over time for some known interval; meanwhile, under Assumption 8(c), returns to scale in production technology does not change for some known interval of inputs.

Proposition 4. *Suppose that Assumptions 1–8 hold for time* t *and* $t+1$. *Then, we can identify the ratio of markups between two periods* t *and* $t+1$, *the ratio of output elasticities between* t *and* $t+1$, *and the ratio of TFP deviation from the mean between* t *and* $t+1$.

Proof. Suppose that Assumption 8(a) holds. Let $\text{var}(\eta_t)$ and $\text{var}(\eta_{t+1})$ be the variance of η_t and η_{t+1} identified under the period-specific normalization in Assumption 4 for t and $t+1$, respectively. From (30) and (45), $\text{var}(\eta_t) = b_t^2 \text{var}(\eta_t^*)$ and $\text{var}(\eta_{t+1}) = b_{t+1}^2 \text{var}(\eta_{t+1}^*)$. From $\text{var}(\eta_t^*) = \text{var}(\eta_{t+1}^*)$, b_{t+1}/b_t is identified as $b_{t+1}/b_t = \sqrt{\text{var}(\eta_t)/\text{var}(\eta_{t+1})}$.

Let $\partial f_t(x_t, z_t^s)/\partial q$ and $\partial f_{t+1}(x_{t+1}, z_{t+1}^s)/\partial q$ be those elasticities identified under the period-specific normalization in Assumption 4 for t and $t+1$, respectively, and $\partial f_t^*(x_t, z_t^s)/\partial q$ and $\partial f_{t+1}^*(x_{t+1}, z_{t+1}^s)/\partial q$ be the true elasticities. From (45), $\partial f_t(x_t, z_t^s)/\partial q = b_t \partial f_t^*(x_t, z_t^s)/\partial q$ and $\partial f_{t+1}(x_{t+1}, z_{t+1}^s)/\partial q = b_{t+1} \partial f_{t+1}^*(x_{t+1}, z_{t+1}^s)/\partial q$ hold.

Suppose that Assumption 8(b) holds. Then, $\partial f_t^*(x, z^s)/\partial q = \partial f_{t+1}^*(x, z^s)/\partial q$ for some input $q \in \{m, k, l\}$ and $x \in \mathscr{B}$. Then, b_{t+1}/b_t is identified as $b_{t+1}/b_t = (\partial f_{t+1}(x, z^s)/\partial q)/(\partial f_t(x, z^s)/\partial m)$ for $x \in \mathscr{B}$.

Suppose that Assumption 8(c) holds, implying

$$1 = \frac{\partial f_{t+1}^*(x, z^s)/\partial m + \partial f_{t+1}^*(x, z^s)/\partial k + \partial f_{t+1}^*(x, z^s)/\partial l}{\partial f_t^*(x, z^s)/\partial m + \partial f_t^*(x, z^s)/\partial k + \partial f_t^*(x, z^s)/\partial l} \text{ for } x \in \mathscr{B}.$$

Then, b_{t+1}/b_t is identified as

$$\frac{b_{t+1}}{b_t} = \frac{\partial f_{t+1}(x, z^s)/\partial m + \partial f_{t+1}(x, z^s)/\partial k + \partial f_{t+1}(x, z^s)/\partial l}{\partial f_t(x, z^s)/\partial m + \partial f_t(x, z^s)/\partial k + \partial f_t(x, z^s)/\partial l} \quad \text{for } x \in \mathscr{B}.$$

□

3.4.2 Local Constant Returns to Scale

We consider the following local constant returns to scale that strengthens Assumption 8(c).

Assumption 9. *(Local Constant Returns to Scale) For some known interval $\mathscr{B}$ of $\mathscr{X}$ and some known point $z^s \in \mathscr{Z}_s$, the sum of the output elasticities of the three inputs evaluated at $z^s_{it} = z^s_t$ equals to 1 for all $x \in \mathscr{B}$.*

Assumption 9 is stronger than Assumption 8(c) but weaker than those used in some studies of markup estimation. In particular, markups are often estimated as the ratio of revenue $\exp(r_{it})$ to total cost TC_{it} under the assumption of a linear cost function $TC_{it} = MC_{it} y_{it}$ with constant marginal cost MC_{it}. Such a linear cost function requires stronger conditions than Assumption 9: (i) global constant returns to scale for all $x \in \mathscr{X}$, (ii) full flexibility of all inputs, and (iii) price-taking behavior in all input markets. By contrast, under Assumption 9, marginal cost may increase with output, especially in the short run when dynamic inputs such as capital entail adjustment costs.

With Assumption 9, the scale normalization parameter b_t can be identified for all periods as follows. Let $f_t(x_t, z^s_t)$ be the identified production function under Assumption 4 and $f^*_t(x_t)$ be the true one where $f_t(x_t, z^s_t) = a_t + b_t f^*_t(x_t, z^s_t)$ from (45). For $x \in \mathscr{B}$, we have

$$b_t = b_t \left(\frac{\partial f^*_t(x, z^s)}{\partial m} + \frac{\partial f^*_t(x, z^s)}{\partial k} + \frac{\partial f^*_t(x, z^s)}{\partial l} \right) = \frac{\partial f_t(x, z^s)}{\partial m} + \frac{\partial f_t(x, z^s)}{\partial k} + \frac{\partial f_t(x, z^s)}{\partial l}.$$

Given that we have identified the scale parameter b_t in (45), we have established the following proposition.

Proposition 5. *Suppose that Assumptions 1–7 and 9 hold. Then, $\varphi_t(\cdot)$, $f_t(\cdot)$, and $\psi_t(\cdot)$ can be identified up to location. The levels of markup and output elasticities can be identified. Output quantity, output price, and TFP can be identified up to location.*

3.4.3 Location Normalization

Suppose that scale normalization b_t is already identified—for example, from Proposition 5. Define

$$\tilde{\varphi}^{-1}_t(r_{it}, z^d_{it}, u_{it}) := \varphi^{-1}_t(r_{it}, z^d_{it}, u_{it})/b_t,\ \tilde{f}_t(x_{it}, z^s_{it}) := f_t(x_{it}, z^s_{it})/b_t,\ \tilde{\omega}_{it} := \omega_{it}/b_t,$$
$$\tilde{a}_{1t} := a_{1t}/b_t, \text{ and } \tilde{a}_{2t} := a_{2t}/b_t. \tag{46}$$

Then, (45) is written as

$$\tilde{\varphi}^{-1}_t(r_{it}, z^d_{it}, u_{it}) = \tilde{a}_{1t} + \tilde{a}_{2t} + \varphi^{*-1}_t(r_{it}, z^d_{it}, u_{it}),\ \tilde{f}_t(x_{it}, z^s_{it}) = \tilde{a}_{1t} + f^*_t(x_{it}, z^s_{it}),\ \tilde{\omega}_{it} = \tilde{a}_{2t} + \omega^*_{it}. \tag{47}$$

From (45), the growth rates (log differences) of the identified output and TFP between t and $t+1$ are related to their true values as follows:

$$\begin{aligned}
\tilde{\varphi}_{t+1}^{-1}(r_{it+1}, z_{it+1}^d, u_{it+1}) - \tilde{\varphi}_t^{-1}(r_{it}, z_{it}^d, u_{it}) &= \tilde{a}_{1t+1} + \tilde{a}_{2t+1} - \tilde{a}_{1t} - \tilde{a}_{2t} \\
&\quad + \varphi_{t+1}^{*-1}(r_{it+1}, z_{it+1}^d, u_{it+1}) - \varphi_t^{*-1}(r_{it}, z_{it}^d, u_{it}), \\
\tilde{f}_{t+1}(x_{it+1}, z_{it+1}^s) - \tilde{f}_t(x_{it}, z_{it}^s) &= \tilde{a}_{1t+1} - \tilde{a}_{1t} + f_t^*(x_{it+1}, z_{it+1}^s) - f_t^*(x_{it}, z_{it}^s), \\
\tilde{\omega}_{it+1} - \tilde{\omega}_{it} &= \tilde{a}_{2t+1} - \tilde{a}_{2t} + \omega_{it+1}^* - \omega_{it}^*. \qquad (48)
\end{aligned}$$

Therefore, to identify the growth rates of output and TFP, we need to identify the changes in the location parameters. To do so, we can use an industry-level producer price index P_t^*, which is often available as data, to identify the change in the location parameters. Suppose that P_t^* is a Laspeyres index

$$P_t^* := \frac{\sum_{i \in \tilde{N}} \exp(p_{it}^* + y_{i0}^*)}{\sum_{i \in \tilde{N}} \exp(p_{i0}^* + y_{i0}^*)}, \qquad (49)$$

where $\tilde{N}$ is a known set (or a random sample) of products. p_{i0}^* and y_{i0}^* are firm i's log true price and log true output at the base period, respectively. The following argument holds for forms of a price index (other than Laspeyres) as long as the price index is a known function of prices that is homogenous of degree 1, which is typically satisfied.

Assumption 10. *(a) The industry-level producer price index P_t^* is known as data. (b) For some known point $(\bar{x}, \bar{z}^s) \in \mathcal{X} \times \mathcal{Z}_s$ and the true production functions of t and $t+1$, $f_t^*(\cdot)$ and $f_{t+1}^*(\cdot)$, satisfy $f_t^*(\bar{x}, \bar{z}^s) = f_{t+1}^*(\bar{x}, \bar{z}^s)$.*

Assumption 10(b) is innocuous, implying that any output change between t and $t+1$ when inputs are fixed at $\bar{x}$ is attributed to a TFP change.

Using the aggregate price index, we can identify the change in the location parameters and identify the growth of TFP and output.

Proposition 6. *Suppose Assumptions 1–7, 9, and 10 hold. Then, the true growth rate of output $\varphi_{t+1}^{*-1}(r_{it+1}, z_{it+1}^d, u_{it+1}) - \varphi_t^{*-1}(r_{it}, z_{it}^d, u_{it})$ and that of TFP $\omega_{it+1}^* - \omega_{it}^*$ can be identified for each firm.*

Proof. Let $\tilde{p}_{it} := r_{it} - \tilde{\varphi}_t^{-1}(r_{it}, z_{it}^d, u_{it})$ and $\tilde{y}_{it} := \tilde{\varphi}_t^{-1}(r_{it}, z_{it}^d, u_{it})$ be an output price and an output quantity identified under the normalization in (46) and Assumption 4, respectively. Using these, we calculate an industry-level producer price index with them:

$$P_t := \frac{\sum_{i \in \tilde{N}} \exp(\tilde{p}_{it} + \tilde{y}_{i0})}{\sum_{i \in \tilde{N}} \exp(\tilde{p}_{i0} + \tilde{y}_{i0})}.$$

From (47) and (49), P_t is written as

$$P_t = \frac{\sum_{i\in\tilde{N}} \exp(-(\tilde{a}_{1t}+\tilde{a}_{2t}) + p^*_{it} + \tilde{a}_{1,0} + \tilde{a}_{2,0} + y^*_{i0})}{\sum_{i\in\tilde{N}} \exp(p^*_{i0} + y^*_{i0})}$$
$$= \exp(\tilde{a}_{1,0} + \tilde{a}_{2,0} - (\tilde{a}_{1t}+\tilde{a}_{2t}))P^*_t.$$

Therefore, $\tilde{a}_{1t+1} + \tilde{a}_{2t+1} - \tilde{a}_{1t} - \tilde{a}_{2t}$ is identified as:

$$\tilde{a}_{1t+1} + \tilde{a}_{2t+1} - \tilde{a}_{1t} - \tilde{a}_{2t} = \ln P^*_{t+1} - \ln P_{t+1} - \left(\ln P^*_t - \ln P_t\right) \tag{50}$$

From (48), we identify the output growth rate $\varphi^{*-1}_{t+1}(r_{it+1}, z^d_{it+1}, u_{it+1}) - \varphi^{*-1}_t(r_{it}, z^d_{it}, u_{it})$.

Evaluating the second equation in (48) at $x_{t+1} = x_t = \bar{x}$ and $z^s_{t+1} = z^s_t = \bar{z}^s$in Assumption 10(b), we identify $\tilde{a}_{1t+1} - \tilde{a}_{1t}$ as:

$$\tilde{a}_{1t+1} - \tilde{a}_{1t} = \tilde{a}_{1,t+1} + f^*_{t+1}(\bar{x}, \bar{z}^s) - \left(\tilde{a}_{1,t} + f^*_t(\bar{x}, \bar{z}^s)\right)$$
$$= \tilde{f}_{t+1}(\bar{x}, \bar{z}^s) - \tilde{f}_t(\bar{x}, \bar{z}^s).$$

From (50), $\tilde{a}_{2t+1} - \tilde{a}_{2t}$ is also identified as

$$\tilde{a}_{2t+1} - \tilde{a}_{2t} = \ln P^*_{t+1} - \ln P_{t+1} - \left(\ln P^*_t - \ln P_t\right) - \left(\tilde{f}_{t+1}(\bar{x}, \bar{z}^s) - \tilde{f}_t(\bar{x}, \bar{z}^s)\right).$$

Therefore, from (48), the true TFP growth rate $\omega^*_{it+1} - \omega^*_{it}$ is also identified. □

3.5 Identification of HSA Demand System and Utility Function

Given that we have identified each firm's output price and quantity, it is possible to identify with additional assumptions a system of demand functions and a homothetic utility function of a representative consumer. The identified demand system and the identified utility function can be used to undertake counterfactual analysis and welfare analysis.

3.5.1 HSA demand system

We consider an HSA (homothetic single aggregator) system (**?**) of inverse demand functions.[17] Let N_t be the number of firms in the industry and $\Phi_t := \ln\left(\sum_{i=1}^{N_t} \exp(r_{it})\right)$ be the log of industry expenditure. The HSA system of inverse demand functions for products $i = 1, ..., N_t$ in an

[17] The HSA system can be expressed as a system of direct demand functions or of inverse demand functions. The two systems are self-dual in the sense that either can be derived from the other. **?** and **?** provide excellent reviews on flexible extensions of the CES demand system, including the HSA demand system.

industry is expressed as:

$$p_{it} = s_t^*\left(y_{it} - q_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right) + \Phi_t - y_{it} \text{ for } i = 1, ..., N_t$$

where Φ_t is the log of consumer expenditure (budget) on the industry, $s_t(\cdot, z_{it}^d, u_{it})$ provides the log of the market (budget) share of product i, $\mathbf{y}_t := (y_{1t}, ..., y_{N_t t}) \in \mathscr{Y}^{N_t}$ is a vector of consumption, $\mathbf{z}_t^d := (z_{1t}^d, ..., z_{N_t t}^d)$ is a vector of observable demand shifters and $\mathbf{u}_t := (u_{1t}, ..., u_{N_t t})$ is a vector of demand shocks, and $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ is the aggregate quantity index summarizing interactions across products[18]. Note that $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ is uniquely defined by the market share constraint:

$$1 = \sum_{i=1}^{N_t} \exp\left(s_t^*\left(y_{it} - q_t(\mathbf{y}_t, z_t^d, u_t), z_t^d, u_{it}\right)\right). \tag{51}$$

Since $s_t^*(\cdot)$ is nonparametric, the HSA demand system can nest various demand systems used in the literature such as the CES demand system and the symmetric translog demand system (Feenstra, 2003; Feenstra and Weinstein, 2017).[19]

3.5.2 Identification of the HSA demand system

For identification of a demand system, we make the following assumptions.

Assumption 11. *(a) The good market is monopolistically competitive (without free entry)—that is, each firm takes the quantity index $q_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)$ as given. (b) $\varphi_t^{-1}(r_t, z_t^d, u_t)$ is identified up to location.*

First, the assumption of monopolistic competition follows Klette and Griliches (1996) and De Loecker (2011), with the inverse demand function becoming a symmetric function of the firm's own output, as in (9). Second, the demand elasticity equals $(\mu - 1)/\mu$ when μ is markup. To identify the demand elasticity, we need to fix the scale normalization. Assumption 11 (b) is satisfied when Proposition 5 holds.

An HSA demand system can be identified as follows. Suppose variables in data are in an "initial equilibrium" and we identify the "reduced-form" revenue function $r_{it} = \varphi_t(y_{it}, z_{it}^d, u_{it})$ from the data. Let Φ_t be the total industry expenditure, $\Phi_t = \ln\left(\sum_{i=1}^{N_t} \exp(r_{it})\right)$. We first define

[18] If the utility function is CES, $U(\mathbf{y}_t) = \left[\sum_{i=1}^{N_t} \exp(\rho y_{it})\right]^{1/\rho}$, the inverse demand becomes $p_{it} = \rho\left(y_{it} - \ln U(\mathbf{y}_t)\right) + \Phi_t - y_{it}$, and the quantity index is the same as the utility function, but they are generally different.

[19] See Matsuyama and Ushchev (2020) regarding how the HSA nests the translog demand.

the reduced-form budget share equation as:

$$s_t\left(\cdot, z_{it}^d, u_{it}\right) \equiv \varphi_t\left(\cdot, z_{it}^d, u_{it}\right) - \Phi_t.$$

The structural and reduced-form functions are in the following relationship:

$$s_t\left(y_{it}, z_{it}^d, u_{it}\right) = s_t^*\left(y_{it} - q_{0t}, z_{it}^d, u_{it}\right) \tag{52}$$

where q_{0t} is the level of the quantity index in the initial equilibrium. For a given output vector $\mathbf{y}_t$, we define $\tilde{q}_t(\mathbf{y}_t, z_t^d, u_t)$ by using the reduced-form revenue function:

$$1 = \sum_{i=1}^{N} \exp\left(s_t\left(y_{it} - \tilde{q}_t(\mathbf{y}_t, z_t^d, u_t), u_{it}\right)\right). \tag{53}$$

Then, we obtain a system of inverse demand functions for products $i = 1, ..., N_t$ in an industry:

$$p_{it} = s_t\left(y_{it} - \tilde{q}_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right) + \Phi_t - y_{it} \text{ for } i = 1, ..., N_t$$

From (52), $\tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t) = q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t) - q_{0t}$. Thus, we can identify the change in the quantity index from the initial equilibrium. Since

$$s_t\left(y_{it} - \tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t), z_{it}, u_{it}\right) = s_t^*\left(y_{it} - q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t), z_{it}^d, u_{it}\right),$$

we can use $\left\{s_t(\cdot,), \tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t), \Phi_t\right\}$ constructed above to obtain the value of the structural HSA demand system $\left\{s_t^*(\cdot), q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t), \Phi_t\right\}$.

Applying the result of Matsuyama and Ushchev (2017, Proposition 1 and Remark 3), the following proposition establishes that the HSA demand system constructed above can be derived from a unique consumer preference, and that it is possible to identify an associated utility function. Appendix A.3 supplies the proof.

Proposition 7. *Suppose Assumption 11 holds. (a) There exists a unique monotone, convex, and homothetic rational preference $\succsim$ over $\bar{\mathscr{Y}}$ that generates an HSA demand system $\left\{\varphi_t(\cdot), \tilde{q}_t(\mathbf{y}_t, z_t^d, u_t), \Phi_t\right\}$. (b) This preference $\succsim$ is represented by a homothetic utility function defined by*

$$\ln U_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t) = \tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t) + \sum_{i=1}^{N_t} \int_c^{y_{it} - \tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)} \exp\left(s_t\left(\zeta, z_{it}^d, u_{it}\right)\right) d\zeta$$

for some constant $c > 0$. (c) The identified demand system and preference $\succsim$ do not depend on the *location normalization of* $\varphi_t^{-1}(r_t, z_t^d, u_t)$.

3.5.3 Counterfactual analysis

We conduct a short-run partial equilibrium counterfactual analysis where firms change materials, outputs, and prices, while pre-determined factor inputs (k_{it}, l_{it}), factor prices, and exogenous variables $(z_{it}^d, z_{it}^s, u_{it}, \omega_{it}, p_{mt})$ are fixed.

Monopolistic Competition Equilibrium Using the identified HSA demand system $\{\varphi_t(\cdot), \tilde{q}_t(\mathbf{y}_t, z_t^d, u_t)\}$, we can calculate a monopolistic competition equilibrium (MCE). Define $m_{it} = \chi_{it}(y_{it})$ such that $y_{it} = f_t\left(\chi_{it}(y_{it}), k_{it}, l_{it}, z_{it}^s\right) + \omega_{it}$ for given $(k_{it}, l_{it}, \omega_{it}, z_{it}^s)$. Equilibrium outputs and quantity index $(\mathbf{y}_t^m, \tilde{q}_t^m)$ in an MCE are obtained from the log of the first order condition for (11) and the market share condition (53) as follows:

$$
\begin{aligned}
s_t\left(y_{it}^m - \tilde{q}_t^m, z_{it}^d, u_{it}\right) + \Phi_t + \ln \frac{\partial s_t\left(y_{it}^m - \tilde{q}_t^m, z_{it}^d, u_{it}\right)}{\partial y_{it}} \\
+ \ln \frac{\partial f_t\left(\chi_{it}\left(y_{it}^m\right), k_{it}, l_{it}\right)}{\partial m_{it}} - p_{mt} - \chi_{it}\left(y_{it}^m\right) = 0 \text{ for } i = 1, \ldots, N_t \\
\sum_{i=1}^{N_t} \exp\left(s_t\left(y_{it}^m - \tilde{q}_t^m, z_{it}^d, u_{it}\right)\right) = 1.
\end{aligned}
\tag{54}
$$

The above system can be extended to incorporate policies such as tax and subsidies to investigate their effects.

Welfare Costs of Firm's Market Power In an empirical section below, we quantify the deadweight loss attributable to firm's market power by considering the transition to a counterfactual marginal cost pricing equilibrium (MCPE) i.e. perfect competition. In a MCPE, each firm sets its price equal to its marginal cost. Equilibrium outputs and quantity index $(\mathbf{y}_t^c, \tilde{q}_t^c)$ are obtained from the first order conditions, which sets the markup to one, $\left(\frac{\partial s_t\left(y_{it}^m - \tilde{q}_t^m, z_{it}^d, u_{it}\right)}{\partial y_{it}}\right)^{-1} = 1$, in (54), and the market share condition (53) as follows:

$$
\begin{aligned}
s_t\left(y_{it}^c - \tilde{q}_t^c, z_{it}^d, u_{it}\right) + \Phi_t^c \\
+ \ln \frac{\partial f_t\left(\chi_{it}\left(y_{it}^c\right), k_{it}, l_{it}\right)}{\partial m_{it}} - p_{mt} - \chi_{it}\left(y_{it}^c\right) = 0 \text{ for } i = 1, \ldots, N_t \\
\sum_{i=1}^{N_t} \exp\left(s_t\left(y_{it}^c - \tilde{q}_t^c, z_{it}^d, u_{it}\right)\right) = 1.
\end{aligned}
\tag{55}
$$

The consumer welfare cost of firm's market power can be calculated as the utility change:

$$\ln U_t(\mathbf{y}_t^c, \mathbf{z}_t^d, \mathbf{u}_t) - \ln U_t(\mathbf{y}_t^m, \mathbf{z}_t^d, \mathbf{u}_t) = \tilde{q}_t(\mathbf{y}_t^c, \mathbf{z}_t^d, \mathbf{u}_t) - \tilde{q}_t(\mathbf{y}_t^m, \mathbf{z}_t^d, \mathbf{u}_t) + \sum_{i=1}^{N_t} \int_{y_{it}^m - \tilde{q}_t(\mathbf{y}_t^m, \mathbf{z}_t^d, \mathbf{u}_t)}^{y_{it}^c - \tilde{q}_t(\mathbf{y}_t^c, \mathbf{z}_t^d, \mathbf{u}_t)} \exp\left(s_t\left(\zeta, z_{it}^d, u_{it}\right)\right) d\zeta.$$

An alternative welfare measure is the compensation variation, which is expressed in monetary term. Solving (55) for a given counterfactual log income Φ_t^c, we can express outputs $y_{it}^c(\Phi_t^c)$ in the MCPE as functions of Φ_t^c. Then, we find a counterfactual income Φ_t^{c*} that achieve the same utility as in the benchmark MCE.

$$\ln U_t(\mathbf{y}_t^c(\Phi_t^{c*}), \mathbf{z}_t^d, \mathbf{u}_t) - \ln U_t(\mathbf{y}_t^m, \mathbf{z}_t^d, \mathbf{u}_t) = 0. \tag{56}$$

Then, the compensation variation CV_t is obtained as $CV_t \equiv \exp\left(\Phi_t^c\right) - \exp\left(\Phi_t^m\right)$.

To evaluate the overall welfare change, the consumer's loss may be compared with firms' profit loss. The change in the total profits is the change in the total revenue minus the total material costs:

$$\Pi^c - \Pi^m \equiv \left(\Phi_t^c - \Phi_t^m\right) - \exp\left(p_{mt}\right) \sum_{i=1}^{N_t} \left\{\exp\left(\chi_{it}\left(y_{it}^c\right)\right) - \exp\left(\chi_{it}\left(y_{it}^m\right)\right)\right\}. \tag{57}$$

3.6 Endogenous Labor Input

Identification is possible when a firm chooses l_{it} after observing ω_{it} and u_{it}. In the spirits of Ackerberg et al. (2015) and the dynamic generalized method of moment approach (e.g., Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998, 2000), we provide identification using lagged labor l_{it-1}. Specifically, we assume a firm incurs an adjustment cost of labor input, e.g., costs of recruiting and training new workers. The profit maximization problem for choosing l_{it} and m_{it} at time t is

$$\max_{m,l} \exp(\varphi_t(f_t(m, l, k_{it}, z_{it}^s) + \omega_{it}, z_{it}^d, u_{it})) - \exp(p_t^m + m) - \exp(p_t^l + l) - C(l, l_{it-1}), \tag{58}$$

where p_t^l is the wage and $C(l_{it}, l_{it-1})$ is the adjustment costs.The solution to the problem provides a material demand function $m_{it} = \tilde{\mathbb{M}}_t(\omega_{it}, l_{it-1}, s_{it}, u_{it})$ and a labor demand function $l_{it} = \tilde{\mathbb{L}}_t(\omega_{it}, l_{it-1}, s_{it}, u_{it})$, where $s_{it} := (k_{it}, z_{it}^s, z_{it}^d)$. We also consider a "conditional" material demand function $m_{it} = \mathbb{M}_t(\omega_{it}, l_{it}, s_{it}, u_{it})$ when l_{it} is given, which solves the conditional problem (11).

We assume both $\tilde{\mathbb{M}}_t(\cdot, l_{it-1}, s_{it}, u_{it})$ and $\mathbb{M}_t(\cdot, l_{it}, s_{it}, u_{it})$ are monotonically increasing functions so that there exist their inverse functions

$$\omega_{it} = \mathbb{M}_t^{-1}(m_{it}, l_{it}, s_{it}, u_{it}) = \tilde{\mathbb{M}}_t^{-1}(m_{it}, l_{it-1}, s_{it}, u_{it}).$$

In the first step, we substitute $\omega_{it} = \mathbb{M}_t^{-1}(\cdot)$ into the revenue function to obtain

$$\begin{aligned} r_{it} &= \varphi_t(f_t(m_{it}, l_{it}, k_{it}, z_{it}^s) + \mathbb{M}_t^{-1}(m_{it}, l_{it}, s_{it}, u_{it}), z_{it}^d, u_{it})) \\ &= \phi_t(m_{it}, l_{it}, s_{it}, u_{it}). \end{aligned}$$

The first step identification is

$$\Pr(r_{it} \leq \phi_t(m_{it}, l_{it}, s_{it}, u) | m_{it-v-1}, l_{it-v-1}, s_{it-v}) = u.$$

The IVQR identifies $\phi(\cdot)$ and u_{it}.

In the second step, we formulate a transformation model using $\omega_{it} = \tilde{\mathbb{M}}_t^{-1}(\cdot)$:

$$\begin{aligned} \tilde{\mathbb{M}}_t^{-1}(m_{it}, l_{it-1,} s_{it}, u_{it}) &= h\left(\tilde{\mathbb{M}}_{t-1}^{-1}(m_{it-1}, l_{it-2}, s_{it-1}, u_{it-1}), z_{it}^h\right) + \eta_{it} \\ &= \bar{h}_t\left(m_{it-1}, l_{it-2}, s_{it-1}, u_{it-1}, z_{it}^h\right) + \eta_{it} \end{aligned}$$

Since η_{it} is independent of $v_{it} \equiv (k_{it}, l_{it-1,}, s_{it}, u_{it}, m_{it-1}, k_{it-1}, l_{it-2}, s_{it-1}, u_{it-1}, z_{it-1}^h)$, the conditional CDF of m_{it} on v_{it-1} becomes

$$\begin{aligned} G_{m_t|v_t}(m|v_t) &= G_{\eta_t|v_t}\left(\tilde{\mathbb{M}}_t^{-1}(m_{it}, l_{it-1,} s_{it}, u_{it}) - \bar{h}_t\left(m_{it-1}, l_{it-2}, s_{it-1}, u_{it-1}, z_{it-1}^h\right)|v_t\right) \\ &= G_{\eta_t}\left(\tilde{\mathbb{M}}_t^{-1}(m_{it}, l_{it-1,} s_{it}, u_{it}) - \bar{h}_t\left(m_{it-1}, l_{it-2}, s_{it-1}, u_{it-1}, z_{it-1}^h\right)\right). \end{aligned}$$

Following the same logic of the main text, we can identify $\tilde{\mathbb{M}}_t^{-1}(\cdot)$ and ω_{it} under scale and location normalization. Once we identify u_{it} and ω_{it}, we can also identify $\omega_{it} = \mathbb{M}_t^{-1}(m_{it}, l_{it}, s_{it}, u_{it})$, e.g., by regression $\mathbb{M}_t^{-1}(m_{it}, l_{it}, s_{it}, u_{it}) = E[\omega_{it}|m_{it}, l_{it}, s_{it}, u_{it}]$.

Differentiating $\varphi_t^{-1}(\phi_t(m_{it}, w_{it}, u_{it}), z_{it}^d, u_{it}) = f_t(x_{it}, z_{it}^s) + \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})$ with respect to $q_{it}^s \in \{m_{it}, k_{it}, l_{it}, z_{it}^s\}$ and $q_{it}^d \in \{z_{it}^d, u_{it}\}$ gives the same equations as (37) and (38). Therefore, Proposition 3 holds with the same proof as before.

Role of Adjustment Costs The role of adjustment costs is to create variations in l_{it} for given $(m_{it}, k_{it}, z_{it}, u_{it})$. When l_{it} is a fully flexible input without adjustment costs, the material demand function and the labor demand function become $m_{it} = \mathbb{M}_t^F(\omega_{it}, k_{it}, z_{it}, u_{it})$ and $l_{it} = \mathbb{L}_t^F(\omega_{it}, k_{it}, z_{it}, u_{it})$, respectively. Once $(m_{it}, k_{it}, z_{it}, u_{it})$ are conditioned, $\omega_{it}$is also conditioned so that l_{it} loses its variation and we cannot identify $\phi_t(m_{it}, l_{it}, k_{it}, z_{it}, u_{it})$.

4 Semiparametric Estimator

We develop a semiparametric estimator that is applicable with $T \geq 4$. We assume the Cobb-Douglas production function:

$$f_t(m_{it}, k_{it}, l_{it}) = \theta_m m_{it} + \theta_k k_{it} + \theta_l l_{it}, \tag{59}$$

and TFP follows an AR(1) process:

$$\omega_{it} = \rho \omega_{it-1} + \eta_{it}. \tag{60}$$

The control function becomes separable:

$$\mathbb{M}_t^{-1}(m_{it}, k_{it}, l_{it}, u_{it}) = \lambda_t(m_{it}, u_{it}) - \theta_k k_{it} - \theta_l l_{it}. \tag{61}$$

Substituting (61) into the revenue function, we obtain

$$\varphi_t\left(\omega_{it} + f_t(m_{it}, k_{it}, l_{it}), u_{it}\right) = \varphi_t\left(\theta_m m_{it} + \lambda_t(m_{it}, u_{it}), u_{it}\right) = \phi_t(m_{it}, u_{it}), \tag{62}$$

where ϕ_t depends only on (m_t, u_{it}) and increases in m_t and u_t.

The second step transformation model becomes

$$\lambda_t(m_{it}, u_{it}) = \theta_k k_{it} + \theta_l l_{it} + \rho \lambda_{t-1}(m_{it-1}, u_{it-1}) - \rho \theta_k k_{it-1} - \rho \theta_l l_{it-1} + \eta_{it}. \tag{63}$$

Step 1: Estimation of the Quantile of Demand Shocks The first step estimates $\phi_t(m_{it}, u_{it})$ and u_{it} by IV quantile regression. A traditional approach to IV quantile regression estimates $\phi_t(\cdot, u)$ from the moment condition (29) for a fixed quantile point u. This approach often yields a non-monotonic and non-smooth function in u, which is problematic for our identification using uniquely identified u_{it} and derivatives of ϕ_t. To overcome this, we use the smoothed GMM quantile regression of Firpo, Galvao, Pinto, Poirier, and Sanroman (2022). Their approach stacks moment conditions over all quantile points so we can estimate the smooth sieve function and impose $\partial \phi_t / \partial u_{it} > 0$. For the approximation of $\phi_t(m_{it}, u_{it})$, we employ the basis $B_\phi(m_{it}, \tau)$ that consists of a constant term, a B-spline basis of degree 3 with 2 interior knots in m_{it}, a cubic polynomial in u_{it}, and interactions of the B-spline in m_{it} with u_{it} and u_{it}^2. Firpo et al. (2022) also replace the indicator in (29) with a smooth kernel CDF to ease computation.

We partition $[0, 1]$ into L equal parts and let $\mathbb{T} \equiv \{\tau_1, \ldots, \tau_{L-1}\}$ (e.g. $\mathbb{T} = \{0.01, \ldots, 0.99\}$

for $L = 100$). The moment condition is

$$E\left[\left(K_1\left(\frac{B_\phi(m_{it},\tau)^T\alpha_t - r_{it}}{b_{n1}}\right) - \tau\right)B_{IV}(m_{it-v})\right] = 0 \quad \text{for } \tau \in \mathbb{T}, \tag{64}$$

where $K_1(\cdot)$ is a smooth kernel CDF with bandwidth b_{n1} and $B_{IV}(m_{it-v}) := B_{S_1,K_1}(m_{it-v})$ is the sieve basis of instruments. We use the B-spline basis of degree 3 with 2 interior knots in m_{it-2} as instruments. Following Firpo et al. (2022), we use the rule-of-thumb bandwidth and the kernel CDF of Horowitz (1998):

$$K_1(s) := \left[\frac{1}{2} + \frac{105}{64}\left(s - \frac{5}{3}s^3 + \frac{7}{5}s^5 - \frac{3}{7}s^7\right)\right]1\{s \in [-1,1]\} + 1\{s > 1\}.$$

The number of moment conditions (64) is the number of IVs times the number of quantile $(S_1 + K_1 + 1) \times (L - 1)$. As L is usually a large number, the moment condition (64) typically overidentifies α_t so that we use GMM. Firpo et al. (2022) derive a expression of the optimal GMM weight matrix and showed it does not depend on the parameter α_t so that its estimation completes in one step. Monotonicity in m_{it} and u_{it} is imposed via linear constraints on the derivatives of the basis functions. The demand shocks $\hat{u}_{it}$ are then estimated by numerically inverting $\hat{\phi}_t(m_{it}, \hat{u}_{it}) = r_{it}$. The same procedure is applied to $t - 1$ to estimate $\hat{u}_{it-1}$.

Step 2: Estimation of the control function The second step estimates the transformation model (63). We use the profile likelihood (PL) estimator developed by Linton, Sperlich, and Van Keilegom (2008). From (31) for $q_{it} = m_{it}$ and (61), the conditional density of m_{it} given v_{it} is written as

$$g_{m_t|v_t}(m_{it}|v_{it}) = g_{\eta_t}(\eta_{it})\,\frac{\partial \lambda_t(m_{it}, u_{it})}{\partial m_{it}}.$$

To approximate $\lambda_t(m_{it}, u_{it})$, we use the basis $B_\lambda(m_{it}, u_{it})$ that is the Kronecker product of B-spline bases of degree 3 with 1 interior knot in m_{it} and u_{it}. We do not assume a parametric distribution on η_{it}. Thus, the log-likelihood function is written as

$$\sum_{i=1}^{n}\left\{\ln g_{m_t|v_t}(m_{it}|v_{it})\right\} = \sum_{i=1}^{n}\left\{\ln g_{\eta_t}(\eta_{it}) + \ln \partial_m B_\lambda(m_{it}, u_{it})^T \beta_t\right\}.$$

where $g_{\eta_t}(\eta)$ is the corresponding (Gaussian) kernel density. We obtain estimates of η_{it} as follows. Using tilde to denote temporal estimate, for given β_t, we define $\lambda_{it}(\beta_t) := B_\lambda(m_{it}, u_{it})^T \beta_t$

in the transformation model (63) and project

$$\lambda_{it}(\beta_t) = \theta_k k_{it} + \theta_l l_{it} + \rho\lambda_{it-1}(\beta_{t-1}) - \rho\theta_k k_{it-1} - \rho\theta_l l_{it-1} + \eta_{it}. \tag{65}$$

by OLS to express the residual $\eta_{it}(\beta_t)$ that enters $g_{\eta_t}(\cdot)$. Then, the PL estimator $\tilde{\nu}_t^*$ is defined as

$$\tilde{\nu}_t^* \in \arg\max_{\nu_t^*} \sum_{i=1}^{n} \left\{ \ln g_{\eta_t^*}\left(\eta_{it}^*(\nu_t^*)\right) + \ln \partial_m B_\lambda(m_{it}, \hat{u}_{it})^T \nu_t^* \right\} \text{ subject to } \partial_m B_\lambda(m_{it}, \hat{u}_{it})^T \nu_t^* > 0.$$

Step 3: Estimation of production function, markup, TFP, and output With estimated $\tilde{\nu}_t^*$, we obtain $\tilde{\lambda}_t^*(m_{it}, \hat{u}_{it})$ to then estimate the following transformation model by OLS:

$$\tilde{\lambda}_t^*(m_{it}, \hat{u}_{it}) = \theta_k^* k_{it} + \theta_l^* l_{it} + \rho\, {\nu_{t-1}^*}^T B_\lambda(m_{it-1}, \hat{u}_{it-1}) - \rho\theta_k^* k_{it-1} - \rho\theta_l^* l_{it-1} + \eta_{it}^*.$$

We estimate $\tilde{\theta}_m^*$ as follows:

$$\tilde{\theta}_m^* = \text{median}\left(\frac{\partial_m B_\lambda(m_{it}, \hat{u}_{it})^T \tilde{\nu}_t^*}{\partial_m B_\phi(m_{it}, \hat{u}_{it})^T \hat{\alpha}_t - \frac{\exp(p_t^m + m_{it})}{\exp(r_{it})}} \frac{\exp(p_t^m + m_{it})}{\exp(r_{it})} \right).$$

Then, by De Loecker and Warzynski (2012), we estimate markups as follows:

$$\tilde{\mu}_{it}^* = \frac{\tilde{\theta}_m^*}{\exp(p_t^m + m_{it}) / \exp(r_{it})}.$$

By assuming constant returns to scale (CRS), the scale parameter $\hat{b}_t = \tilde{\theta}_m^* + \tilde{\theta}_k^* + \tilde{\theta}_l^*$ is estimated. Thus, the CRS normalized production parameters $\hat{\theta}_j = \tilde{\theta}_j^* / \hat{b}_t$ for $j \in \{m, k, l\}$ and markups $\hat{\mu}_{it} = \tilde{\mu}_{it}^* / \hat{b}_t$. With the mean-zero restriction naturally inherent from the AR(1) TFP process, the location parameter $\hat{a}_{2t} = n^{-1} \sum_{i=1}^{n} \left[\tilde{\lambda}_t^*(m_{it}, \hat{u}_{it}) - \tilde{\theta}_k^* k_{it} - \tilde{\theta}_l^* l_{it} \right]$ is estimated. The estimated TFP, output, and price are

$$\hat{\omega}_{it} = \frac{\tilde{\lambda}_t^*(m_{it}, \hat{u}_{it}) - \tilde{\theta}_k^* k_{it} - \tilde{\theta}_l^* l_{it} - \hat{a}_{2t}}{\hat{b}_t},$$
$$\hat{y}_{it} = \hat{\omega}_{it} + \hat{\theta}_m m_{it} + \hat{\theta}_l l_{it} + \hat{\theta}_k k_{it},$$
$$\hat{p}_{it} = r_{it} - \hat{y}_{it}.$$

Step 4: Estimation of parametric CoPaTh-HSA demand system Our estimation steps of production function above does not assume any parametric demand system. Thus, in theory, one can estimate a fully nonparametric HSA demand system as described in Section 3.5. However, in our empirical application, we estimate a parametric HSA demand system to obtain more stable estimates from a dataset with a moderate sample size. In particular, we consider a HSA demand system with the CoPaTh (constant pass-through) demand function with incomplete pass-through by Matsuyama and Ushchev (2020):

$$s_t^*(y_{it} - q_t(\mathbf{y}_t, \epsilon_t), \epsilon_{it}) := \delta_t - \frac{1}{\beta_t} \log\left(\frac{\exp(-\beta_t (y_{it} - q_t(\mathbf{y}_t, \epsilon_t)) + \kappa_t) + \epsilon_{it}}{1 + \epsilon_{it}}\right) \tag{66}$$

where the quantity index $q_t(\mathbf{y}_t, \epsilon_t)$ is implicitly defined by the market share constraint (51) for a given output vector $\mathbf{y}_t = (y_{1t}, ..., y_{Nt})$ and a given demand shock vector $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{Nt})$. Appendix A.1.1 derives the log-variable version of the CoPaTh demand (66) from Matsuyama and Ushchev (2020)'s original formulation. As explained in Section 3.5, we estimate the following reduced form revenue function instead of the structural form (66):

$$\varphi_t(y_{it}, \epsilon_{it}) = \Phi_t + \delta_t - \frac{1}{\beta_t} \log\left(\frac{\exp(-\beta_t y_{it} + \gamma_t) + \epsilon_{it}}{1 + \epsilon_{it}}\right)$$

where $\gamma_t \equiv \beta_t q_t(\mathbf{y}_t, \epsilon_t) + \kappa_t$.

The CoPaTh-HSA demand predicts the markup as:

$$\frac{\mathrm{P}_{it}}{\mathrm{MC}_{it}} = \left(\frac{\partial \varphi_t(y_{it}, \epsilon_{it})}{\partial y_{it}}\right)^{-1} = 1 + \epsilon_{it} \exp(\beta_t y_{it} - \gamma_t).$$

When $\beta_t = 0$, the markup is constant and the pass-through is complete. In addition, $\epsilon_{it} = \epsilon$ is common for all i, the demand system is reduced to the conventional CES demand system.

With the estimated outputs $\hat{y}_{it}$ and markups $\hat{\mu}_{it}$ from Step 3, the composite nonlinear least square estimator of demand parameters (β_t, γ_t, δ_t) is defined as,

$$(\hat{\beta}_t, \hat{\delta}_t, \hat{\gamma}_t)' \in \arg\min_{\beta_t, \delta_t, \gamma_t} \sum_i \left(r_{it} - \left(\Phi + \delta_t - \frac{1}{\beta_t} \log\left(\frac{\exp(-\beta_t \hat{y}_{it} + \gamma_t) + \epsilon_{it}}{1 + \epsilon_{it}}\right)\right)\right)^2$$

$$+ \sum_i (\hat{u}_{it} - (\mathrm{quantile}(\epsilon_{it})))^2,$$

s.t.

$$\epsilon_{it} = \frac{\hat{\mu}_{it} - 1}{\exp(\beta_t \hat{y}_{it} - \gamma_t)},$$

$$1 = \sum_i \exp\left(\delta_t - \frac{1}{\beta_t} \log\left(\frac{\exp(-\beta_t \hat{y}_{it} + \gamma_t) + \epsilon_{it}}{1 + \epsilon_{it}}\right)\right).$$

where quantile(ϵ_{it}) is the empirical quantile of ϵ_{it} among all firms and the industry revenue

$\Phi = \log\left(\sum_i \exp(r_{it})\right)$.<<We need to explain why we impose addition restrictions.>> [20]

5 Simulation

This section presents the finite sample performance of our proposed estimator, comparing to that of the ACF method when firms charge small but heterogeneous markups under the HSA demand system. We consider a simple data-generating process (DGP) in which firms set variable markups, calibrated to Chilean manufacturing data. Further details of the DGP and simulation design are provided in the Appendix.

Consider N firms in a market and $t \in \{1,2,...,T\}$ period. Each firm produces one variety of differentiated goods and faces the HSA-CoPaTh demand function (66). The demand shock ϵ_{it} follows an MA1 process: $\epsilon_{it} = 0.5\zeta_{it-1} + \zeta_{it}$, where $\zeta_{it} \sim Unif[0,0.3]$. The production function takes the Cobb-Douglass form:

$$y_{it} = \theta_m m_{it} + \theta_k k_{it} + \theta_l l_{it} + \omega_{it} \tag{67}$$

where ω_{it} follows an AR1 process $\omega_{it} = 0.8\omega_{it-1} + \eta_{it}, \eta_{it} \sim N\left(0,(0.05)^2\right)$. Capital k_{it} and labor l_{it} are predetermined and follow exogenous laws of motion explained in Appendix A.1.3.

For each period, we find equilibrium outputs and quantity index $(\mathbf{y}_t^m, q_t^m)$ in an MCE by solving the first order order conditions and the market share condition analogous to (54):

$$\begin{aligned}
\Phi_t + \delta_t - \beta_t\left(y_{it}^m - q_t^m\right) + \kappa_t + \Xi_{it} - \frac{y_{it}^m}{\theta_m} + \frac{1}{\beta_t}\log(1+\epsilon_{it})& \\
-\left(1+\frac{1}{\beta_t}\right)\ln\left(\exp(-\beta_t\left(y_{it}^m - q_t^m\right) + \kappa_t) + \epsilon_{it}\right) &= 0 \text{ for } i = 1,..,N_t \\
\sum_{i=1}^{N_t}\exp\left(\delta_t - \frac{1}{\beta_t}\ln\left(\frac{\exp(-\beta_t\left(y_{it}^m - q_t^m\right) + \kappa_t) + \epsilon_{it}}{1+\epsilon_{it}}\right)\right) &= 1.
\end{aligned}$$

where $\Xi_{it} = \ln\hat{\theta}_m + (\hat{\theta}_k k_{it} + \hat{\theta}_l l_{it} + \hat{\omega}_{it})/\hat{\theta}_m$ and $p_{mt} = 0$. Appendix A.2.2 show its derivation.

The value of the "reduced-form" parameter $(\alpha_t, \beta_t, \gamma_t)$ is determined by the HSA system's structural parameters $(\Phi_t, \delta_t, \beta_t, \kappa_t) = (20, -6.5, 0.21, 0)$ and the equilibrium quantity index $q_t(\mathbf{y}_t, \epsilon_t)$, such that $(\alpha_t, \beta_t) = (\Phi_t + \delta_t, \beta_t) = (13.5, 0.21)$, while the value of γ_t depends on $q_t(\mathbf{y}_t, \epsilon_t)$, β_t, and κ_t. Production function parameters are $(\theta_m, \theta_k, \theta_l) = (0.4, 0.3, 0.3)$. We simulate 100 replications of $N = 600$ firms and $T = 5$ periods, with the following summary statistics of the resulting markups:

[20]For model-consistency, we estimate the HSA demand system only using the firms with estimated markups $\hat{\mu}_{it} > 1$.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
Markup	1.001	1.146	1.220	1.223	1.295	1.614

Table 1: Summary statistics of Markups in simulated data ($t = 5$)

In addition to our proposed estimator, we also consider the estimator proposed by Ackerberg et al. (2015)(ACF). Gandhi et al. (2020)(GNR) showed the difficulty of identification in the DGP that ACF assumed where a firm-level unobserved shock is a scalar, TFP. The GNR criticism is not applicable for the current DGP with two unobserved shocks. However, to show our point is different from the GNR critique, we employ the ACF method with constant returns to scale (CRS) restriction (i.e., $\theta_m + \theta_k + \theta_k = 1$) that Flynn, Gandhi, and Traina (2019) proposed to address the GNR criticism.[21]

5.1 Result

Figure 1 show the histograms of 100 estimates of $(\theta_m, \theta_k, \theta_l)$ from the ACF method with revenue data and quantity data. While using quantity data yields estimates that are tightly clustered around the true values, using revenue data substantially biases the estimation of the production function. The simulation result confirms the long criticism in the literature against the ad hoc use of revenue data.

[21] In Appendix, we present the details of the DGP and the estimation method as well as basically the same results using the ACF method without imposing CRS.

Figure 1: Production Function Estimation with ACF on Revenue Data and Quantity Data

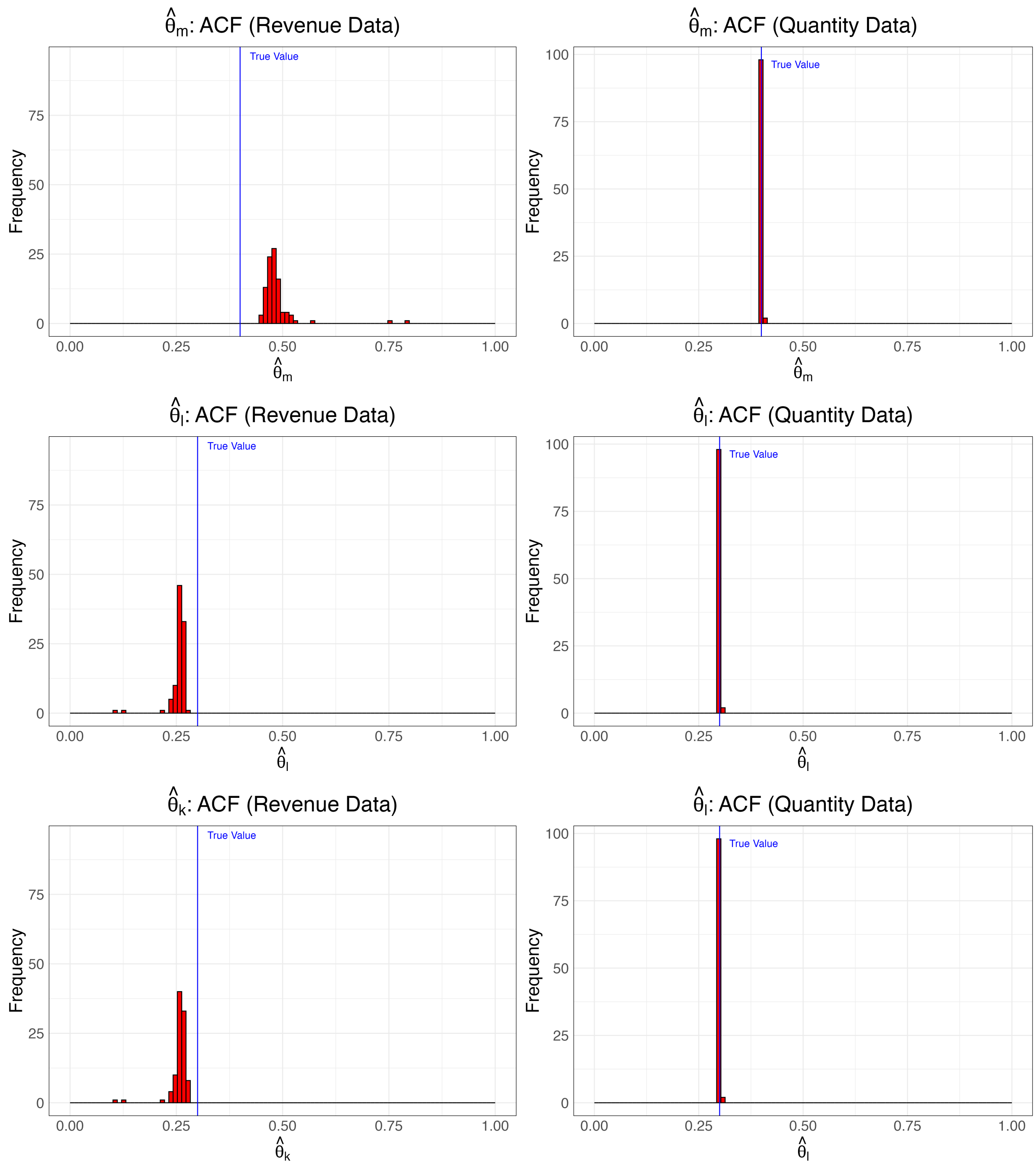


Figure 2 shows the histograms of 100 estimates for $(\theta_m, \theta_k, \theta_l)$ from our proposed estimator and $(\beta_t, \delta_t, \gamma_t)$ at $t = T$ on the HSA demand system. They are tightly clustered around their true values, suggesting that our method recovering the structural parameters very well. Figure 3 shows the scatter plot of true versus estimated TFPs for the first 20 Monte Carlo simulations, and Figure 4 shows the same for markups. The strong alignment of points along the 45-degree lines accompanying with the low RMSEs and high correlations suggest that our method precisely estimates TFPs and markups.

Figure 2: Production Function and Demand System Estimation with Revenue Data

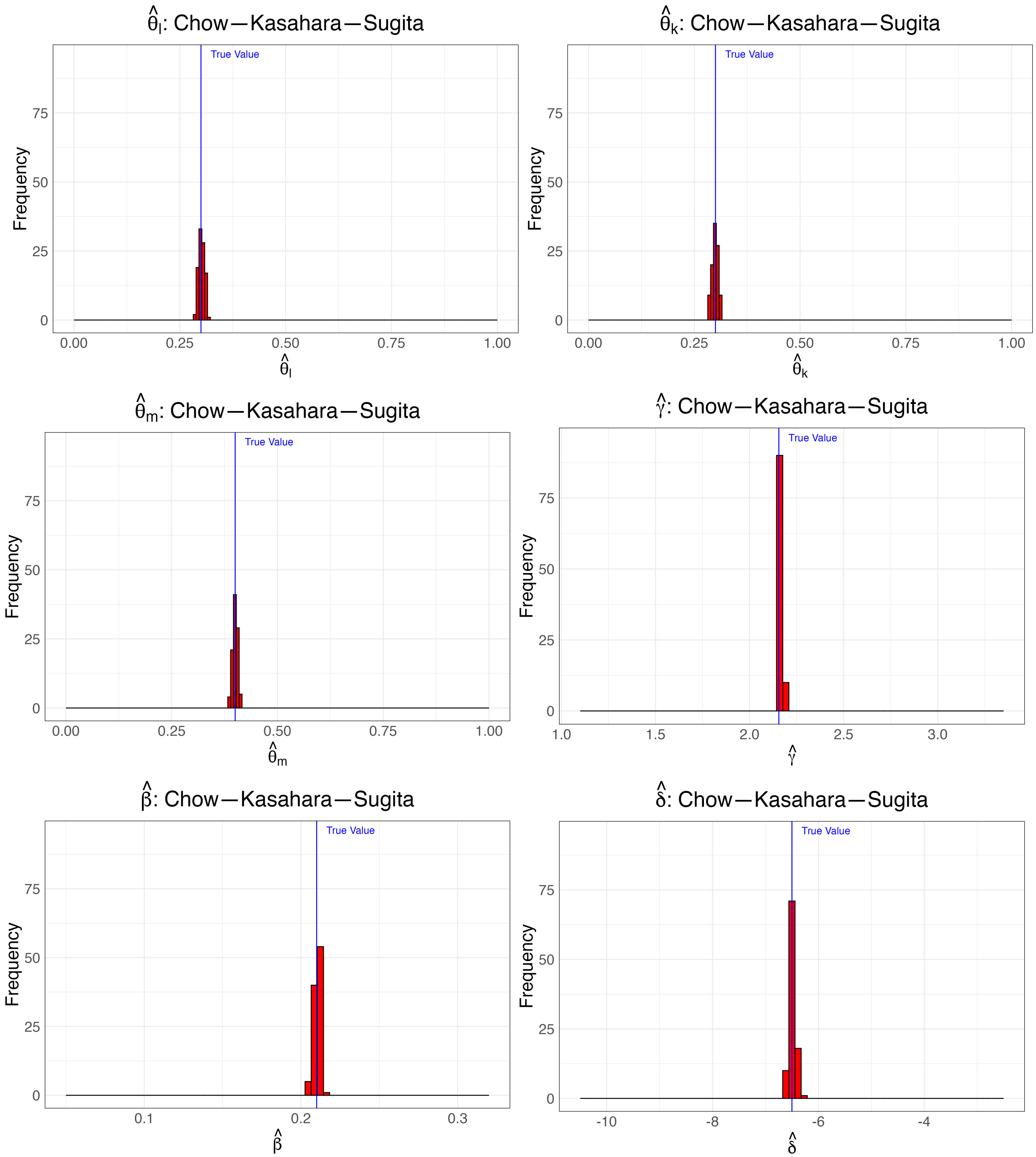


Figure 3: True and estimated TFPs for first 20 MC Simulations

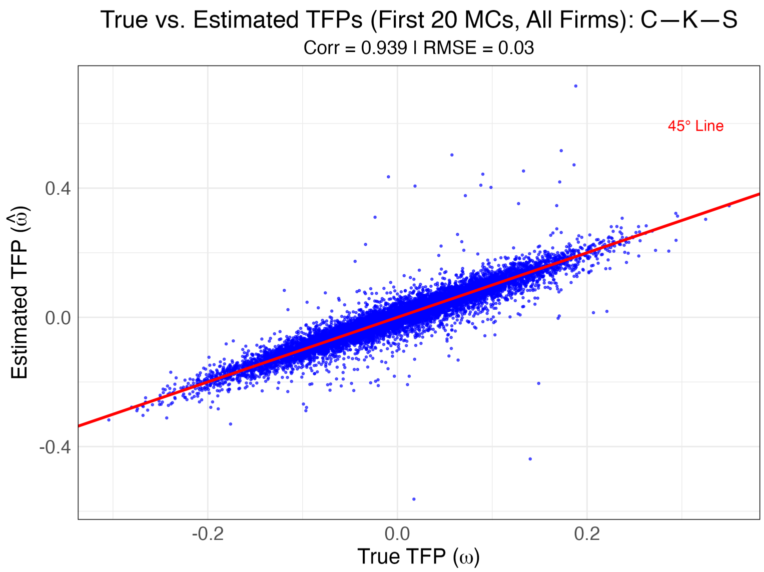
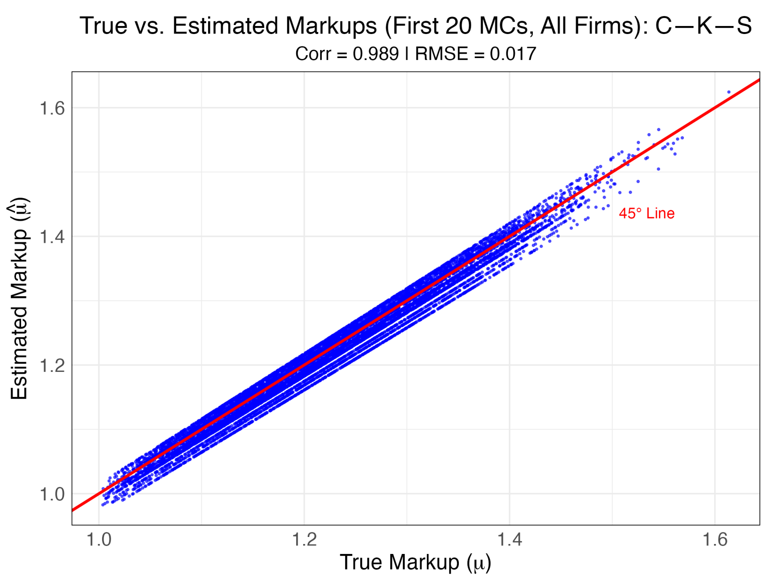


Figure 4: True and estimated Markups for first 20 MC Simulations



6 Empirical Application: Chilean Manufacturing Sector

The semiparametric estimator is applied to the Chilean manufacturing plant dataset, derived from the census conducted by Chile Instituto Nacional de EstadÃstica, covering all plants with 10 or more employees from 1993 to 1996. We define labor input as the number of workers, material input as materials cost, and revenue as income plus the value of capital produced for own use, with all values deflated using appropriate deflators. Capital input is calculated as the sum of deflated values for buildings, machinery, and vehicles, employing the perpetual inventory method with appropriate deflators. Our analysis focuses on the three largest industries with 2-digit SIC codes 31 (Food, Beverage, and Tobacco), 32 (Textiles, Apparel, and Leather Products), and 38 (Metal Products, Electric/Non-electric Machinery, Transport Equipment, and Professional Equipment) in 1996. We drop firms with non-positive capital. Also, firms with material cost-to-revenue ratios of less than 0, above 1, and in the bottom and top 2 percentiles of the distribution are excluded.

6.1 Result

Industry	n	$\hat{\theta}_m$	$\hat{\theta}_k$	$\hat{\theta}_l$	$\hat{\mu}$
31	736	0.848	0.013	0.138	1.386
		(0.031)	(0.010)	(0.031)	(0.052)
32	463	0.756	0.079	0.164	1.503
		(0.049)	(0.032)	(0.045)	(0.099)
38	391	0.672	0.058	0.270	1.628
		(0.067)	(0.037)	(0.062)	(0.167)

Table 2: Chilean Manufacturing plant estimation: Step 1, Step 2, and Step 3 (Industries 31, 32, and 38 in 1996). Standard errors in parentheses with 100 non-parametric bootstrap iterations.

Industry	n	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$
31	698	0.154	1.770	-8.009
		(0.011)	(0.105)	(0.159)
32	409	0.085	0.951	-6.899
		(0.012)	(0.141)	(0.572)
38	347	0.103	1.367	-5.443
		(0.044)	(0.443)	(1.352)

Table 3: Chilean Manufacturing plant estimation: Step 4 (Industries 31, 32, and 38 in 1996). Standard errors in parentheses with 100 non-parametric bootstrap iterations.

Table 2 and 3 demonstrate a successful application of our method to standard empirical data. Notably, we found empirical evidence that β is statistically significantly deviating from zero, suggesting that using CES demand system, which corresponds to the HSA demand system with $\beta = 0$ would be misspecified. Moreover, Figures 5 and 6 presents scatter plots of the observed revenue r_{it} against the fitted revenue by the HSA demand system from Step 4, and quantiles of estimated demand shocks ϵ_{it} from Step 4 against those u_{it} from Step 1, which show a good fit of our HSA demand system by the alignment of points along the 45-degree line.

Figure 5: Observered revenue vs fitted revenue from Step 4

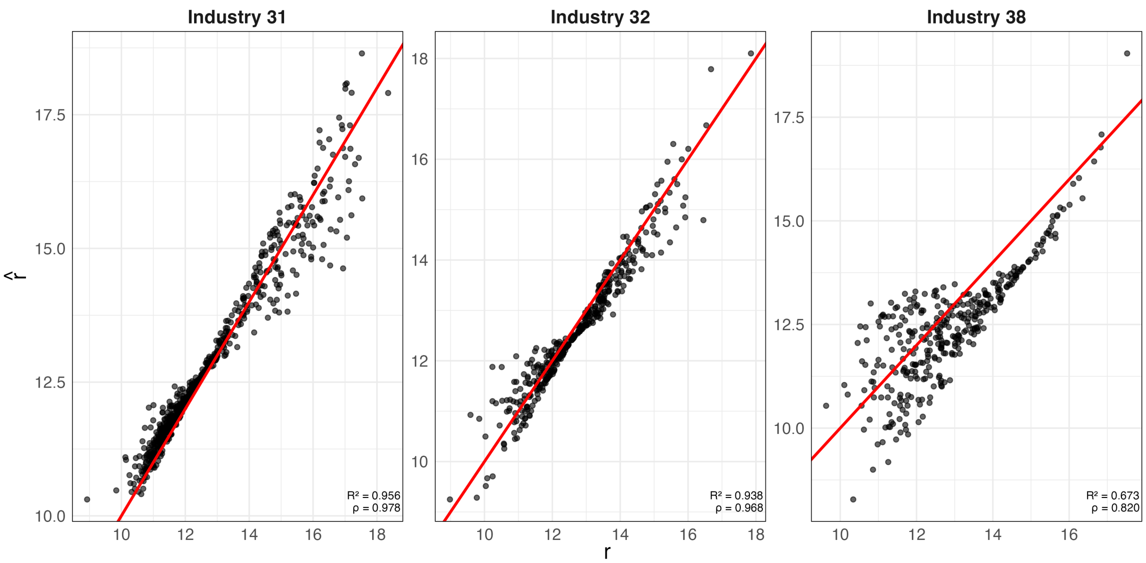
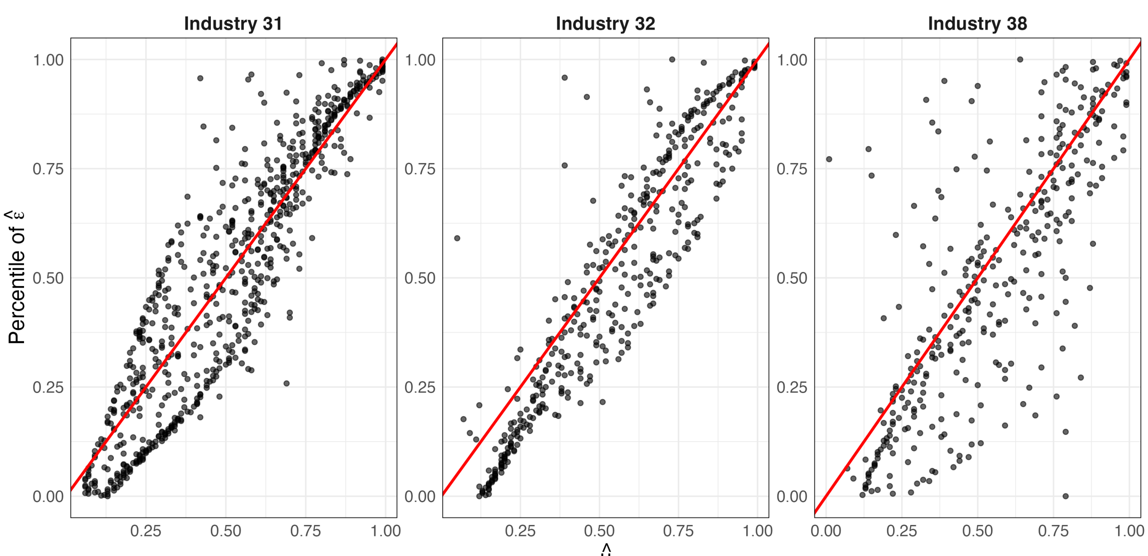


Figure 6: Rank of demand shock from Step 1 vs Step 4



6.2 Counterfactual Welfare Analysis

We quantify the consumer utility loss attributable to firm's market power, calculating the compensation variation for a counterfactual marginal cost pricing equilibrium (MPCE) as described in Section 3.5. First, we calculate a monopolistic competition equilibrium (MCE) as our benchmark, using the estimated structural parameters from Table 2 and 3 and firm-level states $(k_{it}, l_{it}, \hat{\omega}_{it}, \hat{\epsilon}_{it})$. Specifically, we recover y^m_{it} and recalibrate γ^m_t by jointly solving (54), which are simplified as:

$$\Phi_t + \hat{\delta}_t - \hat{\beta}_t y^m_{it} + \gamma^m_t + \Xi_{it} - \frac{y^m_{it}}{\hat{\theta}_m} + \frac{1}{\hat{\beta}_t}\log(1+\hat{\epsilon}_{it})$$
$$-\left(1+\frac{1}{\hat{\beta}_t}\right)\ln\left(\exp(-\hat{\beta}_t y^m_{it} + \gamma^m_t) + \hat{\epsilon}_{it}\right) = 0 \text{ for } i = 1,..,N_t$$
$$\sum_{i=1}^{N_t}\exp\left(\hat{\delta}_t - \frac{1}{\hat{\beta}_t}\ln\left(\frac{\exp(-\hat{\beta}_t y^m_{it} + \gamma^m_t) + \hat{\epsilon}_{it}}{1+\hat{\epsilon}_{it}}\right)\right) = 1$$

where $\Xi_{it} = \ln\hat{\theta}_m + (\hat{\theta}_k k_{it} + \hat{\theta}_l l_{it} + \hat{\omega}_{it})/\hat{\theta}_m$and p_{mt} is normalized to zero. The obtained output vector and parameters exactly satisfy our HSA demand system, which also ensures $\tilde{q}_t\left(y^m_t, \hat{\epsilon}_t\right) = 0$. Using this MCE as a benchmark removes model misspecification bias for counterfactual analysis.

Second, we consider a marginal cost pricing equilibrium (MCPE) for given counterfactual log income Φ^c_t. We find an output vector $\mathbf{y}^c_t(\Phi^c_t)$ and quantity index $\tilde{q}_t(\Phi^c_t)$ by solving (55), which are simplified as:

$$\Phi^c_t + \hat{\delta}_t + \Xi_{it} - \frac{y^c_{it}(\Phi^c_t)}{\hat{\theta}_m} - \frac{1}{\hat{\beta}_t}\ln\left(\frac{\exp(-\hat{\beta}_t\left(y^c_{it}(\Phi^c_t) - \tilde{q}^c_t(\Phi^c_t)\right) + \gamma^m_t) + \hat{\epsilon}_{it}}{1+\hat{\epsilon}_{it}}\right) = 0 \text{ for } i = 1,..,N_t$$
$$\sum_{i=1}^{N_t}\exp\left(\hat{\delta}_t - \frac{1}{\hat{\beta}_t}\ln\left(\frac{\exp(-\hat{\beta}_t\left(y^c_{it}(\Phi^c_t) - \tilde{q}^c_t(\Phi^c_t)\right) + \gamma^m_t) + \hat{\epsilon}_{it}}{1+\hat{\epsilon}_{it}}\right)\right) = 1.$$

We suppose the same income case $\Phi^c_t = \Phi_t$ as our benchmark.

Then, we calculate the compensation variation for transitioning to a MCPE. We find a counterfactual income Φ^{c*}_t that leads to a zero utility change (56):

$$\Delta\ln U^c(\Phi^{c*}_t) = \tilde{q}^c_t(\Phi^{c*}_t) + \sum_i\int_{y^m_{it}}^{y^c_{it}(\Phi^{c*}_t)-\tilde{q}^c_t(\Phi^{c*}_t)}\exp\left(\hat{\delta}_t - \frac{1}{\hat{\beta}_t}\log\left(\frac{\exp(-\hat{\beta}_t\zeta + \gamma^m_t) + \hat{\epsilon}_{it}}{1+\hat{\epsilon}_{it}}\right)\right)d\zeta = 0.$$

The compensation variation is calculated by $CV_t \equiv \exp(\Phi^{c*}_t) - \exp(\Phi_t)$.

Finally, we calculate firms' profit loss. In the case of $\Phi^c_t = \Phi_t$, the total profit change (57) is expressed as

$$\Pi^c - \Pi^m = \sum_{i=1}^{N_t} \exp\left(-\frac{\hat{\theta}_k k_{it} + \hat{\theta}_l l_{it} + \hat{\omega}_{it}}{\hat{\theta}_m}\right)\left\{\exp\left(\frac{y_{it}^m}{\hat{\theta}_m}\right) - \exp\left(\frac{y_{it}^c(\Phi_t)}{\hat{\theta}_m}\right)\right\} \tag{68}$$

since $\chi_{it}(y_{it}) = (y_{it} - \hat{\theta}_k k_{it} - \hat{\theta}_l l_{it} - \hat{\omega}_{it})/\hat{\theta}_m$.

Industry	CV	$\Delta\Pi$	Overall
31	-14.1	-11.0	3.08
	(2.71)	(-2.39)	(0.58)
32	-16.1	-9.89	6.20
	(5.97)	(5.04)	(1.35)
38	-9.85	-4.07	5.78
	(6.75)	(4.55)	(2.92)

Table 4: Compensating Variation, profit loss, and overall welfare change in percentage of industry revenue $\exp(\Phi_t)$ in the transition from original equilibrium to MCPE of Chilean Industries 31, 32, and 38 in 1996 under HSA demand system. Standard errors in parentheses with 100 non-parametric bootstrap iterations.

From Table 4, we found empirical evidence that under our HSA demand system market power in these industries results in consumer's welfare losses of approximately 10%–15% and profit gains of approximately 4%–11%, with overall welfare losses of 3%–6% of industry revenue in the three largest Chilean manufacturing industries in 1996.

7 Concluding Remarks

The current study develops constructive nonparametric identification of production function and markup from revenue data. Our method simultaneously addresses two fundamental identification issues raised in the literature of production function estimation since Marschak and Andrews (1944)—namely, correlations between inputs and TFP, and biases from markup heterogeneity when revenue is used as output. Under standard assumptions, when revenue is modeled as a function of output (rather than a mere proxy for output), firm's observed characteristics and an unobserved demand shock, various economic objects of interest can be identified from revenue data. We develop a semiparametric estimator that is implementable with standard datasets used in the literature. In simulation, our estimator performs very well. We successfully applied our estimator to Chilean manufacturing plant data and found empirical evidence of the misspecification of CES demand system. In counterfactual welfare analysis, the result shows that the market power results in welfare losses of approximately 3%–6% of industry revenue in the three largest Chilean manufacturing industries in 1996.

References

Ackerberg, D., Benkard, C. L., Berry, S., and Pakes, A. (2007), "Chapter 63 Econometric Tools for Analyzing Market Outcomes," Elsevier, vol. 6 of *Handbook of Econometrics*, pp. 4171 – 4276. 1

Ackerberg, D. A., Caves, K., and Frazer, G. (2015), "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83, 2411–2451. 1, 3.1, 3.3.4, 15, 3.6, 5, A.1.4

Arellano, M. and Bond, S. (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies*, 58, 277–297. 3.6

Arellano, M. and Bover, O. (1995), "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics*, 68, 29–51. 3.6

Bartelsman, E. J. and Doms, M. (2000), "Understanding productivity: Lessons from longitudinal microdata," *Journal of Economic literature*, 38, 569–594. 1

Berry, S., Levinsohn, J., and Pakes, A. (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890. 1, 7

Blundell, R. and Bond, S. (1998), "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 87, 115–143. 3.6

— (2000), "GMM estimation with persistent panel data: an application to production functions," *Econometric Reviews*, 19, 321–340. 3.6

Bond, S., Hashemi, A., Kaplan, G., and Zoch, P. (2020), "Some Unpleasant Markup Arithmetic: Production Function Elasticities and Their Estimation from Production Data," NBER Working Paper w27002. 1, 2, 8

Chernozhukov, V. and Hansen, C. (2005), "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261. 1, 5, 3.3.1

Chiappori, P.-A., Komunjer, I., and Kristensen, D. (2015), "Nonparametric Identification and Estimation of Transformation Models," *Journal of Econometrics*, 188, 22–39. 1, 3.1, 3.1, 3.3.2, 14, 3.3.2

De Loecker, J. (2011), "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity," *Econometrica*, 79, 1407–1451. 1, 5, 2, 2, 3.1, 16, 3.5.2

De Loecker, J., Eeckhout, J., and Unger, G. (2020), "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 135, 561–644. 4

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2016), "Prices, Markups, and Trade Reform," *Econometrica*, 84, 445–510. 2, 2

De Loecker, J. and Warzynski, F. (2012), "Markups and Firm-Level Export Status," *American Economic Review*, 102, 2437–71. 1, 2, 8, 3.1, 3.3.3, 4

Doraszelski, U. and Jaumandreu, J. (2018), "Measuring the Bias of Technological Change," *Journal of Political Economy*, 126, 1027–1084. 1

Ekeland, I., Heckman, J. J., and Nesheim, L. (2004), "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112, S60–S109. 1, 7

Feenstra, R. C. (2003), "A homothetic utility function for monopolistic competition models, without constant price elasticity," *Economics Letters*, 78, 79–86. 3.5.1

Feenstra, R. C. and Weinstein, D. E. (2017), "Globalization, markups, and US welfare," *Journal of Political Economy*, 125, 1040–1074. 3.5.1

Firpo, S., Galvao, A. F., Pinto, C., Poirier, A., and Sanroman, G. (2022), "GMM quantile regression," *Journal of Econometrics*, 230, 432–452. 1, 4, 4

Flynn, Z., Gandhi, A., and Traina, J. (2019), "Measuring Markups with Production Data," Unpublished. 6, 5

Foster, L., Haltiwanger, J., and Syverson, C. (2008), "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98, 394–425. 1, 2

Gandhi, A., Navarro, S., and Rivers, D. A. (2020), "On the identification of gross output production functions," *Journal of Political Economy*, 128, 2973–3016. 1, 5, 3.1, 3.3.3, 3.3.4, 5

Griliches, Z. and Mairesse, J. (1999), "Production Functions: The Search for Identification," in *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, ed. Strøm, S., Cambridge University Press, Econometric Society Monographs, pp. 169–203. 1

Hall, R. E. (1988), "The Relation between Price and Marginal Cost in US Industry," *Journal of Political Economy*, 96, 921–947. 1, 2

Heckman, J. J., Matzkin, R. L., and Nesheim, L. (2010), "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78, 1569–1591. 7

Horowitz, J. L. (1996), "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable," *Econometrica*, 103–137. 1

— (1998), "Bootstrap methods for median regression models," *Econometrica*, 1327–1351. 4

Imbens, G. W. and Newey, W. K. (2009), "Identification and estimation of triangular simultaneous equations models without additivity," *Econometrica*, 77, 1481–1512. A.4.1, B.1, B.1, B.1

Kasahara, H. and Rodrigue, J. (2008), "Does the Use of Imported Intermediates Increase Productivity? Plant-Level Evidence," *Journal of Development Economics*, 87, 106–118. 3.1

Katayama, H., Lu, S., and Tybout, J. R. (2009), "Firm-Level Productivity Studies: Illusions and a Solution," *International Journal of Industrial Organization*, 27, 403–413. 1, 2, 9

Klette, T. J. and Griliches, Z. (1996), "The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous," *Journal of Applied Econometrics*, 11, 343–361. 1, 5, 2, 2, 3.1, 16, 3.5.2

Levinsohn, J. and Petrin, A. (2003), "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 317–341. 1, 3.1

Linton, O., Sperlich, S., and Van Keilegom, I. (2008), "Estimation of a semiparametric transformation model," *The Annals of Statistics*, 36, 686–718. 1, 4

Lu, Y. and Yu, L. (2015), "Trade liberalization and markup dispersion: evidence from China's WTO accession," *American Economic Journal: Applied Economics*, 7, 221–53. 2, 2

Marschak, J. and Andrews, W. (1944), "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 12, 143–205. 1, 2, 3.1, 3.1, 7

Matsuyama, K. and Ushchev, P. (2020), "Constant Pass-Through," CEPR Discussion Paper 15475, C.E.P.R. Discussion Papers. 1, 19, 3.5.2, 4, 4, A.1.1, A.1.1, A.1.2, A.3.1, A.1

Matzkin, R. L. (2003), "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71, 1339–1375. 3.1

Nelson, C. R. and Plosser, C. R. (1982), "Trends and random walks in macroeconmic time series: some evidence and implications," *Journal of monetary economics*, 10, 139–162. 3.1

Nishioka, S. and Tanaka, M. (2019), "Measuring Markups from Revenue and Total Cost: An Application to Japanese Plant-Product Matched Data," Rieti Discussion Paper Series 19-E-018. 2, 2

Olley, G. S. and Pakes, A. (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1263–1297. 1

Rovigatti, G. (2017), *prodest: Production Function Estimation*, r package version 1.0.1. A.1.4

Syverson, C. (2011), "What Determines Productivity?" *Journal of Economic Literature*, 49, 326–65. 1

Van Biesebroeck, J. (2003), "Productivity dynamics with technology choice: An application to automobile assembly," *Review of Economic Studies*, 70, 167–198. 1

A Online Appendix (Not for Publication)

A.1 Simulation

A.1.1 CoPaTh-HSA Demand System

We consider the "Incomplete Constant (and Common) Pass-Through" formulation of the CoPaTh-HSA demand system in Matsuyama and Ushchev (2020). With their original notions, the budget share function for product ω is expressed as:

$$S^*_\omega\left(\frac{Y}{Q(\mathbf{Y})}\right) = \gamma_\omega \beta_\omega \left[\left(1-\frac{1}{\sigma_\omega}\right)\left(\frac{Y/Q(\mathbf{Y})}{\gamma_\omega}\right)^{-\Delta} + \frac{1}{\sigma_\omega}\right]^{-1/\Delta} \tag{A.1}$$

where Y is the level of output and $Q(\mathbf{Y})$ is the quantity index which is a function of the output vector $\mathbf{Y}$. The pass-through rate

$$\rho = \frac{\partial \ln P}{\partial \ln MC} = 1 + \frac{\partial \ln \mu}{\partial \ln MC} = \frac{1}{\Delta + 1}$$

is a function of parameter $\Delta \geq 0$ where $\mu = P/MC$ is a markup. When $\Delta = 0$ and $\sigma_\omega = \sigma$, the demand system is reduced to the conventional CES system.

We reformulate it in log variables:

$$\begin{aligned}
s^*_\omega(y) = \ln S^*_\omega(Y) &= \gamma_\omega \beta_\omega \left[\frac{1}{\sigma_\omega} + \left(1-\frac{1}{\sigma_\omega}\right)\left(\frac{\exp(y-q)}{\gamma_\omega}\right)^{-\Delta}\right]^{-1/\Delta} \\
&= \ln(\gamma_\omega\beta_\omega) - \frac{1}{\Delta}\ln\left[\left(1-\frac{1}{\sigma_\omega}\right)\left(\frac{\exp(y-q)}{\exp(\ln\gamma_\omega)}\right)^{-\Delta} + \frac{1}{\sigma_\omega}\right] \\
&= \ln(\gamma_\omega\beta_\omega) - \frac{1}{\Delta}\ln\left[\left(\frac{\sigma_\omega - 1}{\sigma_\omega}\right)\exp(-\Delta(y-q) + \Delta\ln\gamma_\omega) + \frac{1}{\sigma_\omega}\right] \\
&= \ln(\gamma_\omega\beta_\omega) - \frac{1}{\Delta}\ln\left[\frac{\exp(-\Delta(y-q) + \Delta\ln\gamma_\omega) + 1/(\sigma_\omega - 1)}{\sigma_\omega/(\sigma_\omega - 1)}\right]
\end{aligned}$$

Our formulation of the CoPaTh-HSA demand system of inverse demand functions is

$$s^*(y_{it}, \epsilon_{it}) = \delta_t - \frac{1}{\beta_t}\log\left(\frac{\exp(-\beta_t(y_{it} - q_t(\mathbf{y}_t, \epsilon_t)) + \kappa_t) + \epsilon_{it}}{1 + \epsilon_{it}}\right).$$

While the original formulation (A.1) has three firm specific demand shifters $(\gamma_\omega, \beta_\omega, \sigma_\omega)$, we only allow one shifter.

The correspondence between current parameter notations and Matsuyama and Ushchev (2020)'s is as follows:

Current notations	i	δ_t	β_t	ϵ_{it}	$1+\epsilon_{it}$	κ_t
Matsuyama and Ushchev (2020)'s notations	ω	$\ln\gamma_\omega\beta_\omega$	$\Delta = \frac{1-\rho}{\rho}$	$\frac{1}{\sigma_\omega-1}$	$\frac{\sigma_i}{\sigma_i-1}$	$\Delta\ln\gamma_\omega$
Range		$(-\infty,\infty)$	$(0,\infty)$	$(0,\infty)$	$(1,\infty)$	$(-\infty,\infty)$

A.1.2 Structural form of CoPaTh-HSA Demand System

We presented the reduced form of HSA demand system with CoPaTh demand function by Matsuyama and Ushchev (2020) in the main text. The relationship to the structural form of the HSA demand system is given by:

$$\begin{aligned} r_{it} &= \Phi_t + \delta_t - \frac{1}{\beta_t}\log\left(\frac{\exp(-\beta_t(y_{it} - q_t(\mathbf{y}_t,\epsilon_t)) + \kappa_t) + \epsilon_{it}}{1+\epsilon_{it}}\right) \\ &= \Phi + \delta_t - \frac{1}{\beta_t}\log\left(\frac{\exp(-\beta_t\hat{y}_{it} + \gamma_t) + \epsilon_{it}}{1+\epsilon_{it}}\right), \end{aligned}$$

where $\gamma_t = \beta_t q_t(\mathbf{y}_t,\epsilon_t) + \kappa_t$. In the simulation, we set $\kappa_t = 0$.

A.1.3 Data Generating Process

The demand shock ϵ_{it} follows an MA1 process:

$$\epsilon_{it} = \rho_\epsilon \xi_{it-1} + \xi_{it}$$

where $\xi_{it}$and ξ_{it-1} are independent uniform random variables with supports [0,0.3].

Capital and labor are predetermined and follows the following exogenous laws of motion:

$$\begin{aligned} k_{it} &= 0.99k_{it-1} + 0.11\omega_{it-1} + e_{kit},\ e_{kit} \sim N(0,0.25^2),\ k_{i0} \sim N(10,1) \\ l_{it} &= 0.99l_{it-1} + 0.11\omega_{it-1} + e_{lit}, e_{lit} \sim N(0,0.25^2),\ l_{i0} \sim N(10,1). \end{aligned}$$

Summary statistics The following table shows the summary statics of endogenous variables and exogenous variables.

Endogenous variables							
$t=5$	Mean	SD	Min	P25	Median	P75	Max
Markup	1.223	0.101	1.001	1.146	1.220	1.295	1.614
m_{it}	11.323	0.565	8.986	10.951	11.322	11.692	13.937
r_{it}	13.44	0.579	10.98	13.06	13.45	13.82	15.91
Exogenous variables							
$t=5$	Mean	SD	Min	P25	Median	P25	Max
ω_{it}	0.000	0.0832	−0.314	−0.056	−0.000	0.0563	0.358
ϵ_{it}	0.225	0.0971	0.000	0.150	0.225	0.300	0.448
k_{it}	9.511	1.097	4.569	8.771	9.514	10.252	13.953
l_{it}	9.520	1.094	4.875	8.778	9.521	10.257	14.131

A.1.4 Ackerberg et al. (2015) estimation method

We estimate the production function with ACF using the R package `prodest` by (Rovigatti, 2017). Scale parameters are normalized under constant returns to scale (CRS), and location parameters are normalized via the mean-zero restriction on the AR(1) TFP process for the estimates. The initial values in optimization are set to the estimated parameters from our method for empirical application and the true parameters for simulation.

A.2 Calculations and Proofs

A.2.1 A necessary and sufficient condition for Assumption 5

We first derive some derivatives for preparation. From $\varphi_t(y_{it}, z_{it}, u_{it}) = y_{it} + \psi_t(y_{it}, z_{it}, u_{it})$, the demand elasticity is expressed as

$$\frac{\partial \varphi_t}{\partial y_{it}} = 1 - \frac{1}{\sigma_t(y_{it}, z_{it}^d, u_{it})} \Leftrightarrow \sigma_t(y_{it}, z_{it}^d, u_{it}) = \frac{1}{1 - \partial \varphi_t(y_{it}, z_{it}, u_{it})/\partial y_{it}}$$

Their derivatives are

$$\frac{\partial \sigma_t}{\partial y_{it}} = \frac{\partial^2 \varphi_t / \partial y_{it}^2}{(1 - \partial \varphi_t / \partial y_{it})^2} \text{ and } \frac{\partial \sigma_t}{\partial u_{it}} = \frac{\partial^2 \varphi_t / \partial y_{it} \partial u_{it}}{(1 - \partial \varphi_t / \partial y_{it})^2}.$$

Denote the profit by

$$\pi_t(m_{it}, \omega_{it}, u_{it}) := \exp(\varphi_t(f_t(m, k_{it}, l_{it}, z_{it}^s) + \omega_{it}, z_{it} u_{it}), z_{it}^d, u_{it}) - \exp(p_t^m + m)$$

The first order condition for (11) is

$$\frac{\partial \pi_t}{\partial m} = \exp(\varphi_t(f_t(m_{it}, k_{it}, l_{it}, z^s_{it}) + \omega_{it}, z^d_{it}, u_{it})) \frac{\partial \varphi_t(f_t(m_{it}, k_{it}, l_{it}, z^s_{it}) + \omega_{it}, z^d_{it}, u_{it})}{\partial y_{it}} \frac{\partial f_t(m_{it}, k_{it}, l_{it})}{\partial m_{it}} - \exp(p^m_t + m_{it}) = 0.$$

Their cross derivatives are

$$\frac{\partial^2 \pi_t}{\partial m_{it} \partial \omega_{it}} = \exp(r_{it}) \frac{\partial f_t}{\partial m_{it}} \left[\left(\frac{\partial \varphi_t}{\partial y_{it}} \right)^2 + \frac{\partial^2 \varphi_t}{\partial y_{it}^2} \right]$$
$$\frac{\partial^2 \pi_t}{\partial m_{it} \partial u_{it}} = \exp(r_{it}) \frac{\partial f_t}{\partial m_{it}} \left(\frac{\partial \varphi_t}{\partial y_{it}} \frac{\partial \varphi_t}{\partial u_{it}} + \frac{\partial^2 \varphi_t}{\partial y_{it} u_{it}} \right).$$

From the implicit function theorem, the derivatives of the material demand function is

$$\frac{\partial \mathbb{M}_t}{\partial u_{it}} = -\frac{\partial^2 \pi_t / \partial m_{it} \partial u_{it}}{\partial^2 \pi_t / \partial m_{it}^2} \text{ and } \frac{\partial \mathbb{M}_t}{\partial \omega_{it}} = -\frac{\partial^2 \pi_t / \partial m_{it} \partial \omega_{it}}{\partial^2 \pi_t / \partial m_{it}^2}.$$

Differentiating $m_{it} = \mathbb{M}_t \left(\mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}), w_{it}, u_{it} \right)$ by u_{it}, we obtain the derivatives of the inverse function as

$$\begin{aligned}
\frac{\partial \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})}{\partial u_{it}} &= -\frac{\partial \mathbb{M}_t / \partial u_{it}}{\partial \mathbb{M}_t / \partial \omega_{it}} \\
&= -\frac{\partial^2 \pi_t / \partial m_{it} \partial u_{it}}{\partial^2 \pi_t / \partial m_{it} \partial \omega_{it}} \\
&= -\frac{\frac{\partial \varphi_t}{\partial y_{it}} \frac{\partial \varphi_t}{\partial u_{it}} + \frac{\partial^2 \varphi_t}{\partial y_{it} u_{it}}}{\left(\frac{\partial \varphi_t}{\partial y_{it}} \right)^2 + \frac{\partial^2 \varphi_t}{\partial y_{it}^2}}
\end{aligned}$$

Finally, we derive the derivative of ϕ_t with respect to u_{it}:

$$\begin{aligned}
\frac{\partial \phi_t(x_{it}, z_{it}, u_{it})}{\partial u_{it}} &= \frac{\partial \varphi_t}{\partial y_{it}} \frac{\partial \mathbb{M}_t^{-1}}{\partial u_{it}} + \frac{\partial \varphi_t}{\partial u_{it}} \\
&= \left(\left(\frac{\partial \varphi_t}{\partial y_{it}} \right)^2 + \frac{\partial^2 \varphi_t}{\partial y_{it}^2} \right)^{-1} \left(\frac{\partial \varphi_t}{\partial u_{it}} \frac{\partial^2 \varphi_t}{\partial y_{it}^2} - \frac{\partial \varphi_t}{\partial y_{it}} \frac{\partial^2 \varphi_t}{\partial y_{it} \partial u_{it}} \right). \\
&= \left(\left(\frac{\partial \varphi_t}{\partial y_{it}} \right)^2 + \frac{\partial^2 \varphi_t}{\partial y_{it}^2} \right)^{-1} \left(1 - \frac{\partial \varphi_t}{\partial y_{it}} \right)^2 \left(\frac{\partial \varphi_t}{\partial u_{it}} \frac{\partial \sigma_t}{\partial y_{it}} - \frac{\partial \varphi_t}{\partial y_{it}} \frac{\partial \sigma_t}{\partial u_{it}} \right)
\end{aligned}$$

The assumption$\partial \mathbb{M}_t / \partial \omega_{it} > 0$ and $\partial f_t / \partial m_t > 0$ implies $\left(\frac{\partial \varphi_t}{\partial y_{it}}\right)^2 + \frac{\partial^2 \varphi_t}{\partial y_{it}^2} > 0$. Therefore, we have

$$\frac{\partial \phi_t(x_{it}, z_{it}, \epsilon_{it})}{\partial u_{it}} > 0 \Leftrightarrow \frac{\partial \varphi_t}{\partial u_{it}} \frac{\partial \sigma_t}{\partial y_{it}} > \frac{\partial \varphi_t}{\partial y_{it}} \frac{\partial \sigma_t}{\partial u_{it}}.$$

A.2.2 Derivations of Equilibrium Conditions for MCE and MCPE

General HSA Demand System

MCE The profit maximization problem is

$$\max_m \exp\left(\Phi_t + s_t\left(f_t\left(m, k_{it}, l_{it}, z_{it}^s\right) + \omega_{it} - \tilde{q}_t^m, z_{it}^d, u_{it}\right)\right) - \exp\left(p_{mt} + m\right).$$

The first order condition is

$$\exp\left(\Phi_t + s_t\left(y_{it}^m - \tilde{q}_t^m, z_{it}^d, u_{it}\right)\right) \frac{\partial s_t\left(y_{it}^m - \tilde{q}_t^m, z_{it}^d, u_{it}\right)}{\partial y_{it}} \frac{\partial f_t\left(m_{it}^m, k_{it}, l_{it}\right)}{\partial m_{it}} = \exp\left(p_{mt} + m_{it}^m\right).$$

Substituting $m_{it}^m = \chi_{it}\left(y_{it}^m\right)$ and taking the log leads to

$$\begin{aligned}\Phi_t + s_t\left(y_{it}^m - \tilde{q}_t^m, z_{it}^d, u_{it}\right) + \ln \frac{\partial s_t\left(y_{it}^m - \tilde{q}_t^m, z_{it}^d, u_{it}\right)}{\partial y_{it}} \\ + \ln \frac{\partial f_t\left(\chi_{it}\left(y_{it}^m\right), k_{it}, l_{it}\right)}{\partial m_{it}} - p_{mt} - \chi_{it}\left(y_{it}^m\right) = 0.\end{aligned}$$

MCPE The profit maximization problem is

$$\max_m \exp\left(p_{it}^c + f_t\left(m, k_{it}, l_{it}, z_{it}^s\right) + \omega_{it}\right) - \exp\left(p_{mt} + m\right)$$

where the firm takes p_{it}^c as given

From $p_{it}^c + f_t\left(m, k_{it}, l_{it}, z_{it}^s\right) + \omega_{it} = \Phi_t^c + s_t\left(y_{it}^c - \tilde{q}_t^m, z_{it}^c, u_{it}\right)$, the first order condition is

$$\exp\left(\Phi_t^c + s_t\left(y_{it}^c - \tilde{q}_t^m, z_{it}^c, u_{it}\right)\right) \frac{\partial f_t\left(m_{it}^c, k_{it}, l_{it}\right)}{\partial m_{it}} = \exp\left(p_{mt} + m_{it}^c\right).$$

Substituting $m_{it}^m = \chi_{it}\left(y_{it}^m\right)$ and taking the log leads to

$$\begin{aligned}\Phi_t^c + s_t\left(y_{it}^c - \tilde{q}_t^m, z_{it}^c, u_{it}\right) \\ + \ln \frac{\partial f_t\left(\chi_{it}\left(y_{it}^c\right), k_{it}, l_{it}\right)}{\partial m_{it}} - p_{mt} - \chi_{it}\left(y_{it}^c\right) = 0.\end{aligned}$$

CoPaTh-HSA Demand System and Cobb-Douglas Production Function We use

$$s(y_{it}^m, \epsilon_{it}) = \delta_t - \frac{1}{\beta_t} \ln\left(\frac{\exp(-\beta_t y_{it}^m + \gamma_t^m) + \epsilon_{it}}{1 + \epsilon_{it}} \right)$$

where$\tilde{q}_t\left(y_t^m, \hat{\epsilon}_t\right) = 0$ holds in an initial equilibrium.

MCE Since

$$\frac{\partial s_t\left(y_{it}^m, \epsilon_{it}\right)}{\partial y_{it}^m} = \frac{\exp(-\beta_t y_{it}^m + \gamma_t^m)}{\exp(-\beta_t y_{it}^m + \gamma_t^m) + \epsilon_{it}}$$
$$\chi_{it}\left(y_{it}^m\right) = \frac{y_{it}^m - \theta_k k_{it} - \theta_l l_{it} - \omega_{it}}{\theta_m},$$

the first order condition becomes

$$\Phi_t + \delta_t - \frac{1}{\beta_t} \ln\left(\frac{\exp(-\beta_t y_{it}^m + \gamma_t^m) + \epsilon_{it}}{1 + \epsilon_{it}} \right) - \beta_t y_{it}^m + \gamma_t^m$$
$$- \ln\left(\exp(-\beta_t y_{it}^m + \gamma_t^m) + \epsilon_{it} \right) + \ln \theta_m - p_{mt} - \frac{y_{it}^m - \theta_k k_{it} - \theta_l l_{it} - \omega_{it}}{\theta_m} = 0.$$

Letting $\Xi_{it} \equiv \ln \theta_m + (\theta_k k_{it} + \theta_l l_{it} + \omega_{it})/\theta_m$ and $p_{mt} = 0$, it is simplified as

$$\Phi_t + \delta_t - \beta_t y_{it}^m + \gamma_t^m + \Xi_{it} - \frac{y_{it}^m}{\theta_m} + \frac{1}{\beta_t} \ln(1 + \epsilon_{it})$$
$$- \left(1 + \frac{1}{\beta_t}\right) \ln\left(\exp(-\beta_t y_{it}^m + \gamma_t^m) + \epsilon_{it} \right) = 0.$$

MCPE The first order condition becomes

$$\Phi_t + \delta_t - \frac{1}{\beta_t} \ln\left(\frac{\exp(-\beta_t y_{it}^m + \gamma_t^m) + \epsilon_{it}}{1 + \epsilon_{it}} \right)$$
$$+ \ln \theta_m - p_{mt} - \frac{y_{it}^m - \theta_k k_{it} - \theta_l l_{it} - \omega_{it}}{\theta_m} = 0.$$

Letting $\Xi_{it} \equiv \ln \theta_m + (\theta_k k_{it} + \theta_l l_{it} + \omega_{it})/\theta_m$ and $p_{mt} = 0$, it is simplified as

$$\Phi_t + \delta_t + \Xi_{it} - \frac{y_{it}^m}{\theta_m} - \frac{1}{\beta_t} \ln\left(\frac{\exp(-\beta_t y_{it}^m + \gamma_t^m) + \epsilon_{it}}{1 + \epsilon_{it}} \right) = 0.$$

A.3 Identification of Demand Function

A.3.1 Proof for Proposition 7

The proof for Proposition 7 uses the following result of Matsuyama and Ushchev (2017).

Theorem A.1. *(Matsuyama and Ushchev, 2017, Remark 3 and Proposition 1). Consider a mapping* $\mathbf{S}(\mathbf{Y}) := (S_1(Y_1), ..., S_N(Y_N))'$ *from* $\mathbb{R}^N_+$ *to* $\mathbb{R}^N_+$*, which is differentiable almost everywhere, is normalized by*

$$\sum_{i=1}^{N} S_i(Y_i^*) = 1, \tag{A.2}$$

for some point $\mathbf{Y}^* := (Y_1^*, ..., Y_N^*)$ *and satisfies the following conditions*

$$\begin{aligned} S_i'(Y_i)Y_i &< S_i(Y_i) \text{ for } i = 1, ..., N, \\ S_i'(Y_i)S_j'(Y_j) &\geq 0 \text{ for } i, j = 1, ..., N, \end{aligned} \tag{A.3}$$

for all $\mathbf{Y}$ *such that* $\sum_{i=1}^{N} S_i(Y_i) = 1$*. Then, (1) for any such mapping, there exists a unique monotone, convex, continuous, and homothetic rational preference that generates the HSA demand system described by*

$$P_i = \frac{I}{Y_i} S_i\left(\frac{Y_i}{Q(\mathbf{Y})}\right) \text{ for } i = 1, .., N,$$

where $I := \sum_{i=1}^{N} P_i Y_i$ *and* $Q(\mathbf{Y})$ *is obtained by solving*

$$\sum_{i=1}^{N} S_i\left(\frac{Y_i}{Q(\mathbf{Y})}\right) = 1.$$

(2) This homothetic preference is described by a utility function U *which is defined by*

$$\ln U(\mathbf{Y}) = \ln Q(\mathbf{Y}) + \sum_{i=1}^{N} \int_{c_i}^{Y_i/Q(Y)} \frac{S_i(\xi)}{\xi} d\xi, \tag{A.4}$$

where $\mathbf{c} = (c_1, ..., c_N)$ *is a vector of constants such that* $U(\mathbf{c}) = 1$.

Proof for Proposition 7

Proof. (a) We construct $s_t(y_{it} - \tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t), z_{it}^d, u_{it})$ and $\tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ as is explained in the main text. Fix $\mathbf{z}_t^d := (z_{1t}^d, ..., z_{Nt}^d), \mathbf{u}_t := (u_{1t}, ..., u_{Nt})$, and time t. For $\mathbf{Y} \in \bar{\mathscr{Y}}$, define $Q_t(\mathbf{Y}) := \exp\left(\tilde{q}_t(\ln \mathbf{Y}, \mathbf{z}_t^d, \mathbf{u}_t)\right)$ and $S_t(\mathbf{Y}) := \left(S_{1t}(Y_1), ..., S_{N_t t}(Y_{N_t})\right)$ such that $S_{it}(Y_i) \equiv \exp\left(s_t\left(\ln Y_i, z_{it}^d, u_{it}\right)\right) = \exp\left(\varphi_t\left(\ln Y_{it}, z_{it}^d, u_{it}\right) - \Phi_t\right)$.

From Assumption 2 (b) and $y := \ln Y$,

$$0 < \frac{\partial \varphi_t\left(\ln Y, z^d, u\right)}{\partial \ln Y} = 1 + \frac{\partial \psi_t\left(\ln Y, z^d, u\right)}{\partial \ln Y} < 1$$

holds for any (y, z^d, u). The above inequality implies

$$S'_{it}(Y) > 0 \text{ and } S'_{it}(Y)Y < S_{it}(Y) \text{ for all } i \text{ and } Y$$

because

$$\begin{aligned} S'_{it}(Y)Y &= \exp\left(\varphi_t\left(\ln Y, z^d_{it}, u_{it}\right) - \Phi_t\right) \frac{\partial \varphi_t\left(\ln Y, z^d_{it}, u_{it}\right)}{\partial \ln Y} \\ &= S_{it}(Y) \frac{\partial \varphi_t\left(\ln Y, z^d_{it}, u_{it}\right)}{\partial \ln Y}. \end{aligned}$$

Therefore, $\mathbf{S}(\mathbf{Y})$ satisfies the inequalities in (A.3) for all $\mathbf{Y}$ satisfying $\sum_{i=1}^{N_t} S_{it}(Y_i) = 1$.

From Theorem A.1 (1), there exists a unique monotone, convex, continuous, and homothetic rational preference that generates

$$P_{it} = \frac{I_t}{Y_{it}} S_{it}\left(\frac{Y_{it}}{Q_t\left(\mathbf{Y}_t\right)}\right)$$

where $I_t = \exp(\Phi_t)$ is the consumer's budget. Taking the log of the above demand function, we obtain

$$p_{it} = \Phi_t + s_t(y_{it} - \tilde{q}_t(\mathbf{y}_t, \mathbf{z}^d_t, \mathbf{u}_t), z^d_{it}, u_{it}) - y_{it}.$$

(b) Fix $\mathbf{z}^d_t := (z^d_{1t}, ..., z^d_{Nt}), \mathbf{u}_t := (u_{1t}, ..., u_{Nt})$, and time t. For $\mathbf{Y}_t \in \bar{\mathscr{Y}}$, let $U_t := U_t(\mathbf{Y}_t, \mathbf{z}^d_t, \mathbf{u}_t)$ be the utility function of the representative consumer. From Theorem A.1,

$$\begin{aligned} \ln U_t &= \ln Q_t(\mathbf{Y}) + \sum_{i=1}^{N} \int_{c_i}^{Y_i/Q(Y)} \frac{S_i(\xi)}{\xi} d\xi. \\ &= \tilde{q}_t(\ln \mathbf{Y}, \mathbf{z}^d_t, \mathbf{u}_t) + \sum_{i=1}^{N} \int_{c_i}^{Y_i/\exp\left(\tilde{q}_t(\ln \mathbf{Y}, \mathbf{z}^d_t, \mathbf{u}_t)\right)} \frac{\exp\left(s_t\left(\ln \xi, z^d_{it}, u_{it}\right)\right)}{\xi} d\xi \end{aligned}$$

Applying a change in variable $\zeta = \ln \xi + a_t(\tilde{\mathbf{y}}_t)$ and $d\zeta = \frac{d\xi}{\xi}$, we write

$$\ln U_t(\mathbf{y}_t, \mathbf{z}^d_t, \mathbf{u}_t) = \tilde{q}_t(\mathbf{y}_t, \mathbf{z}^d_t, \mathbf{u}_t) + \sum_{i=1}^{N} \int_{\ln c_i}^{y_{it} - \tilde{q}_t(\mathbf{y}_t, \mathbf{z}^d_t, \mathbf{u}_t)} \exp\left(s_t\left(\zeta, z^d_{it}, u_{it}\right)\right) d\zeta.$$

(c) The homothetic preference implies that the market share $P_{it}Y_{it}/I_t$ depends only on a price vector and is independent of income. From

$$\begin{aligned}
1 &= \sum_{i=1}^{N_t} \exp\left(s_t(y_{it} - \tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t), z_{it}^d, u_{it}\right) \\
&= \sum_{i=1}^{N_t} \exp\left(s_t(y_{it} + \gamma - \tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t) - \gamma, z_{it}^d, u_{it}\right) \\
&= \sum_{i=1}^{N_t} \exp\left(s_t(y_{it} + \gamma - \tilde{q}_t(\mathbf{y}_t + \gamma, \mathbf{z}_t^d, \mathbf{u}_t), z_{it}^d, u_{it}\right),
\end{aligned}$$

we have $\tilde{q}_t(\mathbf{y}_t + \gamma, \mathbf{z}_t^d, \mathbf{u}_t) = \tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t) + \gamma$. Since the output $y_{it} = \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})$ is identified up to location, there is $a \in \mathbb{R}$ such that $y_{it} = a + y_{it}^*$ where y_{it}^* is the true output. Note that

$$\begin{aligned}
y_{it} - \tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t) &= a + y_{it}^* - \tilde{q}_t(a + \mathbf{y}_t^*, \mathbf{z}_t^d, \mathbf{u}_t) \\
&= a + y_{it}^* - \tilde{q}_t(\mathbf{y}_t^*, \mathbf{z}_t^d, \mathbf{u}_t) - a \\
&= y_{it}^* - \tilde{q}_t(\mathbf{y}_t^*, \mathbf{z}_t^d, \mathbf{u}_t).
\end{aligned}$$

The utility is expressed as:

$$\begin{aligned}
\ln U_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t) &= \tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t) + \sum_{i=1}^{N} \int_{\ln c_i}^{y_{it} - \tilde{q}_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)} \exp\left(s_t\left(\zeta, z_{it}^d, u_{it}\right)\right) d\zeta. \\
&= \tilde{q}_t(\mathbf{y}_t^* + a, \mathbf{z}_t^d, \mathbf{u}_t) + \sum_{i=1}^{N} \int_{\ln c_i}^{y_{it}^* - \tilde{q}_t(\mathbf{y}_t^*, \mathbf{z}_t^d, \mathbf{u}_t)} \exp\left(s_t\left(\zeta, z_{it}^d, u_{it}\right)\right) d\zeta \\
&= a + \tilde{q}_t(\mathbf{y}_t^*, \mathbf{z}_t^d, \mathbf{u}_t) + \sum_{i=1}^{N} \int_{\ln c_i}^{y_{it}^* - \tilde{q}_t(\mathbf{y}_t^*, \mathbf{z}_t^d, \mathbf{u}_t)} \exp\left(s_t\left(\zeta, z_{it}^d, u_{it}\right)\right) d\zeta \\
&= a + \ln U(\mathbf{y}_t^*, \mathbf{z}_t^d, \mathbf{u}_t).
\end{aligned}$$

Therefore, the log utility function is identified up to the location normalization of $\varphi_t^{-1}(\cdot)$. The identified utility function is a monotonic transformation of the true utility function, which implies both utility functions represent the same consumer preference. □

A.4 Endogenous Characteristics

A.4.1 Proof for Lemma B.1

The following proof follows Imbens and Newey (2009).

Proof. From the monotonicity of Γ_{kt} in Assumption B.1 (iii), we can define the inverse function of Γ_{kt}such that $\kappa_{it}^{k} = \Gamma_{kt}^{-1}\left(\varsigma_{it}, \varpi_{it}, z_{it}^{k}\right)$.For given $(z_{t}^{k}, \varsigma_{t}, \varpi_{t})$

$$
\begin{aligned}
F_{z_t^k|\varsigma_t,\varpi_t}(z_t^k|\varsigma_t,\varpi_t) &= \Pr\left(\Gamma_{kt}(\varsigma_{it},\varpi_{it},\kappa_{it}^k) \le z_t^k|\varsigma_{it}=\varsigma_t,\varpi_{it}=\varpi_t\right) \\
&= \Pr\left(\kappa_{it}^k \le \Gamma_{kt}^{-1}\left(\varsigma_t,\varpi_t,z_t^k\right)|\varsigma_{it}=\varsigma_t,\varpi_{it}=\varpi_t\right) \\
&= F_{\kappa_t^k}\left(\Gamma_{kt}^{-1}\left(\varsigma_t,\varpi_t,z_t^k\right)\right). \text{ (from } \kappa_{it}^k \perp (\varsigma_{it},\varpi_{it}))
\end{aligned}
$$

Therefore, we have

$$
\xi_{it}^k = F_{\kappa_t^k}\left(\Gamma_{kt}^{-1}\left(\varsigma_{it},\varpi_{it},z_{it}^k\right)\right) = F_{\kappa_t^k}\left(\kappa_{it}^k\right).
$$

Consider an arbitrary point (ξ_t, η_t) on the support of (ξ_{it}, η_{it}). Let $(\kappa_t^s, \kappa_t^d) = \left(F_{\kappa_t^s}^{-1}(\xi_t^s), F_{\kappa_t^d}^{-1}(\xi_t^d)\right)$. Since $F_{\kappa_t^k}$ is strictly monotonic, the conditional expectations given $\xi_{it} = \xi_t$ are identical to those given $(\kappa_{it}^s, \kappa_t^d) = (\kappa_{it}^s, \kappa_t^d)$. For any bounded function $a(v_{it})$ of v_{it}, the independence of $(\varsigma_{it}, \varpi_{it})$ and $\left(\kappa_{it}^s, \kappa_{it}^d, \eta_{it}\right)$ implies

$$
\begin{aligned}
&E\left[a(v_{it})|\xi_{it}=\xi_t,\eta_{it}=\eta_t\right] \\
&= E\left[a(v_{it})|\kappa_{it}^s=\kappa_t^s,\kappa_{it}^d=\kappa_t^s,\eta_{it}=\eta_t\right] \\
&= \int a\left(\Gamma_{st}(\varsigma_{it}\varpi_{it},\kappa_t^s),\Gamma_{dt}(\varsigma_{it},\varpi_{it},\kappa_t^d),\varpi_{it}\right)F_{\varsigma_t,\varpi_t}(d(\varsigma_{it},\varpi_{it})) \\
&= E\left[a(v_{it})|\kappa_{it}^s=\kappa_t^s,\kappa_{it}^d=\kappa_t^s\right] \\
&= E\left[a(v_{it})|\xi_{it}=\xi_t\right].
\end{aligned}
$$

For any bounded functions $a(v_{it})$ and $b(\eta_{it})$, we have

$$
\begin{aligned}
E\left[a(v_{it})b(\eta_{it})|\xi_{it}=\xi_t\right] &= E\left[E\left[a(v_{it})b(\eta_{it})|\xi_{it}=\xi_t,\eta_{it}=\eta_t\right]|\xi_{it}=\xi_t\right] \\
&= E\left[b(\eta_{it})E\left[a(v_{it})|\xi_{it}=\xi_t,\eta_{it}=\eta_t\right]|\xi_{it}=\xi_t\right] \\
&= E\left[b(\eta_{it})E\left[a(v_{it})|\xi_{it}=\xi_t\right]|\xi_{it}=\xi_t\right] \\
&= E\left[b(\eta_{it})|\xi_{it}=\xi_t\right]E\left[a(v_{it})|\xi_{it}=\xi_t\right]
\end{aligned}
$$

Thus, $\eta_{it} \perp v_{it}|\xi_{it}$. □

B Alternative Settings

B.1 Endogenous Firm Characteristics

Firm characteristics (z^s_{it}, z^d_{it}) may correlate with ζ_{it} and η_{it}. In step 1, we can use $(z^s_{it-\upsilon-1}, z^d_{it-\upsilon-1})$ in place of $(z^s_{it-\upsilon}, z^d_{it-\upsilon})$ as instrument variables to construct the moment condition similar to (29). In step 2, we consider the nonparametric control function approach by Imbens and Newey (2009) using instrument variables in the triangular model setting. We assume there exist instrument variables $\varsigma_{it} = (\varsigma^s_{it}, \varsigma^d_{it})$, unknown functions Γ_{kt}, and unobservable scalars κ^k_{it} such that

$$z^k_{it} = \Gamma_{kt}(\varsigma_{it}, \varpi_{it}, \kappa^k_{it}), (k = s, d)$$

where $\varpi_{it} := (l_{it}, k_{it}, u_{it}, w_{it-1}, u_{it-1}, z^h_{it-1})$.

Assumption B.1. *For $k = s, d$, (i) $\varsigma^k_{it} \perp (\eta_{it}, \kappa^k_{it})$ (ii) κ^k_{it} is a scalar and $\kappa^k_{it} \perp (\varsigma_{it}, \varpi_{it})$.(iii) Γ_{kt} is strictly increasing in κ^k_t (iv) The CDF of κ^k_{it}, $F_{\kappa^k_t}\left(\kappa^k_{it}\right)$, is strictly increasing on the support of κ^k_{it}.*

Let $F_{z^k_t|\varsigma_t,\varpi_t}(z^k_t|\varsigma_t, \varpi_t)$ be the CDF of z^k_{it} conditional on $(\varsigma_{it}, \varpi_{it}) = (\varsigma_t, \varpi_t)$. Define $\xi^k_{it} := F_{z^k_t|\varsigma_t,\varpi_t}(z^k_{it}|\varsigma_{it}, \varpi_{it})$ and $\xi_{it} := (\xi^s_{it}, \xi^d_{it})$. Imbens and Newey (2009) showed ξ_{it}can be used as control variables, that is, η_{it} becomes independent of v_{it} conditional on ξ_{it}. Appendix provides a proof that follows Imbens and Newey (2009).

Lemma B.1. *(Imbens and Newey, 2009,Theorem 1) $\eta_{it} \perp v_{it}|\xi_{it}$.*

From Lemma B.1, the conditional distribution of m_{it} given (v_{it}, ξ_{it}) satisfies

$$\begin{aligned} G_{m_t|v_t,\xi_t}(m_{it}|v_{it}, \xi_{it}) &= G_{\eta_t|v_t,\xi_t}\left(\mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it}) - \bar{h}_t\left(m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}\right)|v_{it}, \xi_{it}\right) \\ &= G_{\eta_t|\xi_t}\left(\mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it}) - \bar{h}_t\left(m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}\right)|\xi_{it}\right), \end{aligned}$$

Taking the derivatives of both sides with respect to $q_{it} \in \{m_{it}, k_{it}, l_{it}, z^s_{it}, z^d_{it}, u_{it}\}$ and $q_{it-1} \in \{m_{it-1}, k_{it-1}, l_{it-1}, z^s_{it-1}, z^d_{it-1}, z^h_{it-1}, u_{it-1}\}$, we obtain

$$\frac{\partial G_{m_t|v_t}(m_{it}|v_{it}, \xi_{it})}{\partial q_{it}} = \frac{\partial \mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it})}{\partial q_{it}} g_\eta(\eta_{it}|\xi_{it}), \tag{B.5}$$

$$\frac{\partial G_{m_t|v_t}(m_{it}|v_{it}, \xi_{it})}{\partial q_{it-1}} = -\frac{\partial \bar{h}\left(m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}\right)}{\partial q_{it-1}} g_\eta(\eta_{it}|\xi_{it}), \tag{B.6}$$

where $\eta_{it} = \mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it}) - \bar{h}_t\left(m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}\right)$.

By taking the ratio of (B.5) and (B.6), we obtain (33) again. Therefore, following the same steps in the proof for Proposition 3, we can identify $\mathbb{M}_t^{-1}(\cdot)$ up to scale and location and $G_\eta(\cdot)$ up to the scale normalization of η_t. Once $\phi_t(m_{it}, w_{it}, u_{it})$ and $\mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})$ are identified, the step 3 can identify the same objects as before.

B.2 Discrete Firm Characteristics

B.2.1 Exogenous Characteristics

Supposez_{it}^s, $z_{it}^d$and z_{it}^h are discrete variable and have finite support$\mathscr{Z}_s := \{z_s^1, ..., z_s^{J_s}\}$,$\mathscr{Z}_d := \{z_d^1, ..., z_d^{J_d}\}$ and $\mathscr{Z}_h := \{z_h^1, ..., z_h^{J_h}\}$ In step1, the identification of the IVQR model does not require the continuity of firm characteristics.Therefore, this section proves Propositions 2 and 3.The following assumption modifies Assumption 2 for discrete z_{it}^s and z_{it}^d.

Assumption B.2. *(a) For every $z^s \in \mathscr{Z}_s$,$f_t(\cdot, z^s)$ is continuously differentiable with respect to (m, k, l) on $\mathscr{M} \times \mathscr{K} \times \mathscr{L}$ and strictly increasing in m. (b) For every $\left(z^d, u\right) \in \mathscr{Z}_d \times [0, 1]$, $\varphi_t(\cdot, z^d, u)$ is strictly increasing and invertible with its inverse $\varphi_t^{-1}(r, z^d, u_t)$, which is continuously differentiable with respect to (r, u) on $\mathscr{R} \times [0, 1]$. (c) For every $(k, l, z^s, z^d, u) \in \mathscr{K} \times \mathscr{L} \times \mathscr{Z}_s \times \mathscr{Z}_d \times [0, 1]$, $\mathbb{M}_t(\cdot, k, l, z^s, z^d, u)$ is strictly increasing and invertible with its inverse $\mathbb{M}_t^{-1}(m, k, l, z^s, z^d, u)$, which is continuously differentiable with respect to (m, k, l, z^s, z^d, u) on $\mathscr{M} \times \mathscr{K} \times \mathscr{L} \times \mathscr{Z}_s \times \mathscr{Z}_d \times [0, 1]$. (d) $(\zeta_{it}, ..., \zeta_{it-v})$ are independent from η_{it}.*

The following assumption modifies Assumption 6 for discrete z_{it}^s and z_{it}^d.

Assumption B.3. *(a) The distribution $G_\eta(\cdot)$ of η is absolutely continuous with a density function $g_\eta(\cdot)$ that is continuous on its support. (b) η_{it} is independent of $v_{it} := (w_{it}, u_{it}, m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h)' \in \mathscr{V}$ with $E[\eta_{it}|v_{it}] = 0$. (c) $(x_{it}, x_{it-1}, u_{it}, u_{it-1})$ is continuously distributed on $\mathscr{X}^2 \times [0, 1]^2$. (d) The support Ω of ω is an interval $[\underline{\omega}, \bar{\omega}] \subset \mathbb{R}$ where $\underline{\omega} < 0$ and $1 < \bar{\omega}$. (e) For every $z^h \in \mathscr{Z}_h$, $h(\cdot, z^h)$ is continuously differentiable with respect to ω on Ω. (f) The set $\mathscr{A}_{q_{t-1}} := \{(m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h) \in \mathscr{M} \times \mathscr{W} \times [0, 1] \times \mathscr{Z}_h : \partial G_{m_t|v_t}(m_{it}|v_{it})/\partial q_{it-1} \neq 0$ for all $(m_{it}, w_{it}, u_t) \in \mathscr{M} \times \mathscr{W} \times [0, 1]\}$ is nonempty for some $q_{it-1} \in \{m_{it-1}, k_{it-1}, l_{it-1}, u_{it-1}\}$. (g) For each $(m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h) \in \mathscr{M} \times \mathscr{W} \times [0, 1] \times \mathscr{Z}_h$, it is possible to find $(\tilde{m}_t, \tilde{w}_t, \tilde{u}_t) \in \mathscr{M} \times \mathscr{W} \times [0, 1]$ such that $\partial G_{m_t|v_t}(\tilde{m}_t|\tilde{w}_t, \tilde{u}_t, m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h)/\partial m_t > 0$.*

The following proposition establishes the identification of $\mathbb{M}_t^{-1}(\cdot)$.

Proposition B.1. *Suppose that Assumptions 3, 4, B.2, and B.3 hold. Then, we can identify $\mathbb{M}_t^{-1}(\cdot)$ up to scale and location, and identify $G_\eta(\cdot)$ up to scale.*

Proof. Choose normalization points $(m^*_{t1}, k^*_t, l^*_t, u^*_t)$ and $(m^*_{t0}, k^*_t, l^*_t, u^*_t)$ in Assumption 4 as well as $\left(m^*_{t-1}, k^*_{t-1}, l^*_{t-1}, u^*_{t-1}\right) \in \mathscr{X} \times [0,1]$ such that, for $\left(z^s_t, z^s_{t-1}, z^d_t, z^d_{t-1}, z^h_{t-1}\right) \in \mathscr{Z}^2_s \times \mathscr{Z}^2_d \times \mathscr{Z}_h$,

$$\mathbb{M}^{-1}_t(m^*_{t0}, k^*_t, l^*_t, z^s_t, z^d_t, u^*_t) = c_0(z^s_t, z^d_t),\ \mathbb{M}^{-1}_t(m^*_{t1}, k^*_t, l^*_t, z^s_t, z^d_t, u^*_t) = c_1(z^s_t, z^d_t), \tag{B.7}$$

$$\text{and } \bar{h}(m^*_{t-1}, k^*_{t-1}, l^*_{t-1}, z^s_{t-1}, z^d_{t-1}, u^*_{t-1}, z^h_{t-1}) = c_2(z^s_{t-1}, z^d_{t-1}, z^h_{t-1}), \tag{B.8}$$

where $\{c_0(z^s_t, z^d_t), c_1(z^s_t, z^d_t)\}_{(z^s_t, z^d_t)\in\mathscr{Z}_s\times\mathscr{Z}_d}$ and $\{c_2(z^s_{t-1}, z^d_{t-1}, z^h_{t-1})\}_{(z^s_{t-1}, z^d_{t-1}, z^h_{t-1})\in\mathscr{Z}_s\times\mathscr{Z}_d\times\mathscr{Z}_h}$ are unknown constants. Without loss of generality, let $\left(z^{s*}_t, z^{d*}_t\right)$ in Assumption 4 be $z^{s*}_t = z^1_s$ and $z^{d*}_t = z^1_d$. Thus, the normalization in Assumption 4 is imposed as

$$c_0(z^1_s, z^1_d) = 0 \text{ and } c_1(z^1_s, z^1_d) = 1.$$

From the same step in the proof for Proposition 2, we can show that there exist $\left(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)$ and such some $q_{t-1} \in \{k_{t-1}, l_{t-1}, m_{t-1}, u_{t-1}, z^h_{t-1}\}$ that

$$\begin{aligned}\frac{\partial \mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it})}{\partial q_{it}} &= -\frac{\partial \bar{h}\left(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)}{\partial q_{it-1}} \\ &\quad \times \frac{\partial G_{m_t|v_t}\left(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)/\partial q_t}{\partial G_{m_t|v_t}\left(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)/\partial q_{t-1}}\end{aligned} \tag{B.9}$$

for all (m_{it}, w_{it}, u_{it}) and all $q_t \in \{m_t, k_t, l_t, u_t\}$. From (B.7) and (B.11), we have

$$\begin{aligned}1 &= c_1(z^1_s, z^1_d) - c_0(z^1_s, z^1_d) \\ &= \mathbb{M}^{-1}_t(m^*_{t1}, k^*_t, l^*_t, z^s_t, z^d_t, u^*_t) - \mathbb{M}^{-1}_t(m^*_{t0}, k^*_t, l^*_t, z^s_t, z^d_t, u^*_t) \\ &= -\frac{1}{S_{q_{t-1}}} \frac{\partial \bar{h}\left(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)}{\partial q_{t-1}}\end{aligned}$$

and therefore identify $\partial \bar{h}\left(\tilde{x}_{t-1}, \tilde{z}_{t-1}\right)/\partial q_{t-1}$ as

$$\frac{\partial \bar{h}\left(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)}{\partial q_{t-1}} = -\tilde{S}_{q_{t-1}}, \tag{B.10}$$

where

$$\tilde{S}_{q_{t-1}} := \left(\int_{m^*_{t0}}^{m^*_{t1}} \frac{\partial G_{m_t|v_t}\left(m|k^*_t, l^*_t, z^1_s, z^1_d, u^*_t, \tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)/\partial m_t}{\partial G_{m_t|v_t}\left(m|k^*_t, l^*_t, z^1_s, z^1_d, u^*_t, \tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)/\partial q_{t-1}} dm\right)^{-1}.$$

By substituting (B.10) into (B.9), we can identify $\partial \mathbb{M}^{-1}_t(m_t, k_t, l_t, z_t)/\partial q_t$ for $q_t \in$

$\{m_t, k_t, l_t, u_t\}$as

$$\frac{\partial \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})}{\partial q_t} = \tilde{S}_{q_{t-1}} T_{q_t q_{t-1}}(m_{it}, w_{it}, u_{it}), \tag{B.11}$$

where

$$T_{q_t q_{t-1}}(m_{it}, w_{it}, u_{it}) := \frac{\partial G_{m_t|v_t}\left(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}_{it-1}^h\right)/\partial q_t}{\partial G_{m_t|v_t}\left(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}_{it-1}^h\right)/\partial q_{t-1}}.$$

From (B.7) and (B.11), $\mathbb{M}_t^{-1}(x_t, z_t)$ is written as

$$\mathbb{M}_t^{-1}(x_t, z_t^s, z_t^s, u_t) = c_0(z_t^s, z_t^d) + \Lambda_m(x_t, z_t^s, z_t^d, u_t), \tag{B.12}$$

where

$$\begin{aligned}
&\Lambda_m\left(x_t, z_t^s, z_t^d, u_t\right) \\
&:= \tilde{S}_{q_{t-1}}\left\{\int_{m_{t0}^*}^{m_t} T_{m_t q_{t-1}}(s, k_t, l_t, z_t^s, z_t^d, u_t)ds + \int_{k_t^*}^{k_t} T_{k_t q_{t-1}}(m_{t0}^*, s, l_t, z_t^s, z_t^d, u_t)ds\right. \\
&\left. + \int_{l_t^*}^{l_t} T_{l_t q_{t-1}}(m_{t0}^*, k_t^*, s, z_t^s, z_t^d, u_t)ds + \int_{u_t^*}^{u_t} T_{l_t q_{t-1}}(m_{t0}^*, k_t^*, l^*, z_t^s, z_t^d, s)ds\right\}.
\end{aligned}$$

From Assumption B.3 (g), for a given point $(m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h) \in \mathscr{M} \times \mathscr{W} \times [0,1] \times \mathscr{Z}_h$, we can find some point $(\tilde{m}_t, \tilde{w}_t, \tilde{u}_t) \in \mathscr{M} \times \mathscr{W} \times [0,1]$ such that $\partial G_{m_t|v_t}(\tilde{m}_t|\tilde{w}_t, \tilde{u}_t, m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h)/\partial m_t > 0$. . Then, (B.9) identifies $\partial \bar{h}(x_{t-1}, z_{t-1})/\partial q_{t-1}$ as

$$\begin{aligned}
\frac{\partial \bar{h}\left(m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h\right)}{\partial q_{t-1}} = &-\frac{\partial G_{m_t|v_t}(\tilde{m}_t|\tilde{w}_t, \tilde{u}_t, m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h)/\partial q_{t-1}}{\partial G_{m_t|v_t}(\tilde{m}_t|\tilde{w}_t, \tilde{u}_t, m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h)/\partial m_t} \\
&\times \frac{\partial \mathbb{M}_t^{-1}(\tilde{m}_t, \tilde{w}_t, \tilde{u}_t)}{\partial m_t}.
\end{aligned}$$

Repeating this, we can identify $\partial \bar{h}(x_{t-1}, z_{t-1})/\partial q_{t-1}$ for all $(x_{t-1}, z_{t-1}) \in \mathscr{X} \times \mathscr{Z}$. From (B.7) and (B.10), we can write $\bar{h}_t(x_{t-1}, z_{t-1})$ as

$$\bar{h}_t(m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h) = c_2(z_{t-1}^s, z_{t-1}^d, z_{t-1}^h) + \Lambda_{\bar{h}}(m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h) \tag{B.13}$$

with

$$\begin{aligned}\Lambda_{\bar{h}}(m_{t-1},w_{t-1},u_{t-1},z_{t-1}^{h}) :=& \int_{m_{t-1}^{*}}^{m_{t-1}} \frac{\partial \bar{h}_t(s,k_{t-1},l_{t-1},z_{t-1}^{s},z_{t-1}^{d},u_{t-1}^{*},z_{t-1}^{h})}{\partial m_{t-1}}ds \\ &+ \int_{k_{t-1}^{*}}^{k_{t-1}} \frac{\partial \bar{h}_t(m_{t-1}^{*},s,l_{t-1},z_{t-1}^{s},z_{t-1}^{d},u_{t-1}^{*},z_{t-1}^{h})}{\partial k_{t-1}}ds \\ &+ \int_{l_{t-1}^{*}}^{l_{t-1}} \frac{\partial \bar{h}_t(m_{t-1}^{*},k_{t-1}^{*},s,z_{t-1}^{s},z_{t-1}^{d},u_{t-1}^{*},z_{t-1}^{h})}{\partial l_{t-1}}ds \\ &+ \int_{u_{t-1}^{*}}^{u_{t-1}} \frac{\partial \bar{h}_t(m_{t-1}^{*},k_{t-1}^{*},l_{t-1}^{*},z_{t-1}^{s},z_{t-1}^{d},s,z_{t-1}^{h})}{\partial u_{t-1}}ds.\end{aligned}$$

Therefore, we can identify $\mathbb{M}_t^{-1}(m,k_t,l_t,z_t)$ and $\bar{h}_t\left(x_{t-1},z_{t-1}\right)$ up to $\left\{c_0(z_t^s,z_t^d),c_2(z_{t-1}^s,z_{t-1}^d,z_{t-1}^h)\right\}_{z\in\mathscr{Z}}$.

Define $\widetilde{H}_t(z_t^s,z_t^d,z_{t-1}^s,z_{t-1}^d,z_{t-1}^h) := E[\Lambda_m(x_t,z_t^s,z_t^d,u_t) - \Lambda_{\bar{h}}(m_{t-1},w_{t-1},u_{t-1},z_{t-1}^h)|z_t^s,z_t^d,z_{t-1}^s,z_{t-1}^d,z_{t-1}^h]$. To determine $\left\{c_0(z_t^s,z_t^d)\right\}_{(z_t^s,z_t^d)\in\mathscr{Z}_s\times\mathscr{Z}_d}$ and $\left\{c_2(z_{t-1}^s,z_{t-1}^d,z_{t-1}^h)\right\}_{(z_{t-1}^s,z_{t-1}^d,z_{t-1}^h)\in\mathscr{Z}_s\times\mathscr{Z}_d}$, we evaluate

$$\begin{aligned}0 &= E\left[\eta_t|z_t^s,z_t^d,z_{t-1}^s,z_{t-1}^d,z_{t-1}^h\right] \\ &= E\left[\mathbb{M}_t^{-1}(m,k_t,l_t,z_t) - \bar{h}_t\left(x_{t-1},z_{t-1}\right)|z_t^s,z_t^d,z_{t-1}^s,z_{t-1}^d,z_{t-1}^h\right] \\ &= \widetilde{H}_t(z_t^s,z_t^d,z_{t-1}^s,z_{t-1}^d,z_{t-1}^h) + c_0(z_t^s,z_t^d) - c_2(z_{t-1}^s,z_{t-1}^d,z_{t-1}^h)\end{aligned}$$

at different values of $(z_t^s,z_t^d,z_{t-1}^s,z_{t-1}^d,z_{t-1}^h)$. First, evaluating $E\left[\eta_t|z_t^s,z_t^d,z_{t-1}^s,z_{t-1}^d,z_{t-1}^h\right]=0$ at $(z_t^s,z_t^d)=(z_s^1,z_d^1)$, and noting that $c_0(z_s^1,z_d^1)=0$, we have

$$c_2(z_{t-1}^s,z_{t-1}^d,z_{t-1}^h) = \widetilde{H}_t(z_s^1,z_d^1,z_{t-1}^s,z_{t-1}^d,z_{t-1}^h).$$

Therefore, $c_2(z)$ is identified for all $z\in\mathscr{Z}$. Second, evaluating $E\left[\eta_t|z_t^s,z_t^d,z_{t-1}^s,z_{t-1}^d,z_{t-1}^h\right]=0$ at $(z_{t-1}^s,z_{t-1}^d,z_{t-1}^h)=\left(z_s^1,z_d^1,z_h^1\right)$, we identify $c_0(z_t^s,z_t^d)$ as

$$\begin{aligned}c_0(z_t^s,z_t^d) &= c_2(z_s^1,z_d^1,z_h^1) - \widetilde{H}_t(z_t^s,z_t^d,z_s^1,z_d^1,z_h^1) \\ &= \widetilde{H}_t(z_s^1,z_d^1,z_s^1,z_d^1,z_h^1) - \widetilde{H}_t(z_t^s,z_t^d,z_s^1,z_d^1,z_h^1).\end{aligned}$$

Given that $\{c_0(z_t^s,z_t^d)\}_{(z_t^s,z_t^d)\in\mathscr{Z}_s\times\mathscr{Z}_d}$ and $\{c_2(z_{t-1}^s,z_{t-1}^d,z_{t-1}^h)\}_{(z_{t-1}^s,z_{t-1}^d,z_{t-1}^h)\in\mathscr{Z}_s\times\mathscr{Z}_d}$ are identified, we can identify $\mathbb{M}_t^{-1}(m_t,k_t,l_t,z_t)$ and $\bar{h}_t(x_{t-1},z_t)$ from (B.12) and (B.13). From $\eta_{it}=\omega_{it}-\bar{h}_t(x_{it-1},z_{it-1})$, we can identify the distribution of η_t, $G_{\eta_t}(\eta)$. □

Proposition B.2. *Suppose that Assumptions 3, 4, B.2, B.3, and 7 hold. Then, we can identify $\varphi_t^{-1}(\cdot)$ and $f_t(\cdot)$ up to scale and location and each firm's markup $\partial\varphi_t^{-1}(\bar{r}_{it},z_{it})/\partial r_t$ up to scale.*

Proof. From (37) and (39), the markup $\partial\varphi_t^{-1}(r_{it},z_{it}^d,u_{it})/\partial r_t$ is identified as (41).From ϕ_t and

(41), the markup function $\mu_t(m_t, w_t, u_t)$ is also identified as a function of (m_t, w_t, u_t) as (42). Substituting (42) into (37), we identify $\partial f_t(x_t, z_t^s)/\partial q_t$ for $q_t^s \in \{m_t, k_t, l_t\}$ as (43).

Define $c_f(z_t^s) := f_t(m_{t0}^*, k_t^*, l_t^*, z_t^s)$ for the point (m_{t0}^*, k_t^*, l_t^*) in normalization (14) and Assumption 4. Integrating $\partial f_t(x_t, z_t^s)/\partial q_t$ for $q_t^s \in \{m_t, k_t, l_t\}$ identifies $f_t(m_t, k_t, l_t, z_t^s)$ up to the constant $c_f(z_t^s)$:

$$f_t(m_t, k_t, l_t, z_t^s) = c_f(z_t^s) + \Lambda_f(m_t, k_t, l_t, z_t^s)$$

where

$$\begin{aligned}\Lambda_f(m_t, k_t, l_t, z_t^s) = &\int_{m_{t0}^*}^{m_t} \frac{\partial f_t(s, k_t, l_t, z_t^s)}{\partial m_t} ds + \int_{k_t^*}^{k_t} \frac{\partial f_t(m_{t0}^*, s, l_t, z_t^s)}{\partial k_t} ds \\ &+ \int_{l_t^*}^{l_t} \frac{\partial f_t(m_{t0}^*, k_t^*, s, z_t^s)}{\partial l_t} ds.\end{aligned}$$

From the normalization (14) and Assumption 4, we have

$$\begin{aligned}&\varphi_t^{-1}\left(\phi_t\left(m_{t0}^*, k_t^*, l_t^*, z_t^{s*}, z_t^{d*}, u_t^*\right), z_{it}^{d*}, u_t^*\right) \\ =&f_t(m_{t0}^*, k_t^*, l_t^*, z_t^{s*}) + \mathbb{M}_t^{-1}(m_{t0}^*, k_t^*, l_t^*, z_t^{s*}, z_t^{d*}, u_t^*) = 0\end{aligned}$$

and

$$\begin{aligned}&\varphi_t^{-1}\left(\phi_t\left(m_{t0}^*, k_t^*, l_t^*, z_t^s, z_t^{d*}, u_t^*\right), z_{it}^{d*}, u_t^*\right) \\ =&f_t(m_{t0}^*, k_t^*, l_t^*, z_t^s) + \mathbb{M}_t^{-1}(m_{t0}^*, k_t^*, l_t^*, z_t^s, z_t^{d*}, u_t^*) \\ =&c_f(z_t^s) + \mathbb{M}_t^{-1}(m_{t0}^*, k_t^*, l_t^*, z_t^s, z_t^{d*}, u_t^*).\end{aligned}$$

The integration of $\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})/\partial r_t$ leads to

$$\begin{aligned}&\int_{\phi_t\left(m_{t0}^*, k_t^*, l_t^*, z_t^{s*}, z_t^{d*}, u_t^*\right)}^{\phi_t\left(m_{t0}^*, k_t^*, l_t^*, z_t^s, z_t^{d*}, u_t^*\right)} \frac{\partial \varphi_t^{-1}(s, z_{it}^{d*}, u_{it}^*)}{\partial r_t} ds \\ =&\varphi_t^{-1}\left(\phi_t\left(m_{t0}^*, k_t^*, l_t^*, z_t^s, z_t^{d*}, u_t^*\right), z_{it}^{d*}, u_t^*\right) \\ =&c_f(z_t^s) + \mathbb{M}_t^{-1}(m_{t0}^*, k_t^*, l_t^*, z_t^s, z_t^{d*}, u_t^*).\end{aligned}$$

Since $\phi_t(\cdot)$, $\partial\varphi_t^{-1}(\cdot)/\partial r_t$ and $\mathbb{M}_t^{-1}(\cdot)$, $c_f(z_t^s)$ is identified as

$$c_f(z_t^s) = \int_{\phi_t\left(m_{t0}^*,k_t^*,l_t^*,z_t^{s*},z_t^{d*},u_t^*\right)}^{\phi_t\left(m_{t0}^*,k_t^*,l_t^*,z_t^s,z_t^{d*},u_t^*\right)} \frac{\partial\varphi_t^{-1}(s,z_{it}^{d*},u_{it}^*)}{\partial r_t} ds$$
$$-\mathbb{M}_t^{-1}(m_{t0}^*,k_t^*,l_t^*,z_t^s,z_t^{d*},u_t^*).$$

Thus, $f_t(x_t,z_t^s)$ is identified.

For given $(r_t,z_t^d) \in \mathscr{R} \times \mathscr{Z}_d$, $B_t\left(r_t,z_t^d,u_t\right) := \left\{\left(x_t,z_t^s\right) \in \mathscr{X} \times \mathscr{Z}_s : \phi_t\left(x_t,z_t^s,z_t^d,u_t\right) = r_t\right\}$ is non-empty by the construction of $\mathscr{R}$. The output quantity $\varphi_t^{-1}(r_t,z_t,u_t)$ for any $(r_t,z_t,u_t) \in \mathscr{R} \times \mathscr{Z} \times [0,1]$ is identified by

$$\varphi_t^{-1}(r_t,z_t^d,u_t) = f_t(x_t,z_t^s) + \mathbb{M}_t^{-1}(m_t,w_t,u_t) \text{ for } \left(x_t,z_t^s\right) \in B_t(r_t,z_t^d,u_t).$$

The output price for individual firms is identified as

$$p_{it} := r_{it} - \varphi_t^{-1}(r_{it},z_{it}^d,u_{it}).$$

□

B.2.2 Endogenous Characteristics

Firm characteristics (z_{it}^s,z_{it}^d) may correlate with u_{it} and η_{it}.In step 1, we can use $(z_{it-\upsilon}^s,z_{it-\upsilon}^d)$ instead of $(z_{it-\upsilon-1}^s,z_{it-\upsilon-1}^d)$ as instrument variables to construct the moment condition similar to (29). In step 2, we consider the control variable approach as in subsection B.1. Using the same steps in subsection B.1, we can derive (B.5) and (B.6) for continuous variables $q_{it} \in \{m_{it},k_{it},l_{it},u_{it}\}$ and $q_{it-1} \in \{m_{it-1},k_{it-1},l_{it-1},u_{it-1}\}$. Taking their ratios, we have (33). Therefore, we can apply the same steps in subsection B.2.1 to establish identification.