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

Chun Pang Chow[†] Hiroyuki Kasahara[‡] Yoichi Sugita[§]

February 19, 2026

Abstract

We establish nonparametric identification of production functions, total factor productivity (TFP), price markups, and firms' output prices and quantities, as well as consumer demand, using firm-level revenue data, without observing output quantity, in a monopolistically competitive environment with a fully nonparametric demand system. This result overturns the widely held view—formalized by Bond, Hashemi, Kaplan, and Zoch (2021)—that output elasticities and markups are not nonparametrically identifiable from revenue data without quantity information. Under the additional restriction that demand satisfies the homothetic single-aggregator (HSA) structure of Matsuyama and Ushchev (2017), we further nonparametrically identify the representative consumer's utility function from firm-level revenue data. This new identification result enables counterfactual welfare analysis without parametric assumptions on preferences. We propose a semiparametric estimator that is feasible for standard firm-level datasets under a Cobb–Douglas production specification. Monte Carlo simulations show that the estimator performs well, while treating revenue as output induces substantial bias. Applying the estimator to Chilean manufacturing data, we reject the CES specification in favor of HSA, and find that market power reduces welfare by approximately 3%–6% of industry revenue in the three largest manufacturing industries in 1996.

Acknowledgement: We are grateful for comments made at seminars and conferences regarding an earlier version of this work. We thank Zheng Han, Kohei Hashinokuchi, Yoko Sakamoto, and Makoto Tanaka for excellent research assistance. Sugita acknowledges financial support from JSPS KAKENHI (grant numbers: 17H00986, 19H01477, 19H00594, 20H01498, 22H00060). Hiroyuki Kasahara acknowledges financial support from the SSHRC Insight Grant.

[†]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 derivatives of the function ϕ [with respect to inputs] will be random for . . . differences in the prices paid or received by various firms if we drop, as we realistically must, the assumption of perfect competition.*
>
> — Marschak and Andrews (1944, p. 145)

The estimation of production functions and markups is a central tool in empirical analyses of market outcomes, underpinning research on firm-level productivity (Bartelsman and Doms, 2000; Syverson, 2011), aggregate productivity and misallocation (Olley and Pakes, 1996; Hsieh and Klenow, 2009), technological change (Van Biesebroeck, 2003; Doraszelski and Jaumandreu, 2018), and the evolution of market power (Hall, 1988; De Loecker and Warzynski, 2012; De Loecker et al., 2020).[1] A common assumption underlying many estimation methods is that firms' output quantities are observable. In practice, however, most firm-level datasets contain only revenue, and researchers typically deflate revenue using an industry-level price deflator.[2]

Following Marschak and Andrews (1944)'s pioneering critique,[3] a large body of subsequent research has shown that replacing output quantity with revenue can severely bias estimates of production function parameters (e.g., Klette and Griliches, 1996; De Loecker, 2011), TFP (e.g., Foster, Haltiwanger, and Syverson, 2008; Katayama, Lu, and Tybout, 2009), and markups (Bond, Hashemi, Kaplan, and Zoch, 2021). Bond, Hashemi, Kaplan, and Zoch (2021) formalize this difficulty, arguing that "in the usual setting in which the researcher observes only revenue, and does not have separate information on the price and quantity of output, the output elasticity for a flexible input is not identified non-parametrically from estimation of the revenue production function" (p. 2). Under a nonparametric demand function, the identification challenge has two dimensions: revenue becomes a nonparametric function of inputs, unobserved TFP, and an unobserved demand shock; and flexible inputs may be correlated with both TFP and the demand shock.[4] Despite these concerns, researchers continue to use revenue in place of quantity due to

[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;Doraszelski and Jaumandreu, 2013, 2018;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] Marschak and Andrews were acutely aware that economic data typically come in the form of monetary values (sales, value added) rather than physical counts (tons, bushels). In the introduction to Marschak and Andrews (1944), they explicitly flag the danger of conflating these concepts. In Footnote 3, attached to the definition of net output (x_0), they write: "[S]ome changes in notation will be needed later to distinguish between physical output; (gross) revenue; net revenue ('net output'): see §§4 and 9."

[4] As noted by Klette and Griliches (1996), Marschak and Andrews (1944) were the first to recognize these two identification challenges and to criticize the practice of replacing output quantities with revenue in production function estimation.

the scarcity of firm-level quantity data.[5]

In this paper, we establish that, under monopolistic competition, nonparametric identification of the production function, total factor productivity, consumer demand, and counterfactual welfare effects is in fact possible using firm-level revenue data, without observing output quantity. We consider the same demand-side setting as Bond et al. (2021), in which an individual firm faces a nonparametric demand function that depends on its output, observable characteristics, and an unobserved demand shock; for identification, we allow this shock to be transitory.[6] Our identification proof is constructive: it yields closed-form mappings from observables to the objects of interest. The proof combines standard assumptions maintained in the proxy variable literature (e.g., Levinsohn and Petrin, 2003), such as strict monotonicity of input demand, with the first-order condition approach of Doraszelski and Jaumandreu (2013, 2018) and Gandhi et al. (2020), while imposing additional nonparametric restrictions on firms' demand functions.

To address the two identification challenges described above, we develop a three-step approach that combines the control function method of Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves, and Frazer (2015) with the first-order condition approach of Doraszelski and Jaumandreu (2013, 2018) and Gandhi, Navarro, and Rivers (2020) in a novel way. The key insight is to treat the control function—the inverse of a material demand function, which serves as a proxy for TFP—not merely as an auxiliary device, but as an object of nonparametric identification in its own right. In the first step, we identify the unobserved transitory demand shock that nonlinearly affects revenue by combining the control function as in Ackerberg et al. (2015) with the instrumental variable quantile regression of Chernozhukov and Hansen (2005). Intuitively, this step separates demand-side variation from productivity variation in revenue, exploiting the fact that lagged information shifts inputs and TFP but not the transitory demand shock. In the second step, we identify 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). This step recovers TFP (up to normalization) from the dynamics of inputs and the demand shock, without requiring output quantity data, because the control function maps observable inputs and the identified demand shock into unobserved productivity. In the third step, we identify the production function, markups, and the demand function using the first-order condition for materials and the control function identified in the second step.

Our method identifies several key objects from revenue data. In our main setting, markups

[5]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.

[6]In the appendix, we also consider identification under persistent demand shocks (e.g., AR(1)) by utilizing lagged firm characteristics (such as R&D) as instruments or by explicitly modeling persistent unobserved quality heterogeneity.

and output elasticities are identified up to scale, while the output price, output quantity, TFP, gross production function, and consumer demand function are identified up to scale and location—normalizations that are standard in nonparametric settings. Proposition C.1 in Appendix C formally characterizes this equivalence class, showing that any existing identification result for output elasticities or markup levels from revenue data necessarily imposes location and/or scale normalizations, whether explicitly or implicitly. Identification is cross-sectional, allowing these objects to vary over time. With the additional assumption of *local* constant returns to scale, we 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.[7]

By assuming the homothetic single-aggregator (HSA) demand system of Matsuyama and Ushchev (2017), we further identify the firms' demand system and the representative consumer's utility function nonparametrically, without imposing parametric functional form restrictions.[8] The HSA class—which nests CES while permitting variable markups, incomplete pass-through, and non-monotonic relationships between firm size and markups (see Matsuyama, 2023; Matsuyama, 2025 for comprehensive reviews)—has become a leading framework for monopolistic competition with heterogeneous firms. To our knowledge, the nonparametric identification of the consumer's utility function within the HSA class from firm-level data is a new result; it enables counterfactual welfare analysis under flexible demand without the specification errors that arise from imposing CES.

Our result bridges two literatures that have largely operated in isolation: the production function literature, which recovers supply-side objects from revenue data, and the demand estimation literature, which recovers consumer-side objects from market-level data. In the existing production function literature, empirical measures of productivity or markups constructed from revenue data neither permit welfare evaluation nor counterfactual analysis, because the underlying consumer demand system and utility function remain unidentified. Our result enables counterfactual welfare analysis even when only revenue data are available. In particular, we show how to compute a counterfactual marginal-cost-pricing equilibrium and compare it with the observed monopolistic-competition equilibrium, thereby enabling a structural evaluation of firms' market power and its welfare consequences through counterfactual changes in prices, quantities, consumer utility, and firm profits.

While the nonparametric identification results establish informational sufficiency in principle,

[7] Flynn, Gandhi, and Traina (2019) impose *global* constant returns to scale to identify a production function. In Subsection 2.5.2, we clarify the distinction between *local* and *global* constant returns to scale.

[8] In the Appendix, we establish identification of two additional families of homothetic demand systems proposed by Matsuyama and Ushchev (2017), the homothetic demand systems with direct implicit additivity (HDIA) and those with indirect implicit additivity (HIIA), and discuss how to conduct counterfactual analyses using them.

practical estimation with moderate sample sizes requires additional structure. We develop a semiparametric estimator that assumes a Cobb-Douglas production function but leaves the demand system unrestricted. The estimation proceeds in three steps. First, we nonparametrically estimate the transitory demand shock using the smooth GMM IV quantile regression of Firpo, Galvao, Pinto, Poirier, and Sanroman (2022), which ensures quantile monotonicity. Second, we estimate the control function via the profile likelihood estimator of Linton, Sperlich, and Van Keilegom (2008). Third, we recover the production function, markups, and TFP. This three-step procedure provides a standalone estimate of the production function without parametric demand assumptions. In a fourth step, for counterfactual welfare analysis and testing the CES restriction, we estimate the CoPaTh-HSA demand system of Matsuyama and Ushchev (2020).

Simulation results show that our estimator performs well in recovering structural parameters, markups, and TFP. 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 in favor of the HSA demand system. Our counterfactual welfare analysis reveals that market power results in welfare losses of approximately 3%–6% of industry revenue in the three largest Chilean manufacturing industries in 1996. To put this in context, these losses exceed standard Harberger-triangle calculations and are broadly consistent with the welfare costs of markups estimated by Edmond et al. (2023) and the productivity losses from misallocation analyzed by Baqaee and Farhi (2020), though our estimates are derived from a different structural framework.

Our analysis contributes to a growing literature that addresses the revenue-versus-quantity problem in production function estimation. Klette and Griliches (1996), De Loecker (2011), and Gandhi et al. (2020) impose a CES demand structure under which log revenue is linear in log inputs and log TFP; in contrast, we identify the demand system and the representative consumer's utility function nonparametrically, without relying on parametric assumptions. Flynn, Gandhi, and Traina (2019) achieve identification under global constant returns to scale; we require only local constant returns to scale for certain normalizations. Recent work by Demirer (2025) studies factor misallocation using revenue data, Doraszelski and Jaumandreu (2021) reexamine the internal consistency of the De Loecker and Warzynski (2012) method under market power, De Ridder et al. (forthcoming) provides practical guidance on production function estimation, and Kirov et al. (2025) develops alternative approaches to measuring markups. Our framework complements these contributions by providing nonparametric identification of both the production and demand sides from revenue data. On the demand side, our identification of the HSA utility function contributes to the growing literature on non-CES demand systems under monopolistic competition (Matsuyama and Ushchev, 2017; Matsuyama, 2023; Matsuyama, 2025), providing a new empirical foundation for counterfactual welfare analysis within this class.[9]

The remainder of this paper is organized as follows. Subsection 2.1 introduces the model and setting, while Subsection 2.2 illustrates our three-step identification strategy through a parametric example. Subsection 2.3 establishes the main nonparametric identification results, and Subsection 2.4 discusses alternative settings. Subsection 2.5 introduces additional assumptions to fix scale and location normalizations, and Subsection 2.6 discusses the identification of the demand system and counterfactual analysis. Section 3 presents our semiparametric estimator. Section 4 reports simulation results. Section 5 provides an empirical application. Section 6 concludes. The Appendix contains proofs and simulation details, and presents identification results for several extensions, including endogenous labor input, endogenous or discrete firm characteristics, persistent demand shocks, unobserved quality heterogeneity, and two additional families of homothetic demand systems proposed by Matsuyama and Ushchev (2017).

2 Identification

2.1 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

$$y_{it} = f_t\big(x_{it}, z^s_{it}\big) + \omega_{it}, \tag{1}$$

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).[10] Firm-level productivity ω_{it} follows a first-order Markov process given by

$$\omega_{it} = h_t\big(\omega_{it-1}, z^h_{it-1}\big) + \eta_{it}, \qquad \eta_{it} \overset{iid}{\sim} G_{\eta_t}, \tag{2}$$

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

[9] 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.

[10] Throughout the paper, the subscript t on functions (e.g., f_t, ψ_t) indicates that the structural form is time-varying (common to all firms at time t), while the subscript it denotes firm-specific realizations.

import status as in Kasahara and Rodrigue, 2008).

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

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

where z_{it}^d 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.

We assume the demand shock ϵ_{it} has limited persistence to facilitate identification via lagged instruments, and interpret ϵ_{it} as the demand fluctuation remaining after conditioning on observable characteristics. Specifically, ϵ_{it} is generated by

$$\epsilon_{it} = \Upsilon_t(\zeta_{it}, \zeta_{it-1}, ..., \zeta_{it-\upsilon}), \qquad \zeta_{it-s} \stackrel{iid}{\sim} F_{\zeta_t} \quad \text{for } s = 0, 1, .., \upsilon. \tag{3}$$

Therefore, conditional on z_{it}^d, 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: 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 effect on future productivity in (2). This difference between the demand and supply shock specifications in (2) and (3) captures the idea that demand shocks are temporary while supply shocks are permanent (e.g., Nelson and Plosser, 1982). Appendix F.4 provides an alternative identification approach and shows that identification remains possible under persistent demand shocks when a supply-side instrument is available. Appendix F.5 discusses how productivity ω_{it} may capture persistent differences in product quality across firms, where the firm's output y_{it} can be interpreted as a quality-adjusted measure of output quantity; from this perspective, heterogeneous demands attributable to quality differences are captured by ω_{it}.

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_t} be the c.d.f. of ϵ_{it}. Without loss of generality, we transform ϵ_{it} to a uniform variable, using $u_{it} := F_{\epsilon_t}(\epsilon_{it})$,

$$p_{it} = \tilde{\psi}_t(y_{it}, z_{it}^d, F_{\epsilon_t}^{-1}(u_{it})) = \psi_t(y_{it}, z_{it}^d, u_{it}), \;\; u_{it} \sim \text{Unif}(0, 1). \tag{4}$$

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

The inverse demand function (4) 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 (4) implicitly imposes two key assumptions. First, $\psi_t(\cdot, z_{it}^d, u_{it})$ is common across firms once we control for observed demand characteristics z_{it}^d and a transitory scalar unobserved demand shock u_{it}. Second, $\psi_t(\cdot, z_{it}^d, u_{it})$ 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 $\psi_t(y_{it}, z_{it}^d, u_{it}, a_t)$, 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 (1), the observed revenue relates to output and input as follows:

$$\begin{aligned} r_{it} &= \varphi_t(y_{it}, z_{it}^d, u_{it}) && (5) \\ &= \varphi_t\left(f_t(m_{it}, k_{it}, l_{it}, z_{it}^s) + \omega_{it}, z_{it}^d, u_{it}\right) && (6) \end{aligned}$$

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

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 $(\omega_{it}, u_{it}, z_{it}^s, z_{it}^d)$ but is independent of η_{is} and ζ_{is} for $s \geq t+1$. (c) $\left(z_{it}^s, z_{it}^d, z_{it-1}^h\right)$ 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).[11] In Appendix F.1, we present identification results when l_{it} is also endogenous. The continuity requirement in Assumption 1(c) can be relaxed, but the exogeneity of $\left(z_{it}^s, z_{it}^d, z_{it-1}^h\right)$ remains an important—albeit potentially strong—assumption, though it is commonly maintained in the empirical literature.[12] Assumption 1(d) is also standard in most empirical applications.[13] Appendix F.6 shows that the identification strategy extends to such environments when material prices vary across firms as a function of observed characteristics.

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

$$\mathbb{M}_t\left(\omega_{it}, k_{it}, l_{it}, z_{it}^s, z_{it}^d, u_{it}\right) \in \arg\max_m \exp(\varphi_t(f_t(m, k_{it}, l_{it}, z_{it}^s) + \omega_{it}, z_{it}^d, u_{it})) - \exp(p_t^m + m), \quad (7)$$

where p_t^m denotes the logarithm of the material price at time t.

[11] Stochastic independence between (l_{it}, k_{it}) and η_{it} is stronger than the standard mean independence assumption $E[\eta_{it} \mid \mathscr{I}_{it}] = 0$, where $\mathscr{I}_{it}$ is the firm's information set at the beginning of period t. Most identification results below require only mean independence; full stochastic independence is used in the IVQR step (Proposition 1) and the control function argument (Proposition 2).

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

[13] We treat deflated expenditures on materials as measures of inputs. This abstracts from unobserved heterogeneity in material prices. For instance, geographically segmented input markets may induce systematic price differences across regions.

Equation (6) 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 $\varphi_t(\cdot, z_{it}^d, u_{it})$ and $f_t(\cdot)$. From the first-order condition for profit maximization, $P_{it}\left(1 + \partial\psi_t(y_{it}, z_{it}^d, u_{it})/\partial y_{it}\right) = MC_{it}$, where P_{it} and MC_{it} denote the price and marginal cost of output, respectively, the elasticity of revenue with respect to output equals the inverse of the markup:

$$\frac{\partial\varphi_t(y_{it}, z_{it}^d, u_{it})}{\partial y_{it}} = \frac{MC_{it}}{P_{it}}. \tag{8}$$

Thus, the revenue elasticity relates to the output elasticity via markup:

$$\frac{\partial\varphi_t(f_t(x_{it}) + \omega_{it}, z_{it}^d, u_{it})}{\partial v_{it}} = \frac{MC_{it}}{P_{it}}\frac{\partial f_t(x_{it})}{\partial v_{it}} \text{ for } v_{it} \in \{m_{it}, k_{it}, l_{it}\}.$$

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 $\varphi_t^{-1}(r, z^d, u)$, 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 $\mathbb{M}_t^{-1}(m, k, l, z^s, z^d, u)$, 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}.*

Assumptions 2(a)(b) are standard assumptions about smooth production and demand functions. In Assumption 2(b), the condition $\partial\varphi_t(y_{it}, z_{it}^d, u_{it})/\partial y_{it} > 0$ is equivalent to that the elasticity of demand with respect to price, $-\left(\partial\psi_t(y_{it}, z_{it}^d, u_{it})/\partial y_{it}\right)^{-1}$, 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 $w_{it} := (k_{it}, l_{it}, z_{it}^s, z_{it}^d)$ be observable exogenous variables at t. The inverse function of the material demand function with respect to TFP

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

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

$$\varphi_t^{-1}\left(r_{it}, z_{it}^d, u_{it}\right) = f_t(x_{it}, z_{it}^s) + \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}). \tag{9}$$

Let $v_{it} := (w_{it}, u_{it}, m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h)' \in \mathcal{V} := \mathcal{W} \times [0,1] \times \mathcal{M} \times \mathcal{W} \times [0,1] \times \mathcal{Z}_h$, where $\mathcal{W} := \mathcal{K} \times \mathcal{L} \times \mathcal{Z}_s \times \mathcal{Z}_d$. 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 (9). Then, for any $(a_{1t}, a_{2t}, b_t) \in \mathbb{R}^2 \times \mathbb{R}_{++}$,

$$\begin{aligned} \varphi_t^{-1}\left(r_{it}, z_{it}^d, u_{it}\right) = (a_{1t} + a_{2t}) + b_t \varphi_t^{*-1}\left(r_{it}, z_{it}^d, u_{it}\right),\ f_t(x_{it}, z_{it}^s) = a_{1t} + b_t f_t^*(x_{it}, z_{it}^s), \\ \text{and } \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}) = a_{2t} + b_t \mathbb{M}_t^{*-1}(m_{it}, w_{it}, u_{it}) \end{aligned} \tag{10}$$

also satisfy (9). Hence, the true structure $\{\varphi_t^{*-1}(\cdot), f_t^*(\cdot), \mathbb{M}_t^{*-1}(\cdot)\}$ is observationally equivalent to the structure (10) and is therefore identified only up to location and scale normalization (a_{1t}, a_{2t}, b_t) from restriction (9).

Normalization is unavoidable absent further assumptions. Because the unobserved output level y_{it} enters the revenue function through the unknown nonlinear function φ_t in (5), it has no natural scale or location, so the structural functions are identified only up to a class of transformations. Proposition C.1 in Appendix C formally characterizes this equivalence class for $(\varphi_t^{-1}, f_t, \mathbb{M}_t^{-1})$; accordingly, any existing identification result for output elasticities or markup level from revenue data either explicitly or implicitly imposes location and/or scale normalizations.

We may fix (a_{1t}, a_{2t}, b_t) in (10) 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_{t1}^*, w_t^*, u_t^*) on the support $\mathcal{M} \times \mathcal{W} \times [0,1]$ where $m_{t0}^* < m_{t1}^*$, we denote

$$c_{1t} := f_t(m_{t0}^*, k_t^*, l_t^*, z_t^{s*}),\ c_{2t} = \mathbb{M}_t^{-1}(m_{t0}^*, w_t^*, u_t^*),\ \text{and } c_{3t} := \mathbb{M}_t^{-1}(m_{t1}^*, w_t^*, u_t^*). \tag{11}$$

Note that $\partial \mathbb{M}_t^{-1}/\partial m_{it} > 0$ implies that $c_{2t} < c_{3t}$. Then, there exists a unique one-to-one mapping between (c_{1t}, c_{2t}, c_{3t}) in (11) and (a_{1t}, a_{2t}, b_t) in (10) such that $b_t = (c_{3t} - c_{2t}) / \left(\mathbb{M}_t^{*-1}(m_{t1}^*, w_t^*, u_t^*) - \mathbb{M}_t^{*-1}(m_{t0}^*, w_t^*, u_t^*)\right)$, $a_{1t} = c_{1t} - b_t f_t^*(m_{t0}^*, k_t^*, l_t^*, z_t^{s*})$ and $a_{2t} = c_{2t} - b_t \mathbb{M}_t^{*-1}(m_{t0}^*, w_t^*, u_t^*)$. 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_{t1}^*, w_t^*, u_t^*) such that $c_{1t} = c_{2t} = 0$ and $c_{3t} = 1$ in (11).*

As Chiappori et al. (2015) demonstrates, this choice of normalization makes the identification proofs transparent. In Section 2.5, we discuss how we can use additional restrictions and data to identify the normalization parameters (a_{1t}, a_{2t}, b_t).

2.2 Identification in a Parametric Example: Generalized CES Demand with Heterogeneity

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 with heterogeneity:

$$p_{it} = \alpha_t(u_{it}) + (\rho(u_{it}) - 1) y_{it}, \tag{12}$$

where $\alpha_t(\cdot)$ and $\rho(\cdot)$ are unknown functions, where $0 < \rho(\cdot) \leq 1$.[14] 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} that follows a first-order autoregressive (AR(1)) process:

$$y_{it} = \theta_0 + \theta_m m_{it} + \theta_k k_{it} + \theta_l l_{it} + \omega_{it}, \quad \omega_{it} = h_1 \omega_{it-1} + \eta_{it}, \tag{13}$$

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

$$r_{it} = \alpha_t(u_{it}) + \rho(u_{it})\theta_0 + \rho(u_{it})\theta_m m_{it} + \rho(u_{it})\theta_k k_{it} + \rho(u_{it})\theta_l l_{it} + \rho(u_{it})\omega_{it}. \tag{14}$$

Denote the ratio of material cost to revenue as $s_{it}^m := \frac{\exp(p_t^m + m_{it})}{\exp(r_{it})}$. Then, the first-order condition

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

for (7) can be written as

$$\rho(u_{it})\theta_m = s^m_{it}, \tag{15}$$

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

$$\omega_{it} = \mathbb{M}_t^{-1}(m_{it}, k_{it}, l_{it}, u_{it}) = \beta_t(u_{it}) + \beta_m(u_{it})m_{it} + \beta_k k_{it} + \beta_l l_{it} \tag{16}$$

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

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

$$\begin{aligned} &0 = f_t(0,0,0) = \theta_0,\ 0 = \mathbb{M}_t^{-1}(0,0,0,0.5) = \beta_t(0.5), \\ &1 = \mathbb{M}_t^{-1}(1,0,0,0.5) = \beta_t(0.5) + \beta_m(0.5) \end{aligned} \tag{17}$$

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

Our identification approach follows three steps.

2.2.1 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

$$\begin{aligned} r_{it} =& (\alpha_t(u_{it}) + \rho(u_{it})\beta_t(u_{it})) + \rho(u_{it})(\theta_m + \beta_m(u_{it}))\, m_{it} \\ &+ \rho(u_{it})(\theta_k + \beta_k)\, k_{it} + \rho(u_{it})(\theta_l + \beta_l)\, l_{it} \qquad (18)\\ =& \tilde{\phi}_t(u_{it}) + m_{it}, \qquad (19) \end{aligned}$$

where the second equality uses $\beta_k = -\theta_k$ and $\beta_l = -\theta_l$ from (16), so that $\rho(u_{it})(\theta_k + \beta_k) = \rho(u_{it})(\theta_l + \beta_l) = 0$, and $\tilde{\phi}_t(u_{it}) := \alpha_t(u_{it}) + \rho(u_{it})\beta_t(u_{it})$ with $\tilde{\phi}_t'(u) = -\theta_m\rho'(u)/\rho(u) > 0$ for all u.

From (19), we have $\Pr[r_{it} - m_{it} \le \tilde{\phi}_t(u)] = u \quad$ 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 (15) and the monotonicity of $\rho(\cdot)$ with $\rho'(\cdot) < 0$, the demand shock u_{it} is identified as the quantile of $1/s^m_{it}$. This equivalence arises because the quantile of $r_{it} - m_{it}$ coincides with that of $1/s^m_{it} = \exp(r_{it} - m_{it} - p^m_t)$.

Step 2: Identification of Control Function and TFP The second step identifies the control function $\mathbb{M}_t^{-1}(\cdot)$. Substituting (16) into the AR(1) process (13) leads to

$$\mathbb{M}_t^{-1}(m_{it},k_{it},l_{it},u_{it}) = h_1\mathbb{M}_{t-1}^{-1}(m_{it-1},k_{it-1},l_{it-1},u_{it-1}) + \eta_{it}. \tag{20}$$

Since $\mathbb{M}_t^{-1}(m_{it},k_{it},l_{it},u_{it})$ is linear in m_{it} from (16), we can rearrange (20) as:

$$\begin{aligned} m_{it} &= \gamma(u_{it},u_{t-1}) + \gamma_k(u_{it})k_{it} + \gamma_l(u_{it})l_{it} + \delta_m(u_{it},u_{it-1})m_{it-1} \\ &\quad + \delta_k(u_{it})k_{it-1} + \delta_l(u_{it})l_{it-1} + \tilde{\eta}_{it}, \end{aligned} \tag{21}$$

where

$$\gamma_k(u_{it}) = -\frac{\beta_k}{\beta_m(u_{it})},\ \gamma_l(u_{it}) = -\frac{\beta_l}{\beta_m(u_{it})},\ \delta_k(u_{it}) = \frac{h_1\beta_k}{\beta_m(u_{it})},\ \delta_l(u_{it}) = \frac{h_1\beta_l}{\beta_m(u_{it})}, \tag{22}$$

$$\gamma(u_{it},u_{it-1}) = \frac{-\beta_t(u_{it}) + h_1\beta_{t-1}(u_{it-1})}{\beta_m(u_{it})}, \tag{23}$$

$\tilde{\eta}_{it} = \eta_{it}/\beta_m(u_{it})$, 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}), (21) 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 (21) from the conditional moment restriction $E[\,\tilde{\eta}_{it}|\,v_{it}] = 0$.

Then, because $\beta_m(0.5) = 1$, we can recover (θ_k,θ_l,h_1) from (22) 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 (17) to (23), 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 (22)-(23) 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 (16).

Step 3: Identification of Production Function, Markup, and Demand Function The identification of $\rho(u)$ follows from substituting (15) into $\beta_m(u) = (1-\rho(u)\theta_m)/\rho(u)$, and rearranging the terms, which yields

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

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

The first order condition (15) implies that the revenue share of material expenditure is a function of u_{it}, which we denote by $s(u)$, such that $s_{it}^m = s(u_{it})$. 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 (15) as

$$\theta_m = \frac{s(0.5)}{\rho(0.5)} = \frac{s(0.5)}{1 - s(0.5)}. \tag{24}$$

Appendix D further discusses counterfactual analysis under the CES specification.

The Identification under Normalization In view of the first-order condition $\rho(u_{it})\theta_m = s_{it}^m$, 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 (17). 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 2.5 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 2.1.

2.3 Nonparametric Identification

2.3.1 Step 1: Identification of the Demand Shocks

Substituting $\omega_{it} = \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})$ into the revenue function (6), we can rewrite it as

$$\begin{aligned} r_{it} &= \varphi_t\left(f_t(m_{it}, k_{it}, l_{it}, z_{it}^s) + \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}), z_{it}^d, u_{it}\right) \\ &=: \phi_t\left(m_{it}, w_{it}, u_{it}\right),\ \ u_{it} \sim \text{Unif}(0, 1). \end{aligned} \tag{25}$$

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* $g(m,w)$*,*

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

Assumption 5(a) implicitly imposes restrictions on the shape of the demand function. The Appendix B.1 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 $\sigma_t(y, z^d, u) := -1/\left(\frac{\partial \psi_t(y,z^d,u)}{\partial y}\right) > 0$ denotes 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 (25) can be interpreted as the quantile function of revenue r. Although m_{it} is endogenous and correlated with u_{it}, Assumption 1(a)–(c) and equation (3) 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 $\Pr[r_{it} \leq \phi_t(m_{it}, w_{it}, u) | m_{it-v-1}, w_{it-v}] = u$ for all $u \in [0,1]$.[15]

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 $u \in [0,1]$, $E\left[1\left\{r_{it} \leq \phi_t^1(m_{it}, w_{it}, u)\right\}\middle| m_{it-v-1}, w_{it-v}\right] = E\left[1\left\{r_{it} \leq \phi_t^2(m_{it}, w_{it}, u)\right\}\middle| m_{it-v-1}, w_{it-v}\right]$ a.s. 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

$$E\left[1\{r_{it} \leq \phi_t(m_{it}, w_{it}, u)\} - u | m_{it-v-1}, w_{it-v}\right] = 0 \text{ for } u \in [0,1] \tag{26}$$

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.

[15] This follows because

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

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

2.3.2 Step 2: Identification of Control Function and TFP

From (2), the control function $\omega_{it} = \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})$ satisfies

$$\mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}) = \bar{h}_t\left(m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h\right) + \eta_{it}, \tag{27}$$

where $\bar{h}_t\left(m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h\right) := h_t\left(\mathbb{M}_{t-1}^{-1}(m_{it-1}, w_{it-1}, u_{it-1}), z_{it-1}^h\right)$. As $\partial \mathbb{M}_t^{-1}/\partial m_{it} > 0$, given the values of (w_{it}, u_{it}), the dependent variable in (27) is a monotonic transformation of m_{it}. Therefore, the model (27) 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).[16]

Assumption 6. *(a) The distribution $G_{\eta_t}(\cdot)$ of η_{it} is absolutely continuous with a density function $g_{\eta_t}(\cdot)$ that is continuous on its support. (b) η_{it} is independent of $v_{it} := (w_{it}, u_{it}, m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h)'$ with $E[\eta_{it}|v_{it}] = 0$. (c) The support $\mathscr{M} \times \mathscr{V}$ of the vector (m_{it}, v_{it}') is an open, connected subset of Euclidean space, and the normalization point defined in Assumption 4 lies in the interior of $\mathscr{M} \times \mathscr{V}$. (d) The 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) Let $\mathscr{S}_{m,w,u}$ denote the joint support of (m_{it}, w_{it}, u_{it}). The set*

$$\mathscr{A}_{q_{t-1}} := \left\{(m_{it-1}, w_{it-1}, u_{it-1}, z_{it-1}^h) \in Proj_v(\mathscr{M} \times \mathscr{V}) : \frac{\partial G_{m_t|v_t}(m_{it}|v_{it})}{\partial q_{it-1}} \neq 0 \text{ for all } (m_{it}, w_{it}, u_{it}) \in \mathscr{S}_{m,w,u}\right\}$$

is nonempty for some $q_{it-1} \in \{m_{it-1}, k_{it-1}, l_{it-1}, z_{it-1}^s, z_{it-1}^d, u_{it-1}, z_{it-1}^h\}$, where $Proj_v(\mathscr{M} \times \mathscr{V}) := \left\{v \in \mathbb{R}^{d_v} : \exists m \text{ such that } (m, v) \in \mathscr{M} \times \mathscr{V}\right\}$.

We can relax Assumption 6(b) by allowing z_{it}^h and l_{it} to correlate with η_{it} as discussed in Appendix F. 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.

Assumption 6(c) relaxes the "full support" condition (i.e., that the support is a Cartesian product of intervals) typically required in identification proofs of this type (e.g., Chiappori et al., 2015). By requiring only that the support $\mathscr{M} \times \mathscr{V}$ be an open, connected set, we accommodate data structures where inputs are highly persistent (e.g., $k_{it} \approx k_{it-1}$), which often results in "diagonal" support shapes. As long as the support is connected, identification is achieved via line integration as detailed in the proof.

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

Assumption 6(f) can be interpreted as a generalized rank condition. Suppose $g_{\eta_t}(\eta_{it}) > 0$ for all $\eta_{it} \in \mathbb{R}$. Then, as will be shown below in (29), Assumption 6(f) holds if either

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

holds for some vector of lagged variables $\left(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)$ in the projection $\mathrm{Proj}_v(\mathscr{M} \times \mathscr{V})$ and some instrument $q_{it-1} \in \{m_{it-1}, k_{it-1}, l_{it-1}, z^s_{it-1}, z^d_{it-1}, u_{it-1}, z^h_{it-1}\}$. The latter condition is equivalent to (i) ω_{it-1} has a causal impact on ω_{it} ($\partial h / \partial \omega_{it-1} \neq 0$) and (ii) 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} at some point in the support.

Proposition 2 shows that the control function is identified from the distribution of (m_{it}, v_{it}). The identification of TFP also follows from Proposition 2 as $\omega_{it} = \mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it})$.

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

Proof. The proof of Proposition 2 follows the identification strategy of Chiappori et al. (2015), adapted to the connected support assumption.

In view of equation (27), the conditional distribution of m_{it} given v_{it} satisfies

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

where the 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

$$\frac{\partial G_{m_t|v_t}(m_{it}|v_{it})}{\partial q_{it}} = \frac{\partial \mathbb{M}^{-1}_t(m_{it}, w_{it}, u_{it})}{\partial q_{it}} g_{\eta_t}(\eta_{it}), \tag{28}$$
$$\frac{\partial G_{m_t|v_t}(m_{it}|v_{it})}{\partial q_{it-1}} = -\frac{\partial \bar{h}_t\left(m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}\right)}{\partial q_{it-1}} g_{\eta_t}(\eta_{it}), \tag{29}$$

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} 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 both sides of (29) are non-zero.

Dividing (28) by (29) yields:

$$\frac{\partial \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})}{\partial q_{it}} = -\frac{\partial \bar{h}_t\left(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)}{\partial q_{it-1}} \times \frac{\partial G_{m_t|v_t}\left(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)/\partial q_{it}}{\partial G_{m_t|v_t}\left(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)/\partial q_{it-1}}. \tag{30}$$

Evaluating (30) for $q_{it} = m_{it}$ and applying the normalization Assumption 4, we identify the scaling factor:

$$1 = \mathbb{M}_t^{-1}(m^*_{t1}, w^*_t, u^*_t) - \mathbb{M}_t^{-1}(m^*_{t0}, w^*_t, u^*_t) = -\frac{1}{S_{q_{t-1}}} \frac{\partial \bar{h}_t\left(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)}{\partial q_{t-1}},$$

where

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

Thus, we identify $\partial \bar{h}_t / \partial q_{it-1} = -S_{q_{t-1}}$. Substituting this back into (30), the partial derivative for any argument q_{it} is identified as:

$$\frac{\partial \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})}{\partial q_{it}} = S_{q_{t-1}} \frac{\partial G_{m_t|v_t}\left(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \dots\right)/\partial q_t}{\partial G_{m_t|v_t}\left(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \dots\right)/\partial q_{t-1}}. \tag{31}$$

To recover the level of the function $\mathbb{M}_t^{-1}(\cdot)$, we rely on the Connected Support Assumption 6(c). Let $\mathbf{x}_0 := (m^*_{t0}, k^*_t, l^*_t, z^{s*}_t, z^{d*}_t, u^*_t)$ be the normalization point where $\mathbb{M}_t^{-1}(\mathbf{x}_0) = 0$. For any target point $\mathbf{x} := (m_t, k_t, l_t, z^s_t, z^d_t, u_t)$ in the interior of the support int($\mathscr{S}_{m,w,u}$), there exists a piecewise smooth path $\gamma : [0,1] \to$ int($\mathscr{S}_{m,w,u}$) such that $\gamma(0) = \mathbf{x}_0$ and $\gamma(1) = \mathbf{x}$, lying entirely within the support.

By the Fundamental Theorem of Line Integrals (Stewart, 2012, p. 1075), we identify $\mathbb{M}_t^{-1}(\mathbf{x})$ as:

$$\begin{aligned}\mathbb{M}_t^{-1}(\mathbf{x}) &= \mathbb{M}_t^{-1}(\mathbf{x}_0) + \int_\gamma \nabla \mathbb{M}_t^{-1}(\mathbf{z}) \cdot d\mathbf{z} \\ &= 0 + \int_0^1 \left[\sum_{q \in \{m,k,l,z^s,z^d,u\}} \frac{\partial \mathbb{M}_t^{-1}(\gamma(\tau))}{\partial q_{it}} \frac{d\gamma_q(\tau)}{d\tau}\right] d\tau, \end{aligned} \tag{32}$$

where $\gamma_q(\tau)$ denotes the q-th component of $\gamma(\tau)$. Since the vector field $\nabla \mathbb{M}_t^{-1}$ is conservative, the integral is path-independent. This ensures uniquely identified values for $\mathbb{M}_t^{-1}(\cdot)$ regardless of the specific path chosen, provided the path remains within the connected support of the data.

Finally, from $\omega_{it} = \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})$, we identify $\bar{h}_t(\cdot) = E[\omega_{it}|\cdot]$ and the residual η_{it}, allowing for the identification of the distribution $G_{\eta_t}(\cdot)$. $\square$

Remark 1 (Support Requirements and Persistence). Assumption 6(f) and the integration in (32) require the support of observables to be *connected*, relaxing the "full support" condition discussed in Chiappori et al. (2015). While inputs like capital are persistent and have bounded conditional support, identification remains valid if the union of these supports forms a connected set. Local variation identifies partial derivatives pointwise; global identification is then achieved by integrating ("stitching") these derivatives along any path connecting the normalization point to the point of interest. Thus, high persistence does not preclude identification provided there are no isolated "islands" of data.

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

The final step identifies the production function, markup, and demand function. From $\phi_t(m_{it},w_{it},u_{it}) = \varphi_t\left(f_t(x_{it}, z_{it}^s) + \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}), z_{it}^d, u_{it}\right)$ and the monotonicity of φ_t, 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:

$$\frac{\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})}{\partial r_{it}} \frac{\partial \phi_t(m_{it}, w_{it}, u_{it})}{\partial q_{it}^s} = \frac{\partial f_t(x_{it}, z_{it}^s)}{\partial q_{it}^s} + \frac{\partial \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})}{\partial q_{it}^s}, \tag{33}$$

$$\frac{\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})}{\partial r_{it}} \frac{\partial \phi_t(m_{it}, w_{it}, u_{it})}{\partial q_{it}^d} = \frac{\partial \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})}{\partial q_{it}^d} - \frac{\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})}{\partial q_{it}^d}. \tag{34}$$

Note that $\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})/\partial r_{it} = \left(\partial \varphi_t(y_{it}, z_{it}^d, u_{it})/\partial y_{it}\right)^{-1}$ represents the markup from (8). If the markup $\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})/\partial r_{it}$ were known, then equations (33) and (34) could identify $\partial f_t(x_{it}, z_{it}^s)/\partial q_{it}^s$ and $\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})/\partial q_{it}^d$ 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 Doraszelski and Jaumandreu (2013, 2018) and 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 (7)*

$$\frac{\partial f_t(x_{it}, z_{it}^s)}{\partial m_{it}} = \frac{\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})}{\partial r_{it}} \frac{\exp(p_t^m + m_{it})}{\exp(r_{it})} \tag{35}$$

holds for all firms.

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

$$\frac{\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})}{\partial r_{it}} = \frac{\partial f_t(x_{it}, z_{it}^s)/\partial m_{it}}{\exp(p_t^m + m_{it})/\exp(r_{it})}. \tag{36}$$

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 $\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})/\partial r_{it}$ up to scale.*

Proof. From (33) and (35), the markup $\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})/\partial r_{it}$ is identified for each firm as

$$\frac{\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})}{\partial r_{it}} = \frac{\partial \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})}{\partial m_{it}} \left(\frac{\partial \phi_t(m_{it}, w_{it}, u_{it})}{\partial m_{it}} - \frac{\exp(p_t^m + m_{it})}{\exp(r_{it})} \right)^{-1}. \tag{37}$$

From ϕ_t and (37), the markup is expressed as a function of (m_{it}, w_{it}, u_{it}) as

$$\mu_t(m_{it}, w_{it}, u_{it}) := \frac{\partial \varphi_t^{-1}(\phi_t(m_{it}, w_{it}, u_{it}), z_{it}^d, u_{it})}{\partial r_{it}}, \tag{38}$$

which we can identify from data and the identified markups. Substituting (38) into (33), we identify $\partial f_t(x_{it}, z_{it}^s)/\partial q_{it}$ for $q_{it}^s \in \{m_{it}, k_{it}, l_{it}, z_{it}^s\}$ as follows:

$$\frac{\partial f_t(x_{it}, z_{it}^s)}{\partial q_{it}} = \mu_t(m_{it}, w_{it}, u_{it}) \frac{\partial \phi_t(m_{it}, w_{it}, u_{it})}{\partial q_{it}} - \frac{\partial \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})}{\partial q_{it}}. \tag{39}$$

Consequently, the gradient field ∇f_t is identified everywhere on the support.

To recover the production function level $f_t(x_t, z_t^s)$ from these identified derivatives, we apply the connected support condition. Let $\mathbf{v}_{it} := (m_{it}, k_{it}, l_{it}, z_{it}^s)$ denote the vector of production function arguments. Let $\mathbf{v}^* := (m_{t0}^*, k_t^*, l_t^*, z_t^{s*})$ be the normalization point where $f_t(\mathbf{v}^*) = 0$ (Assumption 4). For any point $\mathbf{v}_{it}$ in the connected support, there exists a path γ connecting $\mathbf{v}^*$ to $\mathbf{v}_{it}$. By the Fundamental Theorem of Line Integrals, we identify f_t as:

$$\begin{aligned} f_t(\mathbf{v}_{it}) &= f_t(\mathbf{v}^*) + \int_\gamma \nabla f_t(\mathbf{z}) \cdot d\mathbf{z} \\ &= 0 + \int_0^1 \left[\sum_{q \in \{m,k,l,z^s\}} \frac{\partial f_t(\gamma(\tau))}{\partial q_{it}} \frac{d\gamma_q(\tau)}{d\tau} \right] d\tau. \end{aligned} \tag{40}$$

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

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

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 2. 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 (31). Furthermore, the proof of Proposition 3 does not rely on the restriction in (34) 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.

2.3.4 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 two 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.,Olley and Pakes, 1996; Levinsohn and Petrin, 2003; 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 $y_{it} < y^*_{it}$.[17] 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 (35) in their first step: $\ln \frac{\partial f_t(x_{it}, z^s_{it})}{\partial m_{it}} = \ln \frac{\exp(p^m_t + m_{it})}{\exp(r_{it})} + \ln \frac{\partial \varphi_t^{-1}(r_{it}, z^d_{it}, u_{it})}{\partial r_{it}}$, assuming perfect competition or identical constant markups where $\partial \varphi_t^{-1}(r_{it}, z^d_{it}, u_{it})/\partial r_{it}$ becomes a common constant for all i. Under imperfect competition with variable markups, when the markup depends on revenue r_{it}, $\partial f_t(x_{it}, z^s_{it})/\partial m_{it}$ cannot be identified solely from the first-order condition.

2.4 Extensions to Alternative Settings

We demonstrate in the Appendix that our identification strategy extends to settings where the baseline assumptions are relaxed. First, we establish identification with endogenous labor inputs in Appendix F.1 by incorporating labor adjustment costs, which allow lagged labor to serve as a valid instrument. Second, we address endogenous firm characteristics in Appendix F.2 by employing a control function approach for triangular systems. Third, Appendix F.3 confirms that our identification results—particularly the instrumental variable quantile regression in Step 1—remain valid when firm characteristics are discrete rather than continuous.

We also address the restrictiveness of assuming limited persistence in demand shocks through two extensions. First, Appendix F.4 demonstrates that identification holds under persistent shocks (e.g., AR(1)) if a lagged supply shifter, such as supply-driven R&D (z^h_{it-1}), serves as the instrument. Because z^h_{it-1} shifts productivity but is excluded from demand, it satisfies the completeness condition despite serial correlation. Second, Appendix F.5 explicitly models persistent unobserved quality. This captures the persistent component of demand heterogeneity, allowing ϵ_{it} to represent only the remaining high-frequency, transitory fluctuations.

2.5 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

[17] 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

$$\varphi_t^{-1}(r_{it}, z_{it}^d, u_{it}) = a_{1t} + a_{2t} + b_t \varphi_t^{*-1}(r_{it}, z_{it}^d, u_{it}),\ f_t(x_{it}, z_{it}^s) = a_{1t} + b_t f_t^*(x_{it}, z_{it}^s),$$
$$\mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}) = a_{2t} + b_t \mathbb{M}_t^{*-1}(m_{it}, w_{it}, u_{it}). \tag{41}$$

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.[18]

2.5.1 Scale Normalization

From (41), 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:

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

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

[18] 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.

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 $t+1$.*

The proof is given in Appendix B.2.

2.5.2 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 $z_{it}^s = z_t^s$ 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 $f_t(x_t, z_t^s)$ be the identified production function under Assumption 4 and $f_t^*(x_t)$ be the true one where $f_t(x_t, z_t^s) = a_t + b_t f_t^*(x_t, z_t^s)$ from (41). 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 (41), 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.*

2.5.3 Location Normalization

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

$$\tilde{\varphi}_t^{-1}(r_{it}, z_{it}^d, u_{it}) := \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})/b_t,\ \tilde{f}_t(x_{it}, z_{it}^s) := f_t(x_{it}, z_{it}^s)/b_t,\ \tilde{\omega}_{it} := \omega_{it}/b_t,$$
$$\tilde{a}_{1t} := a_{1t}/b_t,\ \text{and } \tilde{a}_{2t} := a_{2t}/b_t. \tag{42}$$

Then, (41) is written as

$$\tilde{\varphi}_t^{-1}(r_{it}, z_{it}^d, u_{it}) = \tilde{a}_{1t} + \tilde{a}_{2t} + \varphi_t^{*-1}(r_{it}, z_{it}^d, u_{it}),\ \tilde{f}_t(x_{it}, z_{it}^s) = \tilde{a}_{1t} + f_t^*(x_{it}, z_{it}^s),\ \tilde{\omega}_{it} = \tilde{a}_{2t} + \omega_{it}^*. \tag{43}$$

From (41), the growth rates (log differences) of the identified output and TFP between t and $t+1$ are related to their true values as follows:

$$\begin{aligned}\tilde{\varphi}_{t+1}^{-1}(r_{it+1}, z_{it+1}^d, u_{it+1}) - \tilde{\varphi}_t^{-1}(r_{it}, z_{it}^d, u_{it}) &= \tilde{a}_{1t+1} + \tilde{a}_{2t+1} - \tilde{a}_{1t} - \tilde{a}_{2t} \\ &\quad + \varphi_{t+1}^{*-1}(r_{it+1}, z_{it+1}^d, u_{it+1}) - \varphi_t^{*-1}(r_{it}, z_{it}^d, u_{it}), \\ \tilde{f}_{t+1}(x_{it+1}, z_{it+1}^s) - \tilde{f}_t(x_{it}, z_{it}^s) &= \tilde{a}_{1t+1} - \tilde{a}_{1t} + f_t^*(x_{it+1}, z_{it+1}^s) - f_t^*(x_{it}, z_{it}^s), \\ \tilde{\omega}_{it+1} - \tilde{\omega}_{it} &= \tilde{a}_{2t+1} - \tilde{a}_{2t} + \omega_{it+1}^* - \omega_{it}^*.\end{aligned} \tag{44}$$

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

$$P_t^* := \frac{\sum_{i \in \tilde{N}} \exp(p_{it}^* + y_{i0}^*)}{\sum_{i \in \tilde{N}} \exp(p_{i0}^* + y_{i0}^*)}, \tag{45}$$

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 $(\bar{x}, \bar{z}^s) \in \mathscr{X} \times \mathscr{Z}_s$ 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.*

The proof is given in Appendix B.3.

2.6 Identification of HSA Demand System, Utility Function, and Counterfactual Welfare Effects

Given that we have identified each firm's output price and quantity, it is possible to nonparametrically identify a Homothetic Single Aggregator (HSA) system of demand functions and the associated utility function of a representative consumer under an additional homotheticity shape restriction. Furthermore, we may identify the welfare effects of counterfactual experiments. Matsuyama and Ushchev (2017) further propose two additional families of homothetic demand systems—the homothetic demand system with direct implicit additivity (HDIA) and that with indirect implicit additivity (HIIA). In the Appendix E, we show that the analysis in this subsection extends to the HDIA and HIIA cases. Notably, the HDIA family nests the Kimball aggregator, which is widely used in Macroeconomics, as a special case.

2.6.1 Identification of the HSA demand system and consumer preference

The HSA demand system We adopt the Homothetic Single Aggregator (HSA) demand system proposed by Matsuyama and Ushchev (2017).[19] Let N_t denote the number of firms in the industry. The HSA demand system is characterized by the *structural* budget-share function $\mathfrak{s}_t^*(\cdot, z_{it}^d, u_{it})$, which returns the log market revenue share of product i as a function of its log relative output quantity, defined by $y_{it} - q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$. This relationship is expressed as:

$$\ln\left(\frac{\exp(r_{it})}{\sum_{j=1}^{N_t} \exp\left(r_{jt}\right)}\right) = \mathfrak{s}_t^*\left(y_{it} - q_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right) \quad \text{for } i = 1, ..., N_t, \tag{46}$$

where $\mathbf{y}_t := (y_{1t}, ..., y_{N_t t}) \in \mathscr{Y}^{N_t}$ is a vector of consumption, $\mathbf{z}_t^d := (z_{1t}^d, ..., z_{N_t t}^d)$ is a vector of observable demand shifters, $\mathbf{u}_t := (u_{1t}, ..., u_{N_t t})$ is a vector of demand shocks, and $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$

[19] The HSA system can be expressed as a system of direct demand functions or of inverse demand functions. These two systems are self-dual, meaning either can be derived from the other. Matsuyama (2023) and Matsuyama (2025) provide comprehensive reviews of flexible extensions to the CES demand system, including the HSA framework.

is the aggregate quantity index summarizing interactions across products.[20] In equilibrium, the quantity index $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ is uniquely determined by the adding-up constraint on market shares:

$$1 = \sum_{i=1}^{N_t} \exp\left(\mathfrak{s}_t^*\left(y_{it} - q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t), z_{it}^d, u_{it}\right)\right). \tag{47}$$

Because $\mathfrak{s}_t^*(\cdot)$ is nonparametric, the HSA framework nests various demand systems established in the literature, including the CES and symmetric translog demand systems (Feenstra, 2003; Feenstra and Weinstein, 2017).[21]

Identification of the HSA demand system To identify the demand system, we maintain the following assumptions. Let Φ_t denote the log of aggregate expenditure for a representative consumer, where $\Phi_t = \ln\left(\sum_{i=1}^{N_t} \exp(r_{it})\right)$ in equilibrium.

Assumption 11. *(a) The observed data are generated from the equilibrium corresponding to the baseline aggregate state* $\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)$ *under the HSA demand system. (b) The product market is characterized by monopolistic competition (without free entry); specifically, each firm takes the quantity index* $q_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)$ *and the log of aggregate expenditure* Φ_t *as given. (c) The quantity index at the baseline aggregate state is normalized as* $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t) = 0$*. (d)* $\varphi_t(\cdot)$ *is known over its support.*

Assumption 11(a) explicitly specifies the baseline aggregate state $\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)$ under which the *reduced-form* functions are defined. Assumption 11(b) imposes the standard framework of monopolistic competition (cf. Klette and Griliches, 1996; De Loecker, 2011), implying that N_t is sufficiently large such that single-firm strategic decisions have a negligible impact on aggregate quantities. Assumption 11(c) provides the normalization condition required to uniquely determine the *structural* function $\mathfrak{s}_t^*(\cdot)$, as discussed below. Assumption 11(d) holds when the conditions of Proposition 5 hold.

Treating $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ and Φ_t as fixed, we define the *reduced-form* revenue function (48) from the structural budget-share function:

$$r_{it} = \Phi_t + \mathfrak{s}_t^*\left(y_{it} - q_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right) =: \varphi_t\left(y_{it}, z_{it}^d, u_{it}\right). \tag{48}$$

[20]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 $p_{it} = \rho\left(y_{it} - \ln U(\mathbf{y}_t)\right) + \Phi_t - y_{it}$. Here, $\ln U(\mathbf{y}_t) = q_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)$, meaning the quantity index coincides with the utility function, though they generally differ in the HSA framework.

[21]See Matsuyama and Ushchev (2020) for details on how the HSA nests translog demand.

Correspondingly, the *reduced-form* inverse demand function is related to $\mathfrak{s}_t^*(\cdot)$ by:

$$p_{it} = \Phi_t + \mathfrak{s}_t^*\left(y_{it} - q_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right) - y_{it} =: \psi_t\left(y_{it}, z_{it}^d, u_{it}\right).$$

In this context, the function

$$\mathfrak{s}_t\left(y_{it}, z_{it}^d, u_{it}\right) := \varphi_t\left(y_{it}, z_{it}^d, u_{it}\right) - \Phi_t \tag{49}$$

represents the *reduced-form* budget-share function, which is known conditional on Φ_t and $\varphi_t(\cdot)$.

Conditional on the aggregate state $(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$, the *reduced-form* revenue $\varphi_t(\cdot)$ and budget-share function $\mathfrak{s}_t(\cdot)$ are identified up to location normalization via Proposition 5. When evaluating these reduced-form functions (48) and (49) across different values of $(y_{it}, z_{it}^d, u_{it})$, the quantity index $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ is held constant at the baseline state $(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$. Consequently, these reduced-form functions are not invariant to changes in the aggregate state. As indicated by (48), $\varphi_t(\cdot)$ shifts in response to changes in $\mathbf{y}_t$ because the quantity index $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ depends on the consumption vector $\mathbf{y}_t$.

Nonetheless, we may identify the structural function $\mathfrak{s}_t^*(\cdot)$ from the reduced-form function $\mathfrak{s}_t(\cdot)$ in (49) under the equilibrium constraint (47) as follows. Evaluating (46) and (49) at the fixed baseline state $\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)$, the *reduced-form* and *structural* budget share functions are related as:

$$\mathfrak{s}_t\left(y_{it}, z_{it}^d, u_{it}\right) = \mathfrak{s}_t^*\left(y_{it} - q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t), z_{it}^d, u_{it}\right). \tag{50}$$

Let $\Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) := q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) - q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ denote the change in the quantity index induced by a shift from the baseline state $(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ to a new state $(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)$. Evaluating (50) at the argument $(y_{it}, z_{it}^d, u_{it}) = (\tilde{y}_{it} - \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t), \tilde{z}_{it}^d, \tilde{u}_{it})$ while keeping $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ as constant, we have:

$$\mathfrak{s}_t\left(\tilde{y}_{it} - \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t), \tilde{z}_{it}^d, \tilde{u}_{it}\right) = \mathfrak{s}_t^*\left(\tilde{y}_{it} - q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t), \tilde{z}_{it}^d, \tilde{u}_{it}\right), \tag{51}$$

for any alternative state $(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)$. By substituting (51) into the constraint (47), we can uniquely identify $\Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)$ by solving:

$$1 = \sum_{i=1}^{N_t} \exp\left(\mathfrak{s}_t\left(\tilde{y}_{it} - \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t), \tilde{z}_{it}^d, \tilde{u}_{it}\right)\right). \tag{52}$$

Given the identified quantity index change $\Delta q_t(\cdot)$, we are able to identify the *structural* budget share function $\mathfrak{s}_t^*(\cdot)$ from the *reduced-form* function $\mathfrak{s}_t(\cdot)$ via (51), up to the unknown value of the baseline index $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$.

The preceding discussion demonstrates that $\mathfrak{s}_t^*(\cdot)$ cannot be separately identified from the baseline aggregate quantity index $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$. Specifically, given the identified reduced-form function $\mathfrak{s}_t(\cdot)$ and the change in the index $\Delta q_t(\cdot)$, there exist multiple pairs $\{\mathfrak{s}_t^*(\cdot), q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)\}$ that satisfy (51). For this reason, we impose Assumption 11(c) to uniquely determine $\mathfrak{s}_t^*(\cdot)$ but the fundamental properties of the preference structure to be characterized in Proposition 7 are invariant to this specific normalization.

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 representative consumer preference, and that it is possible to identify an associated utility function. Appendix B.5 supplies the proof.

Proposition 7. *Suppose that the conditions of Proposition 5 and Assumption 11 hold. Then: (a) There exists a unique monotone, convex, and homothetic rational preference $\succsim$ over $\mathcal{Y}^{N_t}$ that generates an HSA demand system $\left\{\mathfrak{s}_t^*(\cdot), q_t(\cdot)\right\}$. (b) This preference $\succsim$ is represented by a homothetic utility function defined by*

$$\ln U_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) = \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) + \sum_{i=1}^{N_t} \int_{b_i}^{\tilde{y}_{it} - \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)} \exp\left(\mathfrak{s}_t\left(\zeta, \tilde{z}_{it}^d, \tilde{u}_{it}\right)\right) d\zeta$$

for some constant vector $\mathbf{b} = (b_1, ..., b_{N_t})$, which is identified from $\{\mathfrak{s}_t(\cdot), \Delta q_t(\cdot)\}$. (c) The identified demand system and preference $\succsim$ do not depend on the location normalization of $\varphi_t^{-1}(r_t, z_t^d, u_t)$.

2.6.2 Counterfactual analysis

We conduct a short-run partial equilibrium counterfactual analysis by maintaining the following assumptions.

Assumption 12. *(a) The pre-determined factor inputs (k_{it}, l_{it}), factor prices, and exogenous variables $(z_{it}^d, z_{it}^s, u_{it}, \omega_{it}, p_t^m)$ are fixed before and after counterfactual interventions. (b) The log of aggregate expenditure Φ_t is fixed before and after counterfactual interventions.*

Under Assumption 12(a), firms respond to counterfactual interventions by adjusting their material inputs, outputs, and prices in the short run. Assumption 12(b) implies that the aggregate income of the representative consumer remains invariant to the counterfactual interventions. This is a reasonable approximation when analyzing a specific industry that is small relative to the aggregate economy. For large-scale interventions, however, this assumption can be relaxed to allow aggregate income to adjust endogenously in response to changes in aggregate profits induced by the intervention.

Monopolistic Competition Equilibrium Using the identified HSA demand system $\{\mathfrak{s}_t(\cdot), \Delta q_t(\cdot)\}$, we calculate a monopolistic competition equilibrium (MCE) as follows. Define the inverse production function $m_{it} = \chi_{it}(y_{it})$ such that $y_{it} = f_t\left(\chi_{it}(y_{it}), k_{it}, l_{it}, z^s_{it}\right) + \omega_{it}$ for given $(k_{it}, l_{it}, \omega_{it}, z^s_{it})$; namely, $\chi_{it}(y_{it}) := f_t^{-1}\left(y_{it} - \omega_{it}, k_{it}, l_{it}, z^s_{it}\right)$.

Equilibrium outputs and the quantity index $(\mathbf{y}^m_t, \Delta q^m_t)$ in an MCE are obtained from the first order condition for profit maximization in (7), equivalently written in view of (49)-(51), and the market share condition (52):

$$\underbrace{\exp\left(\mathfrak{s}_t\left(y^m_{it} - \Delta q^m_t, z^d_{it}, u_{it}\right) + \Phi_t - y^m_{it}\right) \frac{\partial \mathfrak{s}_t\left(y^m_{it} - \Delta q^m_t, z^d_{it}, u_{it}\right)}{\partial y_{it}}}_{\text{Marginal Revenue}} = \underbrace{\frac{\exp\left(p^m_t + \chi_{it}\left(y^m_{it}\right)\right)}{\exp\left(y^m_{it}\right)} \frac{\partial \chi_{it}(y^m_{it})}{\partial y_{it}}}_{\text{Marginal Cost}},$$

$$\sum_{i=1}^{N_t} \exp\left(\mathfrak{s}_t\left(y^m_{it} - \Delta q^m_t, z^d_{it}, u_{it}\right)\right) = 1, \tag{53}$$

for $i = 1, ..., N_t$, where $\frac{\partial \chi_{it}(y^m_{it})}{\partial y_{it}} = \left(\frac{\partial f_t}{\partial m}\right)^{-1}$. The above system can be extended to incorporate policies such as taxes and subsidies to investigate their effects.

Welfare Costs of Firm's Market Power In the 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), in which firms set their prices equal to their marginal costs. Specifically, equilibrium outputs and quantity index $(\mathbf{y}^c_t, \Delta q^c_t)$ are obtained by solving:

$$\underbrace{\exp\left(\mathfrak{s}_t\left(y^c_{it} - \Delta q^c_t, z^d_{it}, u_{it}\right) + \Phi_t - y^c_{it}\right)}_{\text{Price}} = \underbrace{\frac{\exp\left(p^m_t + \chi_{it}\left(y^c_{it}\right)\right)}{\exp\left(y^c_{it}\right)} \frac{\partial \chi_{it}(y^c_{it})}{\partial y_{it}}}_{\text{Marginal Cost}},$$

$$\sum_{i=1}^{N_t} \exp\left(\mathfrak{s}_t\left(y^c_{it} - \Delta q^c_t, z^d_{it}, u_{it}\right)\right) = 1, \tag{54}$$

for $i = 1, ..., N_t$.

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

$$\ln U_t(\mathbf{y}^c_t, \mathbf{z}^d_t, \mathbf{u}_t) - \ln U_t(\mathbf{y}^m_t, \mathbf{z}^d_t, \mathbf{u}_t) = \Delta q^c_t - \Delta q^m_t + \sum_{i=1}^{N_t} \int_{y^m_{it} - \Delta q^m_t}^{y^c_{it} - \Delta q^c_t} \exp\left(\mathfrak{s}_t\left(\zeta, z^d_{it}, u_{it}\right)\right) d\zeta.$$

An alternative welfare measure is the Compensating Variation (CV), which is constructed in monetary terms as follows. For a given counterfactual log income Φ^c_t, define the counterfactual output and quantity index $(\mathbf{y}_t^{c*}(\Phi_t^c), \Delta q^{c*}(\Phi_t^c))$ that solve the price-taking condition:

$$\underbrace{\exp\left(\mathfrak{s}_t\left(y_{it}^{c*} - \Delta q_t^{c*}, z_{it}^d, u_{it}\right) + \Phi_t^c - y_{it}^{c*}\right)}_{\text{Price}} = \underbrace{\frac{\exp\left(p_t^m + \chi_{it}\left(y_{it}^{c*}\right)\right)}{\exp(y_{it}^{c*})} \frac{\partial \chi_{it}(y_{it}^{c*})}{\partial y_{it}}}_{\text{Marginal Cost}}$$

for $i = 1, ..., N_t$ with $\sum_{i=1}^{N_t} \exp\left(\mathfrak{s}_t\left(y_{it}^{c*} - \Delta q_t^{c*}, z_{it}^d, u_{it}\right)\right) = 1$. We then find the counterfactual income Φ_t^{c*} that achieves the same utility as in the benchmark MCE:

$$\ln U_t(\mathbf{y}_t^{c*}(\Phi_t^{c*}), \mathbf{z}_t^d, \mathbf{u}_t) - \ln U_t(\mathbf{y}_t^m, \mathbf{z}_t^d, \mathbf{u}_t) = 0. \tag{55}$$

The compensating variation is defined as $CV_t := \exp\left(\Phi_t^{c*}\right) - \exp\left(\Phi_t\right)$. It measures the change in consumer welfare when moving from an MCPE to an MCE and quantifies the consumer's loss due to firms' market power. Since consumers require less income under competitive pricing to achieve the same utility, $CV_t < 0$: the magnitude $|CV_t|$ is the monetary gain from eliminating market power, and the consumer's welfare improvement from transitioning to an MCPE is $-CV_t > 0$.

To evaluate the overall welfare change from an MCE to an MCPE, the consumer gain, measured by $-CV_t$, can be compared with firms' profit loss. Under Assumption 12(b), aggregate revenue is fixed. Thus, the change in total profits is solely the negative change in total material costs:

$$\Delta\Pi_t := \Pi_t^c - \Pi_t^m = -\sum_{i=1}^{N_t} \exp(p_t^m) \left\{\exp\left(\chi_{it}\left(y_{it}^c\right)\right) - \exp\left(\chi_{it}\left(y_{it}^m\right)\right)\right\}. \tag{56}$$

Note that if $\Pi_t^c < 0$, this counterfactual implies firms operate at a loss, representing a short-run equilibrium before firm exit occurs. The overall welfare change associated with a transition from an MCE to an MCPE is therefore given by $\Delta\Pi_t - CV_t$.

3 Semiparametric Estimator

The identification results in Section 2 are nonparametric: revenue data alone suffice to recover the production function, demand system, and welfare-relevant objects without functional-form restrictions. While a fully nonparametric sieve estimator could in principle be constructed from the identified equations, it would demand sample sizes far exceeding typical manufacturing datasets. We therefore impose Cobb-Douglas production and AR(1) productivity dynamics, following the standard practice of pairing nonparametric identification with parametric specifications (see, e.g., Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020).

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

$$f_t(m_{it}, k_{it}, l_{it}) = \theta_m m_{it} + \theta_k k_{it} + \theta_l l_{it}, \tag{57}$$

and TFP follows an AR(1) process:

$$\omega_{it} = \rho \omega_{it-1} + \eta_{it}. \tag{58}$$

We introduce assumptions on the production function to normalize scale and location parameters. First, (57) normalizes one location parameter by $f_t(0,0,0) = 0$. As another location normalization, we assume $E[\omega_{it}] = 0$, which naturally arises from a stationary AR1 process (58). Finally, we assume the constant returns to scale, $\theta_m + \theta_k + \theta_l = 1$, which normalizes the scale parameter.

The control function becomes separable under the Cobb-Douglas assumption:

$$\mathbb{M}_t^{-1}(m_{it}, k_{it}, l_{it}, u_{it}) = \lambda_t(m_{it}, u_{it}) - \theta_k k_{it} - \theta_l l_{it}. \tag{59}$$

Substituting (59) into the revenue function, we obtain

$$\varphi_t \left(f_t(m_{it}, k_{it}, l_{it}) + \omega_{it}, u_{it}\right) = \varphi_t \left(\theta_m m_{it} + \lambda_t(m_{it}, u_{it}), u_{it}\right) = \phi_t(m_{it}, u_{it}), \tag{60}$$

where ϕ_t depends only on (m_t, u_{it}) and increases in m_t and u_t. In Appendix B.6, we derive a separable control function similar to (59) in a more general setting and show that (60) holds when the production function is separable with respect to m_{it}.

Substituting (59) into (58), we obtain the transformation model as

$$\lambda_t(m_{it}, u_{it}) = \theta_k(k_{it} - \rho k_{it-1}) + \theta_l(l_{it} - \rho l_{it-1}) + \rho \lambda_{t-1}(m_{it-1}, u_{it-1}) + \eta_{it}. \tag{61}$$

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 (26) 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 (26) with a smooth kernel CDF to ease computation.

We partition $[0,1]$ into $L = 100$ equal parts and let $\mathbb{T} = \{0.01, \ldots, 0.99\}$. The moment condition is

$$E\left[\left(\mathcal{K}_1\left(\frac{B_\phi(m_{it},\tau)^T\alpha_t - r_{it}}{b_{n1}}\right) - \tau\right) B_{IV}(m_{it-2})\right] = 0 \quad \text{for } \tau \in \mathbb{T}, \tag{62}$$

where $\mathcal{K}_1(\cdot)$ is a smooth kernel CDF with bandwidth b_{n1} and $B_{IV}(m_{it-2})$ is the B-spline basis of degree $S_1 = 3$ with $K_1 = 2$ 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):

$$\mathcal{K}_1(s) := \left[\frac{1}{2} + \frac{105}{64}\left(s - \frac{5}{3}s^3 + \frac{7}{5}s^5 - \frac{3}{7}s^7\right)\right] 1\{s \in [-1,1]\} + 1\{s > 1\}.$$

The number of moment conditions (62) 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 (62) 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 (61). We use the Profile Likelihood (PL) estimator developed by Linton, Sperlich, and Van Keilegom (2008). From (28) for $q_{it} = m_{it}$ and (59), 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}. Let $\partial_m B_\lambda(m_{it}, u_{it})$ be the derivatives of the B-spline bases $B_\lambda(m_{it}, u_{it})$ so that $\partial_m B_\lambda(m_{it}, u_{it})^T c_t$ approximates $\frac{\partial \lambda_t(m_{it}, u_{it})}{\partial m_{it}}$. Thus, the log-likelihood function is written as

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

where $g_{\eta_t}(\eta)$ is the corresponding (Gaussian) kernel density with the bandwidth chosen by Silverman's Rule. We obtain estimates of η_{it} as follows. For given (c_t, ρ), define $\lambda_{it}(c_t) := B_\lambda(m_{it}, u_{it})^T c_t$, $\tilde{k}_{it}(\rho) := k_{it} - \rho k_{it-1}$, $\tilde{l}_{it}(\rho) := l_{it} - \rho l_{it-1}$, and $\tilde{B}_\lambda(m_{it-1}, u_{it-1}, \rho) := \rho B_\lambda(m_{it-1}, u_{it-1})$, and (61) implies that

$$\lambda_{it}(c_t) = \theta_k \tilde{k}_{it}(\rho) + \theta_l \tilde{l}_{it}(\rho) + \tilde{B}_\lambda(m_{it-1}, u_{it-1}, \rho)^T c_{t-1} + \eta_{it}. \tag{63}$$

Therefore, for each (c_t, ρ), let $\eta_{it}(c_t, \rho)$ be the residual from regressing $\lambda_{it}(c_t)$ on $\tilde{k}_{it}(\rho)$, $\tilde{l}_{it}(\rho)$, and $\tilde{B}_\lambda(m_{it-1}, u_{it-1}, \rho)$. Then, the PL estimator $(\tilde{c}_t, \tilde{\rho})$ is defined as

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

Step 3: Estimation of production function, markup, TFP, and output With estimated $(\tilde{c}_t, \tilde{\rho})$, we estimate $(\theta_k, \theta_l, c_{t-1})$ by regressing $B_\lambda(m_{it}, \hat{u}_{it})^T \tilde{c}_t$ on $\tilde{k}_{it}(\tilde{\rho})$, $\tilde{l}_{it}(\tilde{\rho})$, and $\tilde{B}_\lambda(m_{it-1}, u_{it-1}, \tilde{\rho})$:

$$(\tilde{\theta}_k, \tilde{\theta}_l, \tilde{c}_{t-1}) \in \arg\min_{\theta_k, \theta_l, c_{t-1}} \sum_{i=1}^{n} \left\{ B_\lambda(m_{it}, \hat{u}_{it})^T \tilde{c}_t - \theta_k \tilde{k}_{it}(\tilde{\rho}) - \theta_l \tilde{l}_{it}(\tilde{\rho}) - \tilde{B}_\lambda(m_{it-1}, u_{it-1}, \rho)^T c_{t-1} \right\}^2.$$

From (37) and (39), the material elasticity is identified as $\theta_m = \frac{\partial \lambda_t(m_{it}, u_{it})}{\partial m_{it}} \left(\frac{s_{it}^m}{\partial \phi_t(m_{it}, u_{it})/\partial m_{it} - s_{it}^m} \right)$, where $s_{it}^m = \exp(p_t^m + m_{it}) / \exp(r_{it})$. Following this equation, we estimate θ_m as follows:

$$\tilde{\theta}_m = \text{median} \left\{ \frac{\left(\partial_m B_\lambda(m_{it}, \hat{u}_{it})^T \tilde{c}_t \right) s_{it}^m}{\partial_m B_\phi(m_{it}, \hat{u}_{it})^T \hat{\alpha}_t - s_{it}^m} \right\},$$

where $\partial_m B_\phi(m_{it}, \hat{u}_{it})$ is the derivatives of bases $B_\phi(m_{it}, \tau)$ so that $\partial_m B_\phi(m_{it}, \hat{u}_{it})^T \hat{\alpha}_t$ estimates $\partial \phi_t(m_{it}, u_{it}) / \partial m_{it}$. Using the constant returns to scale, we obtain the normalized production parameters as $\hat{\theta}_j = \tilde{\theta}_j / \tilde{b}_t$ for $j \in \{m, k, l\}$ where $\tilde{b}_t = \tilde{\theta}_m + \tilde{\theta}_k + \tilde{\theta}_l$. Then, we estimate markups as follows:

$$\hat{\mu}_{it} = \frac{\hat{\theta}_m}{s_{it}^m}.$$

With the mean-zero restriction on ω_{it}, the location parameter $\hat{a}_{2t} = n^{-1}\sum_{i=1}^{n}\left[\tilde{\lambda}_{it} - \tilde{\theta}_k k_{it} - \tilde{\theta}_l l_{it}\right]$ is estimated. The estimated TFP, output, and price are

$$\hat{\omega}_{it} = \frac{\tilde{\lambda}_{it} - \tilde{\theta}_k k_{it} - \tilde{\theta}_l l_{it} - \hat{a}_{2t}}{\tilde{b}_t}, \quad \hat{y}_{it} = \hat{\omega}_{it} + \hat{\theta}_m m_{it} + \hat{\theta}_l l_{it} + \hat{\theta}_k k_{it}, \quad \text{and} \quad \hat{p}_{it} = r_{it} - \hat{y}_{it}.$$

Step 4: Estimation of parametric CoPaTh-HSA demand system Our estimation steps of production function above do not assume any parametric demand system. Thus, in theory, one can estimate a fully nonparametric HSA demand system as described in Section 2.6. 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):

$$\mathfrak{s}_t^*(y_{it} - q_t(\mathbf{y}_t, \epsilon_t), \epsilon_{it}) := \delta_t - \frac{1}{\beta_t}\ln\left(\frac{\exp(-\beta_t(y_{it} - q_t(\mathbf{y}_t, \epsilon_t)) + \gamma_t) + \epsilon_{it}}{1 + \epsilon_{it}}\right) \tag{64}$$

where the quantity index $q_t(\mathbf{y}_t, \epsilon_t)$ is implicitly defined by the market share constraint (47) 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 derives the log-variable version of the CoPaTh demand (64) from Matsuyama and Ushchev (2020)'s original formulation.

As explained in Section 2.6, we estimate the following reduced-form revenue function instead of the structural form (64):

$$\varphi_t(y_{it}, \epsilon_{it}) = \Phi_t + \delta_t - \frac{1}{\beta_t}\ln\left(\frac{\exp(-\beta_t y_{it} + \gamma_t) + \epsilon_{it}}{1 + \epsilon_{it}}\right), \tag{65}$$

where we impose the normalization $q_t(\mathbf{y}_t, \boldsymbol{\epsilon}_t) = 0$, as stated in Assumption 11(c), by assuming that the observed data correspond to the baseline aggregate state. Consequently, the implied markup is

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

Taking the limit as $\beta_t \to 0$ on the right-hand side of (65) yields a linear revenue function $\varphi_t(y_{it}, \epsilon_{it}) = \alpha_t + \rho_{it} y_{it}$, where $\alpha_t = \Phi_t + \delta_t$ and $\rho_{it} = \frac{1}{1+\epsilon_{it}}$. Thus, the CoPaTh demand system in (64) nests the generalized CES demand with heterogeneous markups defined in (12). Under this special case, the markup is independent of y_{it} and pass-through is complete. Moreover, if $\epsilon_{it} = \epsilon$ is common across all firms, then the system collapses to the standard CES demand framework with a homogeneous markup.

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

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

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

subject to

$$\epsilon_{it} = \frac{\hat{\mu}_{it} - 1}{\exp(\beta_t \hat{y}_{it} - \gamma_t)} \quad \text{and} \quad 1 = \sum_i \exp\left(\delta_t - \frac{1}{\beta_t} \ln\left(\frac{\exp(-\beta_t \hat{y}_{it} + \gamma_t) + \epsilon_{it}}{1+\epsilon_{it}} \right)\right),$$

where quantile(ϵ_{it}) is the empirical quantile of ϵ_{it} among all firms and $\Phi_t = \ln\left(\sum_i \exp(r_{it})\right)$ is the industry revenue. The estimation utilizes the theoretical relationship between markups and demand shocks and the market share restriction. [22]

4 Simulation

This section presents the finite sample performance of our proposed estimator, comparing to that of the conventional 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 A.

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 (64). The demand shock ϵ_{it} follows an MA(1) process: $\epsilon_{it} = 0.5\zeta_{it-1} + \zeta_{it}$, where $\zeta_{it} \sim \text{Unif}[0, 0.3]$. The production function takes the Cobb-Douglas form (57) 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.2.

The production function parameters are set to $(\theta_m, \theta_k, \theta_l) = (0.4, 0.3, 0.3)$. We choose the structural parameters of the HSA demand system, $(\Phi_t, \delta_t, \beta_t) = (20, -6.5, 0.21)$, and allow parameter γ_t to vary across simulations to satisfy the normalization $q_t(\mathbf{y}_t, \boldsymbol{\epsilon}_t) = 0$ in each simulation. Specifically, for each period, we find equilibrium outputs $\{y^m_{it}\}_{i=1}^{N_t}$ and γ_t in an MCE

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

by solving the first order conditions and the market share condition analogous to (53):

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

where $\Xi_{it} = \ln\theta_m + (\theta_k k_{it} + \theta_l l_{it} + \omega_{it})/\theta_m$ and $p_t^m = 0$. Appendix B.4 show its derivation. We simulate 100 replications of $N = 600$ firms and $T = 5$ periods, with the following summary statistics of the resulting markups:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
Markup	1.001	1.147	1.219	1.223	1.295	1.617

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

In addition to our proposed estimator, we also consider the widely-used estimator proposed by Ackerberg et al. (2015)(ACF) applied with quantity data and the ACF estimator applied with revenue data. 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., $\theta_m + \theta_k + \theta_l = 1$) that Flynn, Gandhi, and Traina (2019) proposed to address the GNR criticism.

Figure 1 shows 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 as output.

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

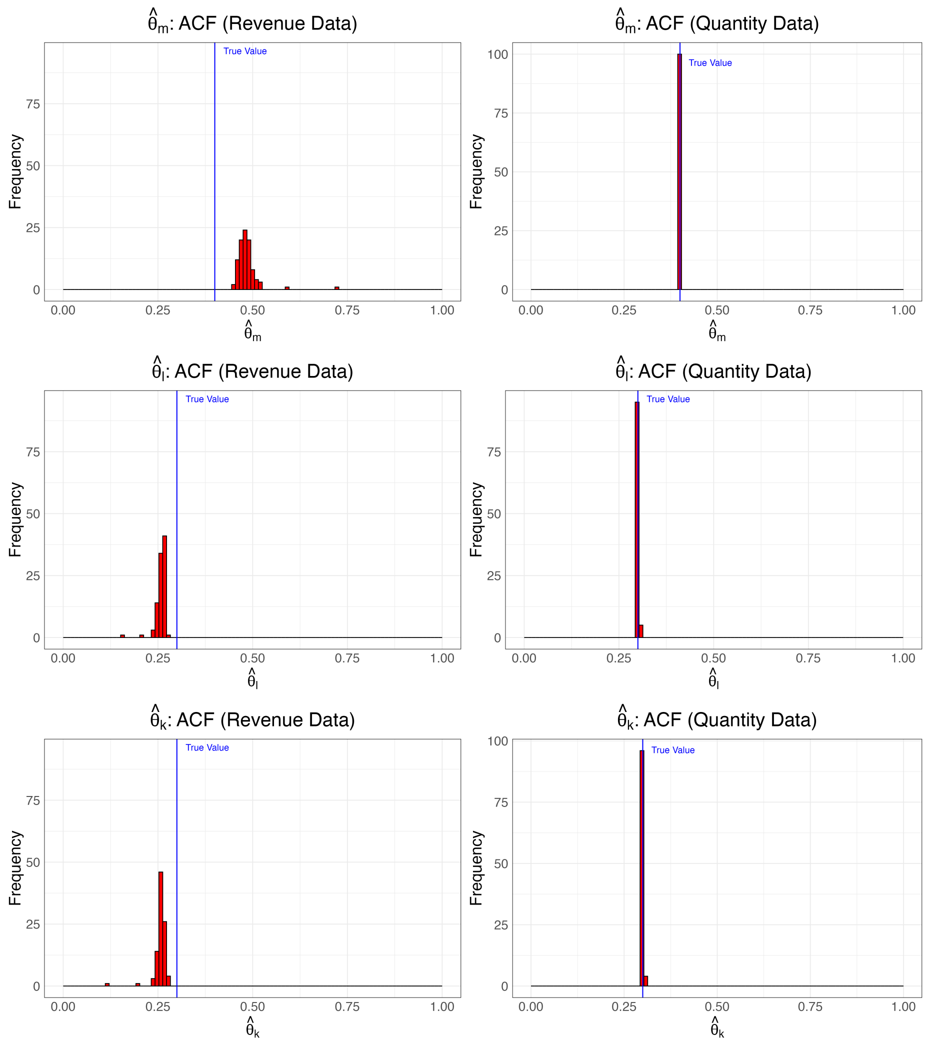


Figure 2 shows the histograms of 100 estimates for $(\theta_m, \theta_k, \theta_l)$ from our proposed estimator and (β_t, δ_t) at $t = T$ on the HSA demand system. Since γ_t varies across simulation, we report the histograms of estimation errors $\hat{\gamma}_t - \gamma_t$. They are tightly clustered around their true values, suggesting that our method recovers the structural parameters very well. Figure 3 shows the scatter plot of true versus estimated TFPs and markups for the first 20 Monte Carlo simulations. 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

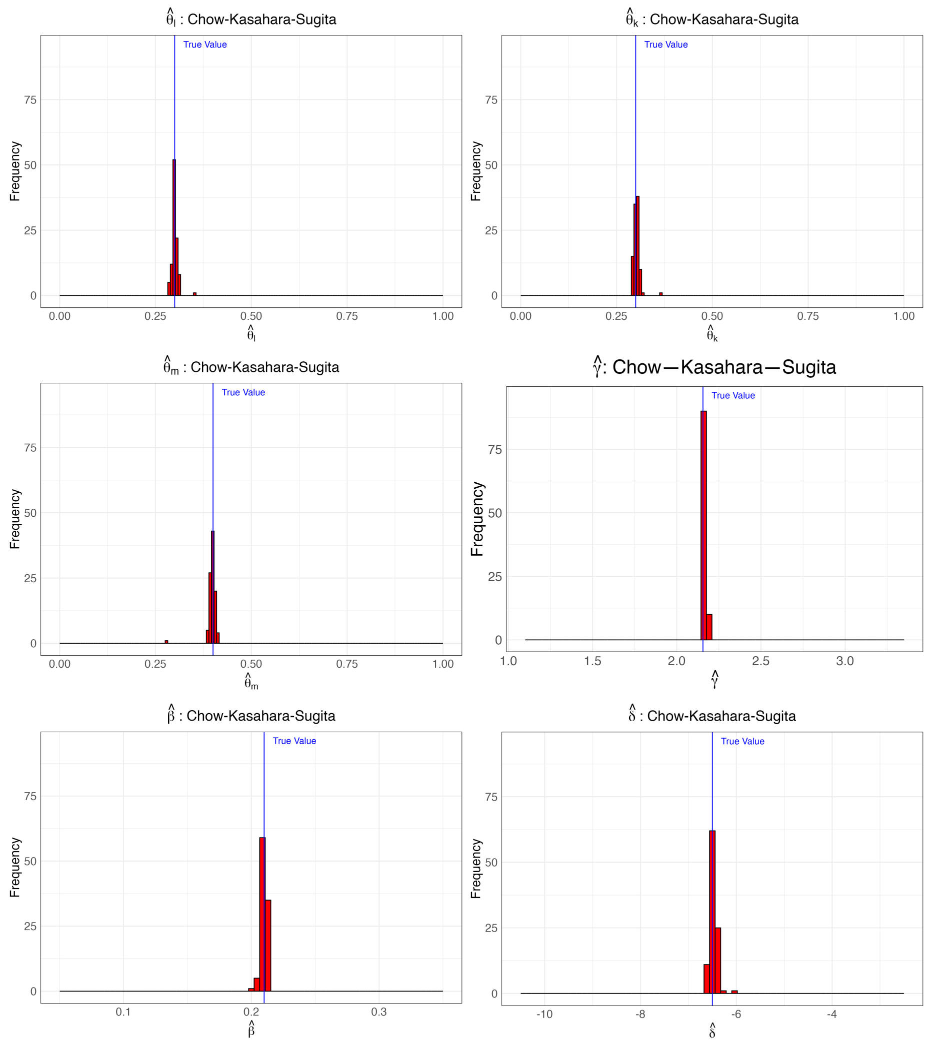
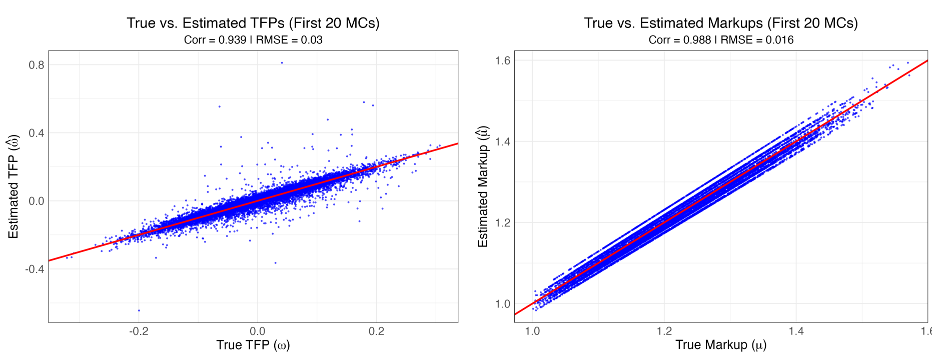


Figure 3: True and estimated TFPs and Markups for first 20 MC Simulations



Notes: The red line indicates the 45-degree line.

5 Empirical Application: Chilean Manufacturing Sector

The semiparametric estimator is applied to the Chilean manufacturing plant dataset, derived from the census conducted by the Chilean Instituto Nacional de EstadÃstica, covering all plants with 10 or more employees from 1993 to 1996. The primary objective of this empirical application is to illustrate the practical feasibility of our method in a well-studied empirical environment, rather than to uncover new stylized facts. The Chilean manufacturing census has been used extensively in the literature, making it a natural benchmark for assessing how our approach performs relative to existing empirical frameworks. In addition, we use this setting to examine whether the commonly imposed CES demand specification is rejected by the data in favor of the more flexible HSA demand system, and to quantify the welfare losses associated with markups through the counterfactual experiments discussed in Section 2.6.

Following the standard approach in this literature, 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 nominal values deflated using industry-specific deflators. Capital input is constructed as the sum of deflated values for buildings, machinery, and vehicles using the perpetual inventory method. Our analysis focuses on the three largest manufacturing industries in 1996, corresponding to 2-digit SIC codes 31 (Food, Beverage, and Tobacco), 32 (Textiles, Apparel, and Leather Products), and 38 (Metal Products, Electric and Non-electric Machinery, Transport Equipment, and Professional Equipment). We exclude plants with non-positive capital, as well as those with material cost-to-revenue ratios below zero, above one, or in the bottom and top two percentiles of the distribution, in order to remove observations that are inconsistent with the production model or likely to reflect reporting errors and extreme measurement noise.

5.1 Result

Industry	n	$\hat{\theta}_m$	$\hat{\theta}_k$	$\hat{\theta}_l$	$\hat{\mu}$
31	736	0.852	0.012	0.136	1.392
		(0.035)	(0.009)	(0.034)	(0.059)
32	463	0.755	0.070	0.174	1.501
		(0.064)	(0.039)	(0.041)	(0.126)
38	391	0.685	0.043	0.272	1.661
		(0.063)	(0.035)	(0.056)	(0.155)

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.

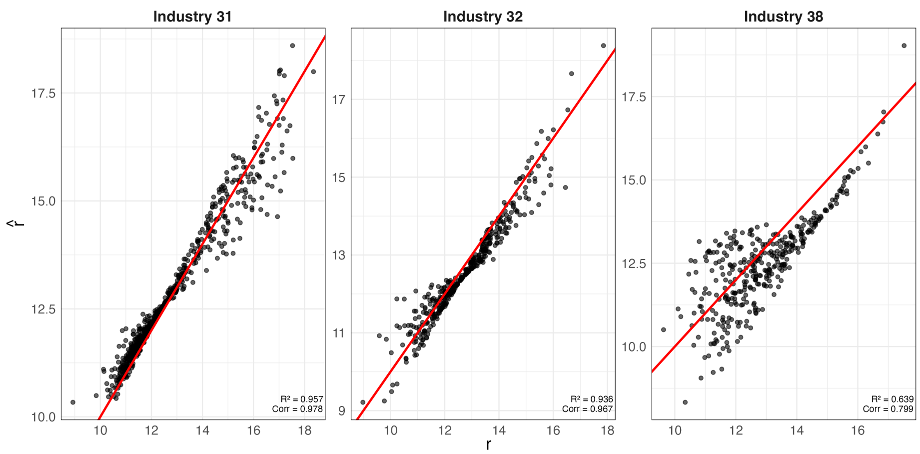
Industry	n	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$
31	700	0.173	1.973	-8.085
		(0.005)	(0.113)	(0.397)
32	408	0.087	0.979	-6.913
		(0.015)	(0.236)	(0.897)
38	351	0.093	1.272	-5.237
		(0.014)	(0.288)	(0.670)

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.

Tables 2 and 3 report estimates of the structural parameters for the Chilean manufacturing industries obtained using our method. Notably, the estimates of β is statistically significantly different from zero. Since the HSA demand system nests the CES demand system as the special case $\beta = 0$, this provides evidence against the CES specification in favor of the more flexible HSA demand system in Chilean manufacturing industries. Appendix G shows that these findings are robust under decreasing returns to scale.

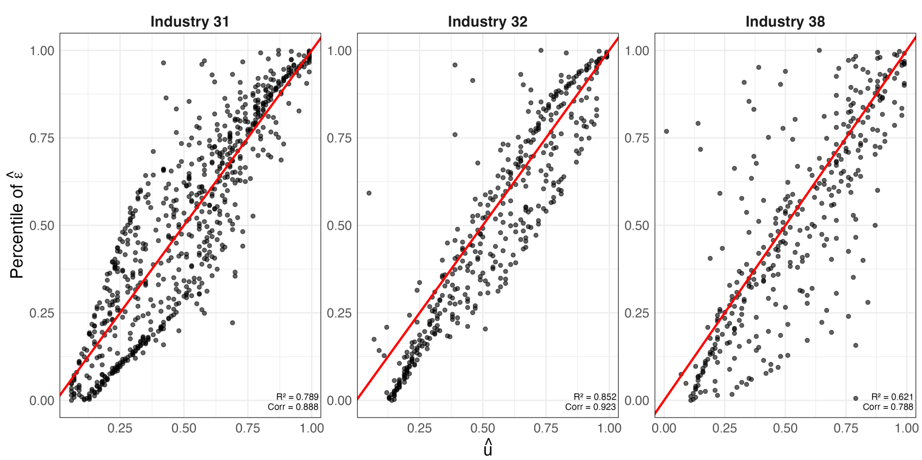
Figures 4 and 5 show scatter plots of observed revenue r_{it} against fitted revenue from the HSA demand system in Step 4, along with quantile–quantile plots comparing the estimated demand shocks ϵ_{it} from Step 4 with the corresponding nonparametric estimates u_{it} from Step 1. Observed revenue aligns closely with fitted revenue under the parametric HSA restriction, and the quantiles of ϵ_{it} generally track those of u_{it} along the 45-degree line. While the fit is not perfect, these patterns provide supportive evidence for the HSA demand specification, particularly for Industries 31 and 32. Appendix G shows that these patterns are preserved under decreasing returns to scale.

Figure 4: Observed revenue vs. fitted revenue from Step 4



Notes: The red line indicates the 45-degree line.

Figure 5: Rank of demand shock from Step 1 vs. Step 4



Notes: The red line indicates the 45-degree line.

5.2 Counterfactual Welfare Analysis

Using the estimated HSA demand system, we quantify the consumer utility loss attributable to firm's market power by calculating the compensation variation for a counterfactual marginal cost pricing equilibrium (MCPE) as described in Section 2.6.

In our counterfactual analysis, we hold fixed the structural primitives: the production function f_t, the demand shock distribution G_u, and the structural taste parameter γ_t. The counterfactual equilibrium under a marginal cost pricing equilibrium (MCPE) recomputes the aggregate index Δq_t^c and firm-level outcomes, ensuring that the equilibrium condition (54). Specifically, we implement the following procedure.

First, we calculate a monopolistic competition equilibrium (MCE), utilizing the estimated structural parameters from Tables 2 and 3 and the firm-level states $(k_{it}, l_{it}, \hat{\omega}_{it}, \hat{\epsilon}_{it})$. Since the observed output values do not exactly satisfy the equilibrium conditions (53)—due to model misspecification and estimation error for γ_t—we recompute the equilibrium output vector $\{y_{it}^m\}_{i=1}^n$ and parameter γ_t^{new} under the normalization $q_t^m = 0$ by jointly solving the conditions in (53), which are simplified as:

$$\Phi_t + \hat{\delta}_t - \hat{\beta}_t y_{it}^m + \gamma_t^{new} + \Xi_{it} - \frac{y_{it}^m}{\hat{\theta}_m} + \frac{1}{\hat{\beta}_t} \ln(1 + \hat{\epsilon}_{it})$$
$$- \left(1 + \frac{1}{\hat{\beta}_t}\right) \ln\left(\exp(-\hat{\beta}_t y_{it}^m + \gamma_t^{new}) + \hat{\epsilon}_{it}\right) = 0 \text{ for } i = 1, .., N_t$$
$$\sum_{i=1}^{N_t} \exp\left(\hat{\delta}_t - \frac{1}{\hat{\beta}_t} \ln\left(\frac{\exp(-\hat{\beta}_t y_{it}^m + \gamma_t^{new}) + \hat{\epsilon}_{it}}{1 + \hat{\epsilon}_{it}}\right)\right) = 1$$

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_t^m is normalized to zero. The resulting output vector and parameters are fully consistent with our HSA demand system, which mitigates model misspecification bias and addresses estimation errors in the counterfactual analysis.

Second, we consider a marginal cost pricing equilibrium (MCPE) for given counterfactual log income Φ_t^c. We find an output vector $\mathbf{y}_t^c(\Phi_t^c)$ and quantity index $\Delta q_t^c(\Phi_t^c)$ by solving (54), which are simplified as:

$$\Phi_t^c + \hat{\delta}_t + \Xi_{it} - \frac{y_{it}^c(\Phi_t^c)}{\hat{\theta}_m} - \frac{1}{\hat{\beta}_t} \ln\left(\frac{\exp(-\hat{\beta}_t \left(y_{it}^c(\Phi_t^c) - \Delta q_t^c(\Phi_t^c)\right) + \gamma_t^{new}) + \hat{\epsilon}_{it}}{1 + \hat{\epsilon}_{it}}\right) = 0 \text{ for } i = 1, .., N_t$$
$$\sum_{i=1}^{N_t} \exp\left(\hat{\delta}_t - \frac{1}{\hat{\beta}_t} \ln\left(\frac{\exp(-\hat{\beta}_t \left(y_{it}^c(\Phi_t^c) - \Delta q_t^c(\Phi_t^c)\right) + \gamma_t^{new}) + \hat{\epsilon}_{it}}{1 + \hat{\epsilon}_{it}}\right)\right) = 1.$$

We assume the same income case $\Phi_t^c = \Phi_t$ for our benchmark, while also analyzing how potential changes in income might affect the equilibrium outcome.

To calculate the compensating variation for transitioning from an MCE to an MCPE, we solve for the counterfactual income Φ_t^{c*} that results in a zero utility change, as defined in (55):

$$\Delta \ln U^c(\Phi_t^{c*}) = \Delta q_t^c(\Phi_t^{c*})$$
$$+ \sum_i \int_{y_{it}^m}^{y_{it}^c(\Phi_t^{c*}) - \Delta q_t^c(\Phi_t^{c*})} \exp\left(\hat{\delta}_t - \frac{1}{\hat{\beta}_t} \ln\left(\frac{\exp(-\hat{\beta}_t(\zeta - q_t^m) + \gamma_t^{new}) + \hat{\epsilon}_{it}}{1 + \hat{\epsilon}_{it}}\right)\right) d\zeta = 0.$$

The compensating variation is then calculated as $CV_t := \exp(\Phi_t^{c*}) - \exp(\Phi_t)$, which quantifies the consumer's welfare change from firm's market power.

Industry	CV	$\Delta\Pi$	Overall
31	-14.7	-11.5	3.20
	(2.65)	(2.43)	(0.648)
32	-18.9	-11.9	6.93
	(7.74)	(7.14)	(1.95)
38	-13.3	-6.12	7.14
	(11.3)	(7.82)	(4.28)

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. Sign convention: $CV_t < 0$ indicates consumers require less income under competition (gain from eliminating market power); $\Delta\Pi_t < 0$ is firms' profit loss; Overall $= \Delta\Pi_t - CV_t > 0$ is the net welfare gain. Standard errors in parentheses with 100 non-parametric bootstrap iterations.

Finally, we calculate firms' profit loss. In the case of $\Phi_t^c = \Phi_t$, the total profit change (56) is expressed as

$$\Delta\Pi := \Pi^c - \Pi^m = \sum_{i=1}^{N_t} \exp\left(-\frac{\hat{\theta}_k k_{it} + \hat{\theta}_l l_{it} + \hat{\omega}_{it}}{\hat{\theta}_m}\right)\left\{\exp\left(\frac{y_{it}^m}{\hat{\theta}_m}\right) - \exp\left(\frac{y_{it}^c(\Phi_t)}{\hat{\theta}_m}\right)\right\} \tag{66}$$

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

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. To put these findings in context, our estimates far exceed the classic Harberger-triangle calculation of less than 0.1% of GDP (Harberger, 1954). They are more consistent with recent general-equilibrium analyses that find substantially larger welfare costs. Baqaee and Farhi (2020) estimate that eliminating misallocation would raise output by approximately 15%, while Edmond et al. (2023) find welfare costs of variable markups of around 7.5% in consumption-equivalent terms. Our estimates, though derived from a different structural framework, are comparable in magnitude—reinforcing the view that the welfare costs of imperfect competition are economically significant. Appendix G shows that these estimates are broadly robust under moderate decreasing returns to scale of 0.9.

6 Concluding Remarks

This paper develops constructive nonparametric identification of production functions and markups from revenue data, simultaneously addressing the two fundamental challenges in production function estimation since Marschak and Andrews (1944) when revenue is used as output: input-TFP correlation and markup heterogeneity bias. By modeling revenue as a function of output, observed characteristics, and an unobserved demand shock, we identify various economic objects of interest under standard assumptions. Our semiparametric estimator, implementable with standard datasets, performs well in simulation. Applied to Chilean manufacturing plant data, we find evidence of CES demand misspecification and estimate that market power generates welfare losses of approximately 3%–6% of industry revenue in the three largest manufacturing industries in 1996.

While our identification results establish that firm-level revenue data suffice to recover production functions and consumer demand without physical quantity data, the proof relies on structural assumptions that merit discussion. First, we assume perfectly competitive input markets; monopsony power would make input prices firm-specific and endogenous, breaking the link between expenditure shares and marginal products. Second, we assume Hicks-neutral, scalar productivity; factor-augmenting technological change requires additional structure or data for identification. Third, we assume monopolistic competition; in oligopolistic markets, strategic interaction makes the first-order conditions underlying our markup identification insufficient without a fully specified equilibrium model. Relaxing any of these assumptions is an important direction for future work.

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.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015), "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83, 2411–2451.

Aczél, J. and Dhombres, J. (1989), *Functional Equations in Several Variables: With Applications to Mathematics, Information Theory and to the Natural and Social Sciences*, vol. 31 of *Encyclopedia of Mathematics and its Applications*, Cambridge, UK: Cambridge University Press.

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.

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

Baqaee, D. R. and Farhi, E. (2020), "Productivity and Misallocation in General Equilibrium," *Quarterly Journal of Economics*, 135, 105–163.

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

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

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

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

Bond, S., Hashemi, A., Kaplan, G., and Zoch, P. (2021), "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data," *Journal of Monetary Economics*, 121, 1–14.

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

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

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

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.

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

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

De Ridder, M., Grassi, B., and Morzenti, G. (forthcoming), "The Hitchhiker's Guide to Markup Estimation: Assessing Estimates from Financial Data," *Econometrica*, accepted 2025.

Demirer, M. (2025), "Production Function Estimation with Factor-Augmenting Technology: An Application to Markups," Conditionally accepted, *Econometrica*.

Doraszelski, U. and Jaumandreu, J. (2013), "R&D and Productivity: Estimating Endogenous Productivity," *The Review of Economic Studies*, 80, 1338–1383.

— (2018), "Measuring the Bias of Technological Change," *Journal of Political Economy*, 126, 1027–1084.

— (2021), "Reexamining the De Loecker & Warzynski (2012) Method for Estimating Markups," Discussion Paper 16027, CEPR.

Edmond, C., Midrigan, V., and Xu, D. Y. (2023), "How Costly Are Markups?" *Journal of Political Economy*, 131, 1619–1675.

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

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

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

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

Flynn, Z., Gandhi, A., and Traina, J. (2019), "Measuring Markups with Production Data," working paper.

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

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

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.

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

Harberger, A. C. (1954), "Monopoly and Resource Allocation," *The American Economic Review*, 44, 77–87.

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

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

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

Hsieh, C.-T. and Klenow, P. J. (2009), "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, 1403–1448.

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

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

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.

Kirov, I., Mengano, P., and Traina, J. (2025), "Measuring Markups with Revenue Data," Working Paper, available at SSRN.

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.

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

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

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.

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

Matsuyama, K. (2023), "Non-CES aggregators: a guided tour," *Annual Review of Economics*, 15, 235–265.

— (2025), "Homothetic Non-CES demand systems with applications to monopolistic competition," *Annual Review of Economics*, 17, 261–292.

Matsuyama, K. and Ushchev, P. (2017), "Beyond CES: Three Alternative Classes of Flexible Homothetic Demand Systems," CEPR Discussion Papers DP12210.

— (2020), "Constant Pass-Through," CEPR Discussion Papers DP 15475.

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

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.

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.

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

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

Stewart, J. (2012), *Calculus: Early Transcendentals*, Boston, MA: Cengage Learning, 7th ed.

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

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

Online Supplemental Appendix

A Simulation

A.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

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

While the original formulation 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}	$1+\epsilon_{it}$	γ_t
Matsuyama and Ushchev (2020)'s notations	ω	$\ln \gamma_\omega \beta_\omega$	$\Delta = \frac{1-\rho}{\rho}$	$\frac{1}{\sigma_\omega - 1}$	$\frac{\sigma_\omega}{\sigma_\omega - 1}$	$\Delta \ln \gamma_\omega$
Range		$(-\infty, \infty)$	$(0, \infty)$	$(0, \infty)$	$(1, \infty)$	$(-\infty, \infty)$

A.2 Data Generating Process

The demand shock ϵ_{it} follows an MA(1) 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:

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

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

Endogenous variables						
$t = 5$	Mean	Min	P25	Median	P75	Max
Markup	1.223	1.001	1.147	1.219	1.295	1.617
m_{it}	11.322	8.690	10.948	11.318	11.691	14.161
r_{it}	13.437	10.712	13.056	13.445	13.823	16.231
Exogenous variables						
$t = 5$	Mean	Min	P25	Median	P75	Max
ω_{it}	0.000	−0.338	−0.056	0.000	0.056	0.323
k_{it}	9.511	4.770	8.771	9.513	10.253	15.055
l_{it}	9.505	4.949	8.770	9.505	10.246	14.252

A.2.1 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.

B Calculations and Proofs

B.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

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

The first order condition for (7) is

$$\begin{aligned}\frac{\partial \pi_t}{\partial m_{it}} =& \exp(\varphi_t(f_t(x_{it}, z_{it}^s) + \omega_{it}, z_{it}^d, u_{it}))\frac{\partial \varphi_t(f_t(x_{it}, z_{it}^s) + \omega_{it}, z_{it}^d, u_{it})}{\partial y_{it}} \frac{\partial f_t(x_{it}, z_{it}^s)}{\partial m_{it}} \\ &- \exp(p_t^m + m_{it}) = 0.\end{aligned}$$

Their cross derivatives are

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

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 $\partial f_t / \partial m_{it} > 0$ implies $\left(\frac{\partial \varphi_t}{\partial y_{it}}\right)^2 + \frac{\partial^2 \varphi_t}{\partial y_{it}^2} > 0$. Therefore, we have

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

B.2 Proof of Proposition 4

Suppose that Assumption 8(a) holds. Let var(η_t) and var(η_{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 (27) and (41), $\text{var}(\eta_t) = b_t^2 \text{var}(\eta_t^*)$ 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 $b_{t+1}/b_t = \sqrt{\text{var}(\eta_{t+1})/\text{var}(\eta_t)}$.

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 (41), $\partial f_t(x_t, z_t^s)/\partial q = b_t \partial f_t^*(x_t, z_t^s)/\partial q$ 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 $b_{t+1}/b_t = (\partial f_{t+1}(x,z^s)/\partial q)/(\partial f_t(x,z^s)/\partial q)$ 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

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

B.3 Proof of Proposition 6

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 (42) 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 (43) and (45), 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:

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

From (44), we identify the output growth rate $\varphi_{t+1}^{*-1}(r_{it+1}, z_{it+1}^d, u_{it+1}) - \varphi_t^{*-1}(r_{it}, z_{it}^d, u_{it})$.

Evaluating the second equation in (44) at $x_{t+1} = x_t = \bar{x}$ and $z_{t+1}^s = z_t^s = \bar{z}^s$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 (B.2), $\tilde{a}_{2t+1} - \tilde{a}_{2t}$ 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 (44), the true TFP growth rate $\omega^*_{it+1} - \omega^*_{it}$ is also identified. $\square$

B.4 Derivations of Equilibrium Conditions for MCE and MCPE

General HSA Demand System

MCE Define the inverse production function $m_{it} = \chi_{it}(y_{it})$ such that $y_{it} = f_t\left(\chi_{it}(y_{it}), k_{it}, l_{it}, z^s_{it}\right) + \omega_{it}$ for given $(k_{it}, l_{it}, \omega_{it}, z^s_{it})$; namely, $\chi_{it}(y_{it}) := f_t^{-1}\left(y_{it} - \omega_{it}, k_{it}, l_{it}, z^s_{it}\right)$. By using this, we rewrite the profit maximization problem with respect to m_{it}

$$\max_m \exp\left(\Phi_t + \mathfrak{s}_t\left(f_t\left(m, k_{it}, l_{it}, z^s_{it}\right) + \omega_{it} - \Delta q^m_t, z^d_{it}, u_{it}\right)\right) - \exp\left(p^m_t + m\right)$$

as the problem with respect to y_{it}:

$$\max_y \exp\left(\Phi_t + \mathfrak{s}_t\left(y - \Delta q^m_t, z^d_{it}, u_{it}\right)\right) - \exp\left(p^m_t + \chi_{it}(y)\right)$$

The first order condition is

$$\exp\left(\Phi_t + \mathfrak{s}_t\left(y^m_{it} - \Delta q^m_t, z^d_{it}, u_{it}\right)\right) \frac{\partial \mathfrak{s}_t\left(y^m_{it} - \Delta q^m_t, z^d_{it}, u_{it}\right)}{\partial y_{it}} = \exp\left(p^m_t + \chi_{it}(y^m_{it})\right) \frac{\partial \chi_{it}(y^m_{it})}{\partial y_{it}}.$$

Noting that the marginal cost is

$$\frac{\partial \exp\left(p^m_t + \chi_{it}(y^m_{it})\right)}{\partial \exp(y_{it})} = \frac{\exp\left(p^m_t + \chi_{it}\left(y^m_{it}\right)\right)}{\exp\left(y^m_{it}\right)} \frac{\partial \chi_{it}(y^m_{it})}{\partial y_{it}},$$

the system of the first order conditions and the market share condition becomes:

$$\underbrace{\exp\left(\mathfrak{s}_t\left(y^m_{it} - \Delta q^m_t, z^d_{it}, u_{it}\right) + \Phi_t - y^m_{it}\right) \frac{\partial \mathfrak{s}_t\left(y^m_{it} - \Delta q^m_t, z^d_{it}, u_{it}\right)}{\partial y_{it}}}_{\text{Marginal Revenue}} = \underbrace{\frac{\exp\left(p^m_t + \chi_{it}\left(y^m_{it}\right)\right)}{\exp\left(y^m_{it}\right)} \frac{\partial \chi_{it}(y^m_{it})}{\partial y_{it}}}_{\text{Marginal Cost}},$$

$$\sum_{i=1}^{N_t} \exp\left(\mathfrak{s}_t\left(y^m_{it} - \Delta q^m_t, z^d_{it}, u_{it}\right)\right) = 1.$$

MCPE The profit maximization problem with respect to y_{it} is

$$\max_y \exp\left(p^c_{it} + y\right) - \exp\left(p^m_t + \chi_{it}(y)\right)$$

where the firm takes p^c_{it} as given. The first order condition is

$$\exp\left(p_{it}^c + y_{it}^c\right) = \exp\left(p_t^m + \chi_{it}\left(y_{it}^c\right)\right)\frac{\partial \chi_{it}(y_{it}^c)}{\partial y_{it}}$$

From $p_{it}^c = \Phi_t^c + \mathfrak{s}_t\left(y_{it}^c - \Delta q_t^c, z_{it}^d, u_{it}\right) - y_{it}^c$ in equilibrium, the system of the first order conditions and the market share condition for the MCPE is

$$\underbrace{\exp\left(\mathfrak{s}_t\left(y_{it}^c - \Delta q_t^c, z_{it}^d, u_{it}\right) + \Phi_t - y_{it}^c\right)}_{\text{Price}} = \underbrace{\frac{\exp\left(p_t^m + \chi_{it}\left(y_{it}^c\right)\right)}{\exp\left(y_{it}^c\right)}\frac{\partial \chi_{it}(y_{it}^c)}{\partial y_{it}}}_{\text{Marginal Cost}},$$

$$\sum_{i=1}^{N_t}\exp\left(\mathfrak{s}_t\left(y_{it}^c - \Delta q_t^c, z_{it}^d, u_{it}\right)\right) = 1,$$

CoPaTh-HSA Demand System and Cobb-Douglas Production Function Consider the CoPaTh-HSA Demand System (64):

$$\mathfrak{s}_t\left(y_{it} - q_t\left(\mathbf{y}_t, \boldsymbol{\epsilon}_t\right), \epsilon_{it}\right) = \delta_t - \frac{1}{\beta_t}\ln\left(\frac{\exp\left(-\beta_t\left(y_{it} - q_t\left(\mathbf{y}_t, \boldsymbol{\epsilon}_t\right)\right) + \gamma_t\right) + \epsilon_{it}}{1 + \epsilon_{it}}\right).$$

MCE Note that $q_t\left(\mathbf{y}_t^m, \boldsymbol{\epsilon}_t\right) = 0$ holds as normalization in the initial MCE. Since

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

the first order condition (53) becomes

$$\begin{aligned}&\exp\left(\delta_t - \frac{1}{\beta_t}\ln\left(\frac{\exp(-\beta_t y_{it}^m + \gamma_t) + \epsilon_{it}}{1 + \epsilon_{it}}\right) + \Phi_t - y_{it}^m\right)\left(\frac{\exp(-\beta_t y_{it}^m + \gamma_t)}{\exp(-\beta_t y_{it}^m + \gamma_t) + \epsilon_{it}}\right)\\ =&\exp\left(p_t^m + \frac{y_{it}^m - \theta_k k_{it} - \theta_l l_{it} - \omega_{it}}{\theta_m} - y_{it}^m\right)\frac{1}{\theta_m}.\end{aligned}$$

Taking the log of both sides yields

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

Letting $\Xi_{it} := \ln\theta_m + (\theta_k k_{it} + \theta_l l_{it} + \omega_{it})/\theta_m$ and $p_t^m = 0$, the system of the first order conditions and the market share condition is simplified as

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

MCPE The first order condition (54) becomes

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

Taking the log of both sides, the system of the first order conditions and the market share condition is simplified as

$$\Phi_t + \delta_t + \Xi_{it} - \frac{y_{it}^c}{\theta_m} - \frac{1}{\beta_t}\ln\left(\frac{\exp\left(-\beta_t\left(y_{it}^c - \Delta q_t^c\right) + \gamma_t\right) + \epsilon_{it}}{1+\epsilon_{it}}\right) = 0 \text{ for } i = 1,..,N_t$$
$$\sum_{i=1}^{N_t}\exp\left(\delta_t - \frac{1}{\beta_t}\ln\left(\frac{\exp\left(-\beta_t\left(y_{it}^c - \Delta q_t^c\right) + \gamma_t\right) + \epsilon_{it}}{1+\epsilon_{it}}\right)\right) = 1.$$

B.5 Proof for Proposition 7

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

Theorem B.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*

$$\sum_{i=1}^{N} S_i(Y_i^*) = 1, \tag{B.3}$$

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

$$\begin{aligned} S_i'(Y_i)Y_i &< S_i(Y_i) \text{ for } i = 1, ..., N, \\ S_i'(Y_i)S_j'(Y_j) &\geq 0 \text{ for } i, j = 1, ..., N, \end{aligned} \tag{B.4}$$

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

$$P_i = \frac{I}{Y_i} S_i\left(\frac{Y_i}{Q(\mathbf{Y})}\right) \text{ for } i = 1, .., N, \tag{B.5}$$

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

$$\sum_{i=1}^{N} S_i\left(\frac{Y_i}{Q(\mathbf{Y})}\right) = 1. \tag{B.6}$$

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

$$\ln U(\mathbf{Y}) = \ln Q(\mathbf{Y}) + \sum_{i=1}^{N} \int_{c_i}^{Y_i/Q(Y)} \frac{S_i(\xi)}{\xi} d\xi, \tag{B.7}$$

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

Proof for Proposition 7

Proof. (a) For any $(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)$, we identify $\Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)$ and $\mathfrak{s}_t^*(\tilde{y}_{it} - q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t), \tilde{z}_{it}^d, \tilde{u}_{it})$ from (51)–(52) as discussed in the main text, where the identification of reduced-form functions $\varphi_t(\cdot)$ and $\mathfrak{s}_t(\cdot)$ at the baseline aggregate state $(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ in the data follows from Proposition 5.

Fix $\tilde{\mathbf{z}}_t^d := (\tilde{z}_{1t}^d, \ldots, \tilde{z}_{Nt}^d)$, $\tilde{\mathbf{u}}_t := (\tilde{u}_{1t}, \ldots, \tilde{u}_{Nt})$, and time t. For given $\tilde{\mathbf{Y}} \in \mathscr{Y}_t^N$, define $\boldsymbol{S}_t(\tilde{\mathbf{Y}}) := \left(S_{1t}(\tilde{Y}_1), \ldots, S_{N_t t}(\tilde{Y}_{N_t})\right)$ with

$$\begin{aligned} S_{it}(\tilde{Y}_{it}) &:= \exp\left(\mathfrak{s}_t^*\left(\ln \tilde{Y}_{it} - q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t), \tilde{z}_{it}^d, \tilde{u}_{it}\right)\right) \\ &= \exp\left(\mathfrak{s}_t\left(\ln \tilde{Y}_{it}, \tilde{z}_{it}^d, \tilde{u}_{it}\right)\right) \\ &= \exp\left(\varphi_t\left(\ln \tilde{Y}_{it}, \tilde{z}_{it}^d, \tilde{u}_{it}\right) - \Phi_t\right) \quad \text{for } i = 1, ..., N_t, \end{aligned} \tag{B.8}$$

where the second and third equalities follow from (50) and (49), respectively. By definition,

$\sum_{i=1}^{N_t} S_{it}(\tilde{Y}_i) = 1$ holds. Also, define

$$Q_t(\tilde{\mathbf{Y}}) := \exp\left(\Delta q_t(\ln \tilde{\mathbf{Y}}, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)\right)$$
$$= \exp\left(q_t(\ln \tilde{\mathbf{Y}}, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)\right) \tag{B.9}$$

where the second equality follows from the normalization $q_t(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t) = 0$ at the baseline aggregate state $(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t)$.

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

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

holds for all i and $\tilde{y}$. The above inequality implies

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

because

$$S'_{it}(\tilde{Y})\tilde{Y} = \exp\left(\varphi_t\left(\ln \tilde{Y}, \tilde{z}_{it}^d, \tilde{u}_{it}\right) - \Phi_t\right) \frac{\partial \varphi_t\left(\ln \tilde{Y}, \tilde{z}_{it}^d, \tilde{u}_{it}\right)}{\partial \ln \tilde{Y}}$$
$$= S_{it}(\tilde{Y}) \frac{\partial \varphi_t\left(\ln \tilde{Y}, \tilde{z}_{it}^d, \tilde{u}_{it}\right)}{\partial \ln \tilde{Y}}.$$

Consequently, $\mathbf{S}(\tilde{\mathbf{Y}})$ satisfies the inequalities in (B.4) for all $\tilde{\mathbf{Y}}$ satisfying $\sum_{i=1}^{N_t} S_{it}(\tilde{Y}_i) = 1$.

Therefore, from Theorem B.1(a), there exists a unique monotone, convex, continuous, and homothetic rational preference that generates the following HSA demand system:

$$\tilde{P}_{it} = \frac{I_t}{\tilde{Y}_{it}} S_{it}\left(\frac{\tilde{Y}_{it}}{Q_t\left(\tilde{\mathbf{Y}}_t\right)}\right) \text{ for } i = 1, .., N,$$
$$\sum_{i=1}^{N} S_{it}\left(\frac{\tilde{Y}_{it}}{Q_t\left(\tilde{\mathbf{Y}}_t\right)}\right) = 1$$

where $\tilde{P}_{it}\tilde{Y}_{it} = \varphi_t\left(\ln \tilde{Y}_{it}, \tilde{z}_{it}^d, \tilde{u}_{it}\right)$ and $I_t = \sum_{i=1}^{N} \tilde{P}_i \tilde{Y}_{it}$ is the consumer's budget. Taking the log of this demand system and using (B.8) and (B.9) with $\tilde{y}_{it} := \ln \tilde{Y}_{it}$, we obtain (46) and (47) as

$$\ln\left(\frac{\exp\left(\tilde{r}_{it}\right)}{\sum_{j=1}^{N_t}\exp\left(\tilde{r}_{jt}\right)}\right) = \mathfrak{s}_t^*\left(\tilde{y}_{it} - q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t), \tilde{z}_{it}^d, \tilde{u}_{it}\right)$$
$$1 = \sum_{i=1}^{N_t}\mathfrak{s}_t^*\left(\tilde{y}_{it} - q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t), \tilde{z}_{it}^d, \tilde{u}_{it}\right)$$

where $\tilde{r}_{it} := \varphi_t\left(\ln \tilde{Y}_{it}, \tilde{z}_{it}^d, \tilde{u}_{it}\right)$.

(b) Fix $\tilde{\mathbf{z}}_t^d := (\tilde{z}_{1t}^d, \ldots, \tilde{z}_{Nt}^d)$, $\tilde{\mathbf{u}}_t := (\tilde{u}_{1t}, \ldots, \tilde{u}_{Nt})$, and time t. For $\tilde{\mathbf{Y}}_t \in \mathscr{Y}_t^N$ and $\tilde{\mathbf{y}}_t = \ln \tilde{\mathbf{Y}}_t$, let $\bar{U}_t(\tilde{\mathbf{Y}}_t) := U_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)$ be the utility function of the representative consumer. From Theorem B.1(b) and using (B.8) and (B.9), the homothetic preference is described as

$$\begin{aligned}\ln \bar{U}_t(\tilde{\mathbf{Y}}_t) &= \ln Q_t(\tilde{\mathbf{Y}}_t) + \sum_{i=1}^{N}\int_{c_i}^{\tilde{Y}_{it}/Q(\tilde{\mathbf{Y}}_t)}\frac{S_{it}(\xi)}{\xi}d\xi.\\ &= \Delta q_t(\ln \tilde{\mathbf{Y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) + \sum_{i=1}^{N}\int_{c_i}^{\tilde{Y}_i/\exp\left(\Delta q_t(\ln \tilde{\mathbf{Y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)\right)}\frac{\exp\left(\mathfrak{s}_t\left(\ln \xi, \tilde{z}_{it}^d, \tilde{u}_{it}\right)\right)}{\xi}d\xi.\end{aligned}$$

Applying a change in variable $\zeta = \ln \xi$ and $d\zeta = \frac{d\xi}{\xi}$, we write

$$\ln U_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) = \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) + \sum_{i=1}^{N}\int_{b_i}^{\tilde{y}_{it} - \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)}\exp\left(\mathfrak{s}_t\left(\zeta, \tilde{z}_{it}^d, \tilde{u}_{it}\right)\right)d\zeta,$$

where $b_i = \ln c_i$ such that $\ln U_t(\mathbf{b}, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) = 0$ for $\mathbf{b} = (b_1, ..., b_{N_t})$ from Theorem B.1(b).

(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. Suppose that $\{\tilde{y}_{it}\}_{i=1}^{N_t}$ be a log consumption vector for given price and income. If the price remains the same and the income is multiplied by $\exp(\lambda)$, then the log-consumption vector becomes $\{\tilde{y}_{it} + \gamma\}_{i=1}^{N_t}$ to keep the same market share for each product. This property implies that

$$\begin{aligned}1 &= \sum_{i=1}^{N_t}\exp\left(\mathfrak{s}_t(\tilde{y}_{it} - \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t), \tilde{z}_{it}^d, \tilde{u}_{it}\right)\\ &= \sum_{i=1}^{N_t}\exp\left(\mathfrak{s}_t(\tilde{y}_{it} + \gamma - \Delta q_t(\tilde{\mathbf{y}}_t + \gamma, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t), \tilde{z}_{it}^d, \tilde{u}_{it}\right).\end{aligned}$$

On the other hand, from

$$
\begin{aligned}
1 &= \sum_{i=1}^{N_t} \exp\left(\mathfrak{s}_t(\tilde{y}_{it} - \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t), \tilde{z}_{it}^d, \tilde{u}_{it}\right) \\
&= \sum_{i=1}^{N_t} \exp\left(\mathfrak{s}_t(\tilde{y}_{it} + \gamma - \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) - \gamma, \tilde{z}_{it}^d, \tilde{u}_{it}\right),
\end{aligned}
$$

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

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

The utility is expressed as:

$$
\begin{aligned}
\ln U_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) &= \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) + \sum_{i=1}^{N} \int_{\ln c_i}^{\tilde{y}_{it} - \Delta q_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)} \exp\left(\mathfrak{s}_t\left(\zeta, \tilde{z}_{it}^d, \tilde{u}_{it}\right)\right) d\zeta. \\
&= \Delta q_t(\mathbf{y}_t^* + a, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) + \sum_{i=1}^{N} \int_{\ln c_i}^{\tilde{y}_{it}^* - \Delta q_t(\mathbf{y}_t^*, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)} \exp\left(\mathfrak{s}_t\left(\zeta, \tilde{z}_{it}^d, \tilde{u}_{it}\right)\right) d\zeta \\
&= a + \Delta q_t(\mathbf{y}_t^*, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t) + \sum_{i=1}^{N} \int_{\ln c_i}^{\tilde{y}_{it}^* - \Delta q_t(\mathbf{y}_t^*, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t)} \exp\left(\mathfrak{s}_t\left(\zeta, \tilde{z}_{it}^d, \tilde{u}_{it}\right)\right) d\zeta \\
&= a + \ln U(\mathbf{y}_t^*, \tilde{\mathbf{z}}_t^d, \tilde{\mathbf{u}}_t).
\end{aligned}
$$

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. □

B.6 Derivation of the Control Function under a Production Function that is Separable in Materials

While the main text focuses on the Cobb–Douglas production function, here we consider a more general production function that is separable in materials:

$$
f_t(m_{it}, k_{it}, l_{it}, z_{it}^s) = f_{1t}(m_{it}) + f_{2t}\left(k_{it}, l_{it}, z_{it}^s\right).
$$

Define a composite variable

$$\Omega_{it} \equiv f_{2t}\left(k_{it}, l_{it}, z_{it}^{s}\right) + \omega_{it}.$$

The first-order condition for problem (7) can then be written as

$$\exp\Big(\varphi_t\big(f_{1t}(m_{it}) + \Omega_{it}, z_{it}^{d}, u_{it}\big)\Big) \frac{\partial \varphi_t\big(f_{1t}(m_{it}) + \Omega_{it}, z_{it}^{d}, u_{it}\big)}{\partial y_{it}} f_{1t}'(m_{it}) = \exp\big(p_t^{m} + m_{it}\big).$$

This condition implicitly defines the material demand function

$$m_{it} = \mathbb{M}_t\big(\Omega_{it}, z_{it}^{d}, u_{it}\big).$$

Since $\mathbb{M}_t\big(\cdot, z_{it}^{d}, u_{it}\big)$ is strictly increasing in Ω_{it} and therefore invertible. Its inverse is given by

$$\Omega_{it} = \lambda_t\big(m_{it}, z_{it}^{d}, u_{it}\big).$$

The control function for ω_{it} is expressed as

$$\omega_{it} = \lambda_t\big(m_{it}, z_{it}^{d}, u_{it}\big) - f_{2t}\big(k_{it}, l_{it}, z_{it}^{s}\big).$$

Substituting this expression into the revenue function yields

$$\begin{aligned}\varphi_t\big(f_t(m_{it}, k_{it}, l_{it}, z_{it}^{s}) + \omega_{it}, z_{it}^{d}, u_{it}\big) &= \varphi_t\Big(f_{1t}(m_{it}) + \lambda_t\big(m_{it}, z_{it}^{d}, u_{it}\big), z_{it}^{d}, u_{it}\Big) \\ &\equiv \phi_t\big(m_{it}, z_{it}^{d}, u_{it}\big).\end{aligned}$$

C Observational Equivalence and Identification up to Location and Scale

Proposition C.1 (Observational Equivalence and Identification up to Location and Scale)**.** *Suppose Assumptions 1–3 hold. Let* $\{\varphi_t^{*-1}(\cdot), f_t^{*}(\cdot), \mathbb{M}_t^{*-1}(\cdot)\}$ *denote the true model structure satisfying*

$$\varphi_t^{-1}(r_{it}, z_{it}^{d}, u_{it}) = f_t(x_{it}, z_{it}^{s}) + \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}). \tag{C.10}$$

Define the equivalence class

$$\begin{aligned}\mathscr{E}_t := \Big\{ \big(\varphi_t^{-1}, f_t, \mathbb{M}_t^{-1}\big) : \exists\, (a_{1t}, a_{2t}, b_t) \in \mathbb{R}^2 \times \mathbb{R}_{++} \text{ such that}\\ \varphi_t^{-1}(r, z^d, u) = (a_{1t} + a_{2t}) + b_t \varphi_t^{*-1}(r, z^d, u),\\ f_t(x, z^s) = a_{1t} + b_t f_t^*(x, z^s),\\ \mathbb{M}_t^{-1}(m, w, u) = a_{2t} + b_t \mathbb{M}_t^{*-1}(m, w, u) \Big\}.\end{aligned}$$

Then:

(i) Every element $(\varphi_t^{-1}, f_t, \mathbb{M}_t^{-1}) \in \mathscr{E}_t$ *satisfies* (C.10) *and generates the same population joint distribution of* $\{r_{is}, m_{is}, v_{is}\}_{s=t-\upsilon-2}^{t}$ *as the true structure.*

(ii) Conversely, if a structure $(\tilde{\varphi}_t^{-1}, \tilde{f}_t, \tilde{\mathbb{M}}_t^{-1})$ *satisfies* (C.10) *and generates the same population joint distribution of* $\{r_{is}, m_{is}, v_{is}\}_{s=t-\upsilon-2}^{t}$, *then* $(\tilde{\varphi}_t^{-1}, \tilde{f}_t, \tilde{\mathbb{M}}_t^{-1}) \in \mathscr{E}_t$.

(iii) The equivalence class $\mathscr{E}_t$ *is exactly identified by the population joint distribution of* $\{r_{is}, m_{is}, v_{is}\}_{s=t-\upsilon-2}^{t}$.

Proof. We prove each part in turn.

Proof of (i). Let $(\varphi_t^{-1}, f_t, \mathbb{M}_t^{-1}) \in \mathscr{E}_t$ with associated constants (a_{1t}, a_{2t}, b_t). We verify that this structure satisfies (C.10) and generates the same joint distribution of observables.

First, we verify that (C.10) holds. By definition of $\mathscr{E}_t$,

$$\begin{aligned}f_t(x_{it}, z_{it}^s) + \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it}) &= \big[a_{1t} + b_t f_t^*(x_{it}, z_{it}^s)\big] + \big[a_{2t} + b_t \mathbb{M}_t^{*-1}(m_{it}, w_{it}, u_{it})\big]\\ &= (a_{1t} + a_{2t}) + b_t \big[f_t^*(x_{it}, z_{it}^s) + \mathbb{M}_t^{*-1}(m_{it}, w_{it}, u_{it})\big]\\ &= (a_{1t} + a_{2t}) + b_t \varphi_t^{*-1}(r_{it}, z_{it}^d, u_{it})\\ &= \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it}),\end{aligned}$$

where the third equality uses the fact that the true structure satisfies (C.10).

Second, we show that the joint distribution of observables is unchanged. Note that (r_{it}, m_{it}, v_{it}) are directly observed. The equivalence class transformation redefines the latent output as $y_{it} = \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it}) = (a_{1t} + a_{2t}) + b_t y_{it}^*$, where $y_{it}^* = \varphi_t^{*-1}(r_{it}, z_{it}^d, u_{it})$ is the latent output under the true structure. Since this is a one-to-one transformation of the latent variable y_{it}^* that does not alter the relationship between observables (r_{it}, z_{it}^d) and the normalized residual u_{it}, the conditional distribution $F_{r|z^d,m,w}$ remains unchanged. Consequently, the joint distribution of all observed variables $\{r_{is}, m_{is}, v_{is}\}_{s=t-\upsilon-2}^{t}$ is identical under both structures.

Proof of (ii). Suppose $(\tilde{\varphi}_t^{-1}, \tilde{f}_t, \tilde{\mathbb{M}}_t^{-1})$ satisfies (C.10) and generates the same population joint distribution as the true structure. We show that there exist constants $(a_{1t}, a_{2t}, b_t) \in \mathbb{R}^2 \times \mathbb{R}_{++}$ such that $(\tilde{\varphi}_t^{-1}, \tilde{f}_t, \tilde{\mathbb{M}}_t^{-1}) \in \mathscr{E}_t$.

Step 1: Residual equivalence. Since both structures generate the same conditional distribution $F_{r|z^d,m,w}$, and both φ_t^* and $\tilde{\varphi}_t$ are strictly monotonic in the demand shock (Assumption 1), the residuals are identified up to monotone transformation. Under the normalization that u_{it} follows a standard uniform distribution, we have $\tilde{u}_{it} = u_{it}$.

Step 2: Affine relationship between latent outputs. Define the latent outputs $y_{it}^* := \varphi_t^{*-1}(r_{it}, z_{it}^d, u_{it})$ and $\tilde{y}_{it} := \tilde{\varphi}_t^{-1}(r_{it}, z_{it}^d, u_{it})$. Both structures satisfy:

$$y_{it}^* = f_t^*(x_{it}, z_{it}^s) + \mathbb{M}_t^{*-1}(m_{it}, w_{it}, u_{it}), \tag{C.11}$$

$$\tilde{y}_{it} = \tilde{f}_t(x_{it}, z_{it}^s) + \tilde{\mathbb{M}}_t^{-1}(m_{it}, w_{it}, u_{it}). \tag{C.12}$$

Define the difference $\Delta(r, z^d, u) := \tilde{\varphi}_t^{-1}(r, z^d, u) - \varphi_t^{*-1}(r, z^d, u) = \tilde{y} - y^*$. Subtracting (C.11) from (C.12):

$$\Delta(r_{it}, z_{it}^d, u_{it}) = \underbrace{\left[\tilde{f}_t(x_{it}, z_{it}^s) - f_t^*(x_{it}, z_{it}^s)\right]}_{=:\Delta_f(x_{it}, z_{it}^s)} + \underbrace{\left[\tilde{\mathbb{M}}_t^{-1}(m_{it}, w_{it}, u_{it}) - \mathbb{M}_t^{*-1}(m_{it}, w_{it}, u_{it})\right]}_{=:\Delta_{\mathbb{M}}(m_{it}, w_{it}, u_{it})}. \tag{C.13}$$

Note that Δ_f depends only on (x, z^s) while $\Delta_{\mathbb{M}}$ depends only on (m, w, u).

Under Assumption 3, there exists variation in (k_{it}, l_{it}) that is independent of (u_{it}, z_{it}^d) conditional on (m_{it}, z_{it}^s). Fix (m, z^s, z^d, u) and vary (k, l). Since $x = (m, k, l)$ and $w = (k, l, z^s, z^d)$, the left-hand side $\Delta(r, z^d, u)$ changes only through changes in r induced by changes in y^* (and hence $\tilde{y}$). However, by (C.13), this change must equal the change in $\Delta_f(m, k, l, z^s)$, since $\Delta_{\mathbb{M}}(m, w, u)$ varies only through its dependence on (k, l) via w.

Taking the partial derivative of (C.13) with respect to k (holding m, l, z^s, z^d, u fixed):

$$\frac{\partial \Delta_f}{\partial k} + \frac{\partial \Delta_{\mathbb{M}}}{\partial k} = \frac{\partial \Delta}{\partial r} \cdot \frac{\partial r}{\partial k}.$$

Since Δ_f depends on $(x, z^s) = (m, k, l, z^s)$ and $\Delta_{\mathbb{M}}$ depends on $(m, w, u) = (m, k, l, z^s, z^d, u)$, both terms on the left may depend on k. However, Δ_f does not depend on (z^d, u) while $\Delta_{\mathbb{M}}$ does not depend on x except through the overlap in (k, l) via w. By varying (z^d, u) while holding (x, z^s) fixed, we deduce that $\partial \Delta_{\mathbb{M}} / \partial k$ cannot depend on (x, z^s) beyond its explicit dependence on (k, l) through w.

Since both structures generate the same joint distribution by assumption, there exists a continuous strictly increasing function g such that $\tilde{y} = g(y^*)$. Using (C.11)–(C.12), this implies

$$g\big(f_t^*(x,z^s) + \mathbb{M}_t^{*-1}(m,w,u)\big) = \tilde{f}_t(x,z^s) + \tilde{\mathbb{M}}_t^{-1}(m,w,u).$$

This is a Pexider functional equation: a continuous function of a sum equals a sum of functions of the separate arguments. Under Assumption 3, the ranges of $f_t^*(\cdot,z^s)$ and $\mathbb{M}_t^{*-1}(\cdot,w,u)$ are non-degenerate intervals when the other argument is held fixed: the supply-side instruments z^s enter f_t^* but not $\mathbb{M}_t^{*-1}$, while the demand quantile u enters $\mathbb{M}_t^{*-1}$ but not f_t^*, so each argument can be varied independently of the other. To see this, set $A = f_t^*(x,z^s)$ and $B = \mathbb{M}_t^{*-1}(m,w,u)$, so the equation reads $g(A+B) = h_1(A) + h_2(B)$ where $h_1 := \tilde{f}_t$ and $h_2 := \tilde{\mathbb{M}}_t^{-1}$, viewed as functions of A and B respectively. By the Pexider theorem (Aczél and Dhombres, 1989, Theorem 9), the unique continuous solution is $g(s) = \beta s + c$, $h_1(A) = \beta A + \alpha_1$, $h_2(B) = \beta B + \alpha_2$ with $c = \alpha_1 + \alpha_2$. In particular, the common slope β applies to both components:

$$\tilde{f}_t(x,z^s) = \alpha_1 + \beta f_t^*(x,z^s), \qquad \tilde{\mathbb{M}}_t^{-1}(m,w,u) = \alpha_2 + \beta \mathbb{M}_t^{*-1}(m,w,u).$$

Step 3: Common scale parameter. The Pexider argument in Step 2 already establishes that $\beta_1 = \beta_2 =: \beta$, so we set $b_t := \beta$. Since both φ_t^* and $\tilde{\varphi}_t$ are strictly increasing in y (higher output yields higher revenue), we have $b_t > 0$. Setting $a_{1t} := \alpha_1$ and $a_{2t} := \alpha_2$, we obtain:

$$\tilde{y}_{it} = (a_{1t} + a_{2t}) + b_t y_{it}^*,$$

which implies $\tilde{\varphi}_t^{-1}(r,z^d,u) = (a_{1t} + a_{2t}) + b_t \varphi_t^{*-1}(r,z^d,u)$.

Hence $(\tilde{\varphi}_t^{-1}, \tilde{f}_t, \tilde{\mathbb{M}}_t^{-1}) \in \mathscr{E}_t$.

Proof of (iii). Part (i) establishes that every element of $\mathscr{E}_t$ is observationally equivalent to the true structure. Part (ii) establishes that every observationally equivalent structure belongs to $\mathscr{E}_t$. Together, $\mathscr{E}_t$ is exactly the set of structures consistent with the population distribution, i.e., $\mathscr{E}_t$ is exactly identified. □

Remark 3. The three-dimensional indeterminacy (a_{1t}, a_{2t}, b_t) arises because the unobserved output level y_{it} enters the revenue function through the unknown nonlinear function φ_t, which has no natural scale or location. The parameter $b_t \in \mathbb{R}_{++}$ governs the common scale of all three structural functions, while a_{1t} and a_{2t} independently shift the location of f_t and $\mathbb{M}_t^{-1}$, respectively. The constraint that b_t is common across f_t and $\mathbb{M}_t^{-1}$ follows from the additive separability of the model.

Remark 4. Under Assumption 4, the normalization conditions

$$f_t(m^*_{t0}, k^*_t, l^*_t, z^{s*}_t) = 0, \quad \mathbb{M}^{-1}_t(m^*_{t0}, w^*_t, u^*_t) = 0, \quad \mathbb{M}^{-1}_t(m^*_{t1}, w^*_t, u^*_t) = 1$$

uniquely pin down (a_{1t}, a_{2t}, b_t), thereby selecting a unique representative from $\mathscr{E}_t$ and achieving point identification of the structural functions.

D Identification of Consumer Preference and Counterfactual Analysis under CES Assumption with Heterogeneity

Under the constant elastic inverse demand function (12), the *structural* budget-share function $\mathfrak{s}^*_t(\cdot)$ takes the linear semi-parametric form:

$$\mathfrak{s}^*_t(y_{it} - q_t(\mathbf{y}_t, \mathbf{u}_t), u_{it}) = \rho(u_{it})(y_{it} - q_t(\mathbf{y}_t, \mathbf{u}_t)) + \delta(u_{it}), \tag{D.14}$$

where $\delta(u_{it})$ captures structural heterogeneity, and $\alpha_t(u_{it}) = \Phi_t - \rho(u_{it})q_t + \delta(u_{it})$ in (12). At the baseline state $(\mathbf{y}_t, \mathbf{u}_t)$, imposing normalization $q_t = 0$ (Assumption 11(c)) allows the reduced-form budget-share function to identify structural parameters directly from revenue data: $r_{it} - \Phi_t = \mathfrak{s}_t(y_{it}, u_{it}) = \rho(u_{it})y_{it} + \delta(u_{it})$. Estimating this equation identifies the elasticity parameter $\rho(u_{it})$ and heterogeneity $\delta(u_{it})$.

For a counterfactual state $(\tilde{\mathbf{y}}_t, \tilde{\mathbf{u}}_t)$, substituting the identified linear form into the market share constraint (52) yields: $1 = \sum_{i=1}^{N_t} \exp(\rho(\tilde{u}_{it})(\tilde{y}_{it} - \Delta q_t) + \delta(\tilde{u}_{it}))$. Since $\rho(\cdot) > 0$, the right-hand side is strictly decreasing in Δq_t, uniquely identifying the quantity index change.

Applying Proposition 7(b) with $\mathfrak{s}_t(\zeta, \tilde{u}_{it}) = \rho(\tilde{u}_{it})\zeta + \delta(\tilde{u}_{it})$ yields the utility function:

$$\begin{aligned} \ln U_t(\tilde{\mathbf{y}}_t, \tilde{\mathbf{u}}_t) &= \Delta q_t + \sum_{i=1}^{N_t} \int_{b_i}^{\tilde{y}_{it} - \Delta q_t} \exp(\rho(\tilde{u}_{it})\zeta + \delta(\tilde{u}_{it}))\, d\zeta \\ &= \Delta q_t + \sum_{i=1}^{N_t} \frac{1}{\rho(\tilde{u}_{it})} \exp(\rho(\tilde{u}_{it})(\tilde{y}_{it} - \Delta q_t) + \delta(\tilde{u}_{it})) + \text{const.} \end{aligned} \tag{D.15}$$

This expression represents utility as a weighted sum of structural market shares. In the homogeneous case ($\rho(\tilde{u}_{it}) = \rho$ and $\delta(\tilde{u}_{it}) = \delta$), the summation simplifies to $1/\rho$ due to the adding-up constraint, recovering the standard CES result $\ln U_t = \Delta q_t + \text{const} = \frac{1}{\rho} \ln \sum_{i=1}^{N_t} \exp(\rho \tilde{y}_{it}) + \text{const.}$

We may also apply the counterfactual framework to the CES specification (D.14). In this case, the derivative of the structural budget-share function with respect to output is constant, $\frac{\partial \mathfrak{s}^*_t}{\partial y} = \rho(u_{it})$. Substituting this derivative into the equilibrium conditions, we can express the

Monopolistic Competition Equilibrium (MCE, where $\mathbb{I}_{MCE}=1$) and the Marginal Cost Pricing Equilibrium (MCPE, where $\mathbb{I}_{MCE}=0$) in a unified system of equations in levels:

$$
\begin{aligned}
&\underbrace{\exp\left(\rho(u_{it})(y_{it}^{\star}-\Delta q_t^{\star})+\delta(u_{it})+\Phi_t+\mathbb{I}_{MCE}\ln\rho(u_{it})-y_{it}^{\star}\right)}_{\text{Marginal Revenue }(\mathbb{I}_{MCE}=1)\text{ or Price }(\mathbb{I}_{MCE}=0)}\\
&\quad=\underbrace{\frac{\exp\left(p_t^m+\chi_{it}\left(y_{it}^{\star}\right)\right)}{\exp(y_{it}^{*})}\frac{\partial\chi_{it}(y_{it}^{\star})}{\partial y_{it}}}_{\text{Marginal Cost}}\quad\text{for } i=1,\dots,N_t,\\
&\sum_{i=1}^{N_t}\exp\left(\rho(u_{it})(y_{it}^{\star}-\Delta q_t^{\star})+\delta(u_{it})\right)=1,
\end{aligned}
\tag{D.16}
$$

where $\mathbb{I}_{MCE}$ is an indicator variable. For the MCE, we set $\mathbb{I}_{MCE}=1$ to solve for $(\mathbf{y}^{\star},\Delta q^{\star})=(\mathbf{y}^m,\Delta q^m)$. For the MCPE, we set $\mathbb{I}_{MCE}=0$ to solve for $(\mathbf{y}^{\star},\Delta q^{\star})=(\mathbf{y}^c,\Delta q^c)$.

Using the utility function derived for the case of CES with heterogeneity, the consumer welfare cost of market power is calculated as:

$$
\begin{aligned}
\ln U_t^c-\ln U_t^m=&\left(\Delta q_t^c-\Delta q_t^m\right)\\
&+\sum_{i=1}^{N_t}\frac{1}{\rho(u_{it})}\left[\exp\left(\mathfrak{s}_t^{*}(y_{it}^c-\Delta q_t^c,u_{it})\right)-\exp\left(\mathfrak{s}_t^{*}(y_{it}^m-\Delta q_t^m,u_{it})\right)\right].
\end{aligned}
$$

In the homogeneous case where $\rho(u_{it})=\rho$, the summation term vanishes because the sum of structural market shares $\sum\exp(\mathfrak{s}_t^{*})$ equals 1 in both equilibria, simplifying the welfare loss to the difference in quantity indices: $\Delta q_t^c-\Delta q_t^m$.

E Alternative Demand Systems

In addition to the HSA demand system, Matsuyama and Ushchev (2017) further propose two additional families of homothetic demand systems:

1. the Homothetic demand system with Direct Implicit Additivity (HDIA), and

2. the Homothetic demand system with Indirect Implicit Additivity (HIIA).

Both systems share the same homotheticity structure but impose implicit additivity either in quantities (HDIA) or prices (HIIA). Below we summarize their reduced-form representations and show how the two quantity (or price) indices are identified.

We consider a shift from the baseline state $(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t)$ to a new state $(\tilde{\mathbf{y}}_t,\mathbf{z}_t^d,\mathbf{u}_t)$ where demand shifters $\left(\mathbf{z}_t^d,\mathbf{u}_t\right)$ remain unchanged.

E.1 HDIA demand system

The structural budget share function of the HDIA has the form:

$$\ln\left(\frac{\exp\left(r_{it}\right)}{\sum_{j=1}^{N_t}\exp\left(r_{jt}\right)}\right)=\upsilon_t^*\left(y_{it}-a_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right),z_{it}^d,u_{it}\right)-b_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right)$$

There are two aggregators $a_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right)$ and $b_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right)$. Aggregator $a_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right) = \ln U(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t)$ is the log of a utility function that generates the demand system and implicitly determined by the additive restriction:

$$1=\sum_{i=1}^{N_t}\Upsilon_t^*\left(y_{it}-a_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right),z_{it}^d,u_{it}\right) \tag{E.17}$$

where $\partial\Upsilon_t^*\left(x,z_{it}^d,u_{it}\right)/\partial x=\exp\left(\upsilon_t^*\left(x,z_{it}^d,u_{it}\right)\right)$. Another aggregator $b_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right)$ is defined by

$$b_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right):=\ln\left[\sum_{i=1}^{N_t}\exp\left(\upsilon_t^*\left(y_{it}-a_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right),z_{it}^d,u_{it}\right)\right)\right]$$

The reduced form revenue function is related to the structural budget share function as:

$$\begin{aligned}r_{it}&=\Phi_t+\upsilon_t^*\left(y_{it}-a_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right),z_{it}^d,u_{it}\right)-b_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right)\\&=:\varphi_t\left(y_{it},z_{it}^d,u_{it}\right).\end{aligned}$$

We define the reduced form budget share function as

$$\upsilon_t\left(y_{it},z_{it}^d,u_{it}\right):=\varphi_t\left(y_{it},z_{it}^d,u_{it}\right)-\Phi_t,$$

which is related to the structural form as

$$\upsilon_t\left(y_{it},z_{it}^d,u_{it}\right)=\upsilon_t^*\left(y_{it}-a_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right),z_{it}^d,u_{it}\right)-b_t\left(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t\right). \tag{E.18}$$

We consider a shift from the baseline state $(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t)$ to a new state $(\tilde{\mathbf{y}}_t,\mathbf{z}_t^d,\mathbf{u}_t)$ where demand shifters $\left(\mathbf{z}_t^d,\mathbf{u}_t\right)$ remain unchanged. Let $\Delta a_t\left(\tilde{\mathbf{y}}_t\right):=a_t(\tilde{\mathbf{y}}_t,\mathbf{z}_t^d,\mathbf{u}_t)-a_t(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t)$ and $\Delta b_t\left(\tilde{\mathbf{y}}_t\right):=b_t(\tilde{\mathbf{y}}_t,\mathbf{z}_t^d,\mathbf{u}_t)-b_t(\mathbf{y}_t,\mathbf{z}_t^d,\mathbf{u}_t)$.

The reduced form budget share functions with the changes in the quantity indices gives the value of the structural budget share function at the new state as follows:

$$
\begin{aligned}
&\upsilon_t\left(\tilde{y}_{it}-\Delta a_t(\tilde{\mathbf{y}}_t), z_{it}^d, u_{it}\right)-\Delta b_t(\tilde{\mathbf{y}}_t) \\
&=\upsilon_t^*\left(\tilde{y}_{it}-a_t\left(\tilde{\mathbf{y}}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, \tilde{u}\right)-b_t\left(\tilde{\mathbf{y}}_t, \mathbf{z}_t^d, \mathbf{u}_t\right).
\end{aligned}
$$

Therefore, as in the HSA case, we can conduct a counterfactual analysis using the reduced form budget share functions and the changes in the quantity indices. The following lemma shows identification of $\Delta a_t(\tilde{\mathbf{y}}_t)$ and $\Delta b_t(\tilde{\mathbf{y}}_t)$.

Lemma E.1. *$\Delta a_t(\tilde{\mathbf{y}}_t)$ is uniquely identified by*

$$
\sum_{i=1}^{N_t}\int_{y_{it}}^{\tilde{y}_{it}-\Delta a_t(\tilde{\mathbf{y}}_t)} \exp\left(\upsilon_t\left(\zeta, z_{it}^d, u_{it}\right)\right) d\zeta = 0.
$$

$\Delta b_t(\tilde{\mathbf{y}}_t)$ is identified by

$$
\Delta b_t(\tilde{\mathbf{y}}_t)=\ln\left[\sum_{i=1}^{N_t}\exp\left\{\upsilon_t\left(\tilde{y}_{it}-\Delta a_t(\tilde{\mathbf{y}}_t), z_{it}^d, u_{it}\right)\right\}\right].
$$

Proof. Let $B_t := \exp\left(b_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)\right)$. From (E.18), we have

$$
\begin{aligned}
&\int_{y_{it}}^{\tilde{y}_{it}-\Delta a_t(\tilde{\mathbf{y}}_t)} \exp\left(\upsilon_t\left(\zeta, z_{it}^d, u_{it}\right)\right) d\zeta \\
=&\frac{1}{B_t}\int_{y_{it}}^{\tilde{y}_{it}-\Delta a_t(\tilde{\mathbf{y}}_t)} \exp\left(\upsilon_t^*\left(\zeta-a_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right)\right) d\zeta \\
=&\frac{1}{B_t}\int_{y_{it}}^{\tilde{y}_{it}-\Delta a_t(\tilde{\mathbf{y}}_t)} \exp\left(\frac{\partial \Upsilon_t^*\left(\zeta-a_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right)}{\partial x}\right) d\zeta \\
=&\frac{1}{B_t}\left[\Upsilon_t^*\left(\tilde{y}_{it}-a_t\left(\tilde{\mathbf{y}}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right)-\Upsilon_t^*\left(y_{it}-a_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right)\right]
\end{aligned}
$$

Thus, from (E.17),

$$
\begin{aligned}
&\sum_{i=1}^{N_t}\int_{y_{it}}^{\tilde{y}_{it}-\Delta a_t(\tilde{\mathbf{y}}_t)} \exp\left(\upsilon_t\left(\zeta, z_{it}^d, u_{it}\right)\right) d\zeta \\
=&\frac{1}{B_t}\left[\sum_{i=1}^{N_t}\Upsilon_t^*\left(\tilde{y}_{it}-a_t\left(\tilde{\mathbf{y}}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right)-\sum_{i=1}^{N_t}\Upsilon_t^*\left(y_{it}-a_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right)\right] \\
=&\frac{1}{B_t}[1-1]=0.
\end{aligned}
$$

Because $\exp(\upsilon_t(\cdot)) > 0$, the left-hand side is strictly decreasing in Δa_t. Thus, the last equation uniquely identifies $\Delta a_t(\tilde{\mathbf{y}}_t)$.

From (E.17), we have

$$\begin{aligned}
\Delta b_t(\tilde{\mathbf{y}}_t) &= \ln\left[\sum_{i=1}^{N_t} \exp\left(\upsilon_t^*\left(\tilde{y}_{it} - a_t\left(\tilde{\mathbf{y}}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right)\right)\right] - b_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right) \\
&= \ln\left[\sum_{i=1}^{N_t} \exp\left\{\upsilon_t\left(\tilde{y}_{it} - \Delta a_t(\tilde{\mathbf{y}}_t), z_{it}^d, u_{it}\right) + b_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)\right\}\right] - b_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right) \\
&= \ln\left[\exp\left(b_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)\right) \sum_{i=1}^{N_t} \exp\left\{\upsilon_t\left(\tilde{y}_{it} - \Delta a_t(\tilde{\mathbf{y}}_t), z_{it}^d, u_{it}\right)\right\}\right] - b_t\left(\mathbf{y}_t, \mathbf{z}_t^d, \mathbf{u}_t\right) \\
&= \ln\left[\sum_{i=1}^{N_t} \exp\left\{\upsilon_t\left(\tilde{y}_{it} - \Delta a_t(\tilde{\mathbf{y}}_t), z_{it}^d, u_{it}\right)\right\}\right]
\end{aligned}$$

□

E.2 HIIA demand system

We consider the direct demand system instead of the inverse demand system. The structural budget share function of the HIIA has two aggregators $a_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)$ and $b_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)$:

$$\ln\left(\frac{\exp(r_{it})}{\sum_{j=1}^{N_t} \exp(r_{jt})}\right) = \upsilon_t^*\left(p_{it} - a_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right) - b_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)$$

where $a_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right) = \ln P_t(\mathbf{p}_t)$ is the log of the ideal price index and implicitly restricted by the following additive restriction:

$$1 = \sum_{i=1}^{N_t} \Upsilon_t^*\left(p_{it} - a_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right) \tag{E.19}$$

where $\partial \Upsilon_t^*\left(x, z_{it}^d, u_{it}\right)/\partial x = \exp\left(\upsilon_t^*\left(x, z_{it}^d, u_{it}\right)\right)$. Another price aggregator$b_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right)$ is defined by

$$b_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right) := \ln\left[\sum_{i=1}^{N_t} \exp\left(\upsilon_t^*\left(p_{it} - a_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right)\right)\right].$$

We obtain the reduced form revenue function as a function of price:

$$\chi_t\left(p_{it}, z_{it}^d, u_{it}\right) := \psi_t^{-1}\left(p_{it}, z_{it}^d, u_{it}\right) + p_{it}$$

where $\psi_t^{-1}\left(p_{it}, z_{it}^d, u_{it}\right)$ is the inverse of the reduced form inverse demand $\psi_t\left(y_{it}, z_{it}^d, u_{it}\right)$ with respect to y_{it}.

The reduced form revenue function is related to the structural budget share function as follows:

$$\begin{aligned} r_{it} &= \Phi_t + \upsilon_t^*\left(p_{it} - a_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right) - b_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right) \\ &=: \chi_t\left(p_{it}, z_{it}^d, u_{it}\right). \end{aligned}$$

We define the reduced form budget share function as

$$\upsilon_t\left(p_{it}, z_{it}^d, u_{it}\right) := \chi_t\left(p_{it}, z_{it}^d, u_{it}\right) - \Phi_t,$$

which is related to the structural form as

$$\upsilon_t\left(p_{it}, z_{it}^d, u_{it}\right) = \upsilon_t^*\left(p_{it} - a_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right) - b_t\left(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t\right). \tag{E.20}$$

We consider a shift from the baseline state $(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ to a new state $(\tilde{\mathbf{p}}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ where $\left(\mathbf{z}_t^d, \mathbf{u}_t\right)$ remain unchanged. Let $\Delta a_t\left(\tilde{\mathbf{p}}_t\right) := a_t(\tilde{\mathbf{p}}_t, \mathbf{z}_t^d, \mathbf{u}_t) - a_t(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ and $\Delta b_t\left(\tilde{\mathbf{p}}_t\right) := b_t(\tilde{\mathbf{p}}_t, \mathbf{z}_t^d, \mathbf{u}_t) - b_t(\mathbf{p}_t, \mathbf{z}_t^d, \mathbf{u}_t)$ be the changes in the price aggregators.

The reduced form budget share functions with those price aggregators gives the value of the structural budget share function at the new state:

$$\begin{aligned} &\upsilon_t\left(\tilde{p}_{it} - \Delta a_t\left(\tilde{\mathbf{p}}_t\right), z_{it}^d, u_{it}\right) - \Delta b_t\left(\tilde{\mathbf{p}}_t\right) \\ &= \upsilon_t^*\left(\tilde{p}_{it} - a_t\left(\tilde{\mathbf{p}}_t, \mathbf{z}_t^d, \mathbf{u}_t\right), z_{it}^d, u_{it}\right) - b_t\left(\tilde{\mathbf{p}}_t, \mathbf{z}_t^d, \mathbf{u}_t\right). \end{aligned}$$

The following lemma shows identification of $\Delta a_t\left(\tilde{\mathbf{p}}_t\right)$ and $\Delta b_t\left(\tilde{\mathbf{p}}_t\right)$.

Lemma E.2. $\Delta a_t\left(\tilde{\mathbf{p}}_t\right)$ *is uniquely identified by*

$$\sum_{i=1}^{N_t} \int_{p_{it}}^{\tilde{p}_{it} - \Delta a_t(\tilde{\mathbf{p}}_t)} \exp\left(\upsilon_t\left(\zeta, z_{it}^d, u_{it}\right)\right) d\zeta = 0.$$

$\Delta b_t\left(\tilde{\mathbf{p}}_t\right)$ *is identified by*

$$\Delta b_t\left(\tilde{\mathbf{p}}_t\right) = \ln\left[\sum_{i=1}^{N_t} \exp\left\{\upsilon_t\left(\tilde{p}_{it} - \Delta a_t\left(\tilde{\mathbf{p}}_t\right), z_{it}^d, u_{it}\right)\right\}\right].$$

Proof. The proof is identical to that of Lemma E.1, replacing quantities with prices. ∎

F Alternative Settings

F.1 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. For given k_{it}, a firm's per-period profit (excluding capital costs) is given by

$$\exp(\varphi_t(f_t(m_{it}, l_{it}, k_{it}, z^s_{it}) + \omega_{it}, z^d_{it}, u_{it})) - \exp(p^m_t + m_{it}) - \exp(p^l_t + l_{it}) - C_t(l_{it}, l_{it-1}),$$

where p^l_t is the wage and $C_t(l_{it}, l_{it-1})$ is the adjustment costs. From the above per-period profit, the solution to a firm's dynamic 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 $s_{it} := (k_{it}, z^s_{it}, z^d_{it})$. 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 (7).

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

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

The first step identification is

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

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

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

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

Since η_{it} is independent of $v_{it} := (k_{it}, l_{it-1,}, s_{it}, u_{it}, m_{it-1}, k_{it-1}, l_{it-2}, s_{it-1}, u_{it-1}, z_{it-1}^h)$, 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 the conditional expectation $\mathbb{M}_t^{-1}(m_{it}, l_{it}, s_{it}, u_{it}) = E[\omega_{it}|m_{it}, l_{it}, s_{it}, u_{it}]$.

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 (33) and (34). 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 and chosen at the same timing as m_{it}, 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 $\phi_t(m_{it}, l_{it}, k_{it}, z_{it}, u_{it})$.

F.2 Endogenous Firm Characteristics

Firm characteristics (z_{it}^s, z_{it}^d) may correlate with ζ_{it} and η_{it}. In step 1, we can use $(z_{it-\upsilon-1}^s, z_{it-\upsilon-1}^d)$ in place of $(z_{it-\upsilon}^s, z_{it-\upsilon}^d)$ as instrument variables to construct the moment condition similar to (26).

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 $\varsigma_{it} = (\varsigma_{it}^s, \varsigma_{it}^d)$, unknown functions Γ_{kt}, and unobservable scalars κ_{it}^k such that

$$z_{it}^k = \Gamma_{kt}(\varsigma_{it}, \varpi_{it}, \kappa_{it}^k), (k = s, d)$$

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

Assumption F.1. *For $k = s, d$, (i) $\varsigma^k_{it} \perp\!\!\!\perp (\eta_{it}, \kappa^k_{it})$ (ii) κ^k_{it} is a scalar and $\kappa^k_{it} \perp\!\!\!\perp (\varsigma_{it}, \varpi_{it})$.(iii) Γ_{kt} is strictly increasing in κ^k_{it} (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}.

Lemma F.3. *(Imbens and Newey, 2009,Theorem 1) $\eta_{it} \perp\!\!\!\perp v_{it}|\xi_{it}$.*

Proof. The following proof follows Imbens and Newey (2009). From the monotonicity of Γ_{kt} in Assumption F.1 (iii), we can define the inverse function of Γ_{kt}such that $\kappa^k_{it} = \Gamma^{-1}_{kt}\left(\varsigma_{it}, \varpi_{it}, z^k_{it}\right)$.For given $(z^k_t, \varsigma_t, \varpi_t)$

$$\begin{aligned}
F_{z^k_t|\varsigma_t,\varpi_t}(z^k_t|\varsigma_t, \varpi_t) &= \Pr\left(\Gamma_{kt}(\varsigma_{it}, \varpi_{it}, \kappa^k_{it}) \le z^k_t|\varsigma_{it} = \varsigma_t, \varpi_{it} = \varpi_t\right) \\
&= \Pr\left(\kappa^k_{it} \le \Gamma^{-1}_{kt}\left(\varsigma_t, \varpi_t, z^k_t\right)|\varsigma_{it} = \varsigma_t, \varpi_{it} = \varpi_t\right) \\
&= F_{\kappa^k_t}\left(\Gamma^{-1}_{kt}\left(\varsigma_t, \varpi_t, z^k_t\right)\right). \text{ (from } \kappa^k_{it} \perp\!\!\!\perp (\varsigma_{it}, \varpi_{it}))
\end{aligned}$$

Therefore, we have

$$\xi^k_{it} = F_{z^k_t|\varsigma_t,\varpi_t}(z^k_{it}|\varsigma_{it}, \varpi_{it}) = F_{\kappa^k_t}\left(\Gamma^{-1}_{kt}\left(\varsigma_{it}, \varpi_{it}, z^k_{it}\right)\right) = F_{\kappa^k_t}\left(\kappa^k_{it}\right).$$

Consider an arbitrary point (ξ_t, η_t) on the support of (ξ_{it}, η_{it}). Let $(\kappa^s_t, \kappa^d_t) = \left(F^{-1}_{\kappa^s_t}(\xi^s_t), F^{-1}_{\kappa^d_t}(\xi^d_t)\right)$. Since $F_{\kappa^k_t}$ is strictly increasing, the conditional expectations given $\xi_{it} = \xi_t$ are identical to those given $(\kappa^s_{it}, \kappa^d_{it}) = (\kappa^s_t, \kappa^d_t)$. For any bounded function $a(v_{it})$ of v_{it}, the independence of $(\varsigma_{it}, \varpi_{it})$ and $\left(\kappa^s_{it}, \kappa^d_{it}, \eta_{it}\right)$ implies

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

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

Let $\mathfrak{X} := \{\xi_{it}\}_{i=1}^{n}$ be the set of ξ_{it} constructed for each observation. We update Assumption 6 so that for every $\xi_{it} \in \mathfrak{X}$, the conditional distribution $G_{\eta_t|\xi_t}(\cdot|\xi_{it})$ of η_{it} on ξ_{it}can satisfy the requirement for $G_{\eta_t}(\cdot)$ in Assumption 6.

Assumption F.2. *(a) For every $\xi_{it} \in \mathfrak{X}$, the conditional distribution $G_{\eta_t|\xi}(\cdot|\xi_{it})$ of η_{it} conditional on ξ_{it} is absolutely continuous with a density function $g_{\eta_t|\xi_t}(\cdot|\xi_{it})$ that is continuous on its support. (b) $E[\eta_{it}|m_{it-1},w_{it-1},u_{it-1},z_{it-1}^h]=0$. (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 $\mathscr{A}_{q_{t-1}} := \{(m_{it-1},w_{it-1},u_{it-1},z_{it-1}^h)\in\mathscr{M}\times\mathscr{W}\times[0,1]\times\mathscr{Z}_h :$ $\partial G_{m_t|v_t,\xi_t}(m_{it}|v_{it},\xi_{it})/\partial q_{it-1}\neq 0$ for all $(m_{it},w_{it},u_t)\in\mathscr{M}\times\mathscr{W}\times[0,1]$ and for every$\xi_{it}\in\mathfrak{X}\}$ is nonempty for some $q_{it-1}\in\{m_{it-1},k_{it-1},l_{it-1},z_{it-1}^s,z_{it-1}^d,u_{it-1},z_{it-1}^h\}$.*

From Lemma F.3, the conditional distribution of m_{it} given (v_{it},ξ_{it}) satisfies

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

Taking the derivatives of both sides with respect to $q_{it}\in\{m_{it},k_{it},l_{it},z_{it}^s,z_{it}^d,u_{it}\}$ and $q_{it-1}\in\{m_{it-1},k_{it-1},l_{it-1},z_{it-1}^s,z_{it-1}^d,z_{it-1}^h,u_{it-1}\}$, we obtain

$$\frac{\partial G_{m_t|v_t,\xi_t}(m_{it}|v_{it},\xi_{it})}{\partial q_{it}} = \frac{\partial\mathbb{M}_t^{-1}(m_{it},w_{it},u_{it})}{\partial q_{it}}g_{\eta_t|\xi_t}(\eta_{it}|\xi_{it}), \tag{F.21}$$

$$\frac{\partial G_{m_t|v_t,\xi_t}(m_{it}|v_{it},\xi_{it})}{\partial q_{it-1}} = -\frac{\partial\bar{h}_t\left(m_{it-1},w_{it-1},u_{it-1},z_{it-1}^h\right)}{\partial q_{it-1}}g_{\eta_t|\xi_t}(\eta_{it}|\xi_{it}), \tag{F.22}$$

where $\eta_{it}=\mathbb{M}_t^{-1}(m_{it},w_{it},u_{it})-\bar{h}_t\left(m_{it-1},w_{it-1},u_{it-1},z_{it-1}^h\right)$.Using Assumption F.2(f), we can choose $q_{it-1}\in\{m_{it-1},k_{it-1},l_{it-1},z_{it-1}^s,z_{it-1}^d,z_{it-1}^h,u_{it-1}\}$ and $(\tilde{m}_{it-1},\tilde{w}_{it-1},\tilde{u}_{it-1},\tilde{z}_{it-1}^h)\in\mathscr{A}_{q_{t-1}}$ such that $\partial G_{m_t|v_t,\xi_t}\left(m_t|k_t,l_t,z_t,u_t,\tilde{m}_{it-1},\tilde{w}_{it-1},\tilde{u}_{it-1},\tilde{z}_{it-1}^h,\xi_{it}\right)/\partial q_{t-1}\neq 0$ for every $\xi_{it}\in\mathfrak{X}$ and 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 (F.21) by (F.22), we derive

$$\frac{\partial \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})}{\partial q_{it}} = -\frac{\partial \bar{h}_t\left(\tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}\right)}{\partial q_{it-1}} \times \frac{\partial G_{m_t|v_t,\xi_t}\left(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}, \xi_{it}\right)/\partial q_t}{\partial G_{m_t|v_t,\xi_t}\left(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \tilde{w}_{it-1}, \tilde{u}_{it-1}, \tilde{z}^h_{it-1}, \xi_{it}\right)/\partial q_{t-1}}. \tag{F.23}$$

We have obtained equation (F.23) similar to (30) again. Therefore, following the same steps in the proof for Proposition 2 by replacing $G_{m_t|v_t}(\cdot|\cdot)$ with $G_{m_t|v_t,\xi_t}(\cdot|\cdot, \xi_{it})$, we can identify $\mathbb{M}_t^{-1}(\cdot)$ up to scale and location. From $\omega_{it} = \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})$ and $E[\eta_{it}|m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}] = 0$, we can identify $\bar{h}_t(m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}) = E\left[\omega_{it}|m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1}\right]$ and $\eta_{it} = \omega_{it} - \bar{h}_t(m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1})$. Thus, we can identify the distribution of η_{it}, $G_{\eta_t}(\cdot)$.

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.

F.3 Discrete Firm Characteristics

F.3.1 Exogenous Characteristics

Suppose z^s_{it}, z^d_{it} and z^h_{it} are discrete variables and have finite support $\mathscr{Z}_s := \{z^1_s, ..., z^{J_s}_s\}$, $\mathscr{Z}_d := \{z^1_d, ..., z^{J_d}_d\}$ and $\mathscr{Z}_h := \{z^1_h, ..., z^{J_h}_h\}$. In Step 1, 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^s_{it} and z^d_{it}.

Assumption F.3. *(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 the continuous arguments (m, k, l, u) on $\mathscr{M} \times \mathscr{K} \times \mathscr{L} \times [0, 1]$ for every fixed $(z^s, z^d) \in \mathscr{Z}_s \times \mathscr{Z}_d$. (d) $(\zeta_{it}, ..., \zeta_{it-\upsilon})$ are independent from η_{it}.*

The following assumption modifies Assumption 6 for discrete z^s_{it} and z^d_{it}, incorporating the relaxed support condition.

Assumption F.4. *(a) The distribution $G_{\eta_t}(\cdot)$ of η is absolutely continuous with a density function $g_{\eta_t}(\cdot)$ that is continuous on its support. (b) η_{it} is independent of $v_{it} := (w_{it}, u_{it}, m_{it-1}, w_{it-1}, u_{it-1}, z^h_{it-1})'$ with $E[\eta_{it}|v_{it}] = 0$. (c) Let $\mathbf{x}_{it} := (m_{it}, k_{it}, l_{it}, u_{it})$. For every fixed tuple of discrete variables $\mathbf{z} := (z^s_{it}, z^d_{it}, z^s_{it-1}, z^d_{it-1}, z^h_{it-1})$, the conditional support of the* *continuous variables* $(\mathbf{x}_{it}, \mathbf{x}_{it-1})$*, denoted* $\mathscr{X}(\mathbf{z})$*, is an open, connected subset of Euclidean space. Furthermore, the normalization points defined in the proof lie in the interior of these supports. (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* $\mathscr{A}_{q_{t-1}}$ *(defined analogously to Assumption 6(f) for the discrete case) is nonempty for some* $q_{it-1} \in \{m_{it-1}, k_{it-1}, l_{it-1}, u_{it-1}\}$*. (g) For each* $(m_{t-1}, w_{t-1}, u_{t-1}, z^h_{t-1}) \in Proj_v(\mathscr{M} \times \mathscr{V})$*, it is possible to find* $(\tilde{m}_t, \tilde{w}_t, \tilde{u}_t)$ *in the support such that* $\partial G_{m_t|v_t}(\tilde{m}_t|\tilde{w}_t, \tilde{u}_t, m_{t-1}, w_{t-1}, u_{t-1}, z^h_{t-1})/\partial m_t > 0$*.*

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

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

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

$$\mathbb{M}_t^{-1}(m^*_{t0}, k^*_t, l^*_t, z^s_t, z^d_t, u^*_t) = c_0(z^s_t, z^d_t),\ \mathbb{M}_t^{-1}(m^*_{t1}, k^*_t, l^*_t, z^s_t, z^d_t, u^*_t) = c_1(z^s_t, z^d_t), \tag{F.24}$$

$$\text{and } \bar{h}_t(m^*_{t-1}, k^*_{t-1}, l^*_{t-1}, z^s_{t-1}, z^d_{t-1}, u^*_{t-1}, z^h_{t-1}) = c_2(z^s_{t-1}, z^d_{t-1}, z^h_{t-1}), \tag{F.25}$$

where $\{c_0(z^s_t, z^d_t), c_1(z^s_t, z^d_t)\}$ and $\{c_2(z^s_{t-1}, z^d_{t-1}, z^h_{t-1})\}$ 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 implies $c_0(z^1_s, z^1_d) = 0$ and $c_1(z^1_s, z^1_d) = 1$.

Following the derivation in the proof of Proposition 2, the partial derivative of the control function with respect to continuous arguments $q_{it} \in \{m_{it}, k_{it}, l_{it}, u_{it}\}$ is identified locally as:

$$\frac{\partial \mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})}{\partial q_{it}} = \tilde{S}_{q_{t-1}} T_{q_t q_{t-1}}(m_{it}, w_{it}, u_{it}), \tag{F.26}$$

where $\tilde{S}_{q_{t-1}}$ is a scalar identified by integration over m at the normalization point (analogous to $S_{q_{t-1}}$ in the main proof) and

$$T_{q_t q_{t-1}}(m_{it}, w_{it}, u_{it}) := \frac{\partial G_{m_t|v_t}(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \dots)/\partial q_{it}}{\partial G_{m_t|v_t}(m_{it}|w_{it}, u_{it}, \tilde{m}_{it-1}, \dots)/\partial q_{it-1}}.$$

To recover the level of $\mathbb{M}_t^{-1}$ from these partial derivatives, we invoke the connected support condition in Assumption F.4(c). Let $\mathbf{x}^*_0 := (m^*_{t0}, k^*_t, l^*_t, u^*_t)$ be the reference point for the continuous variables. For any fixed discrete characteristics (z^s_t, z^d_t) and any target point $\mathbf{x} := (m_{it}, k_{it}, l_{it}, u_{it})$ in the continuous support, there exists a piecewise smooth path $\gamma : [0, 1] \to \mathscr{X}(\mathbf{z})$ such that $\gamma(0) = \mathbf{x}^*_0$ and $\gamma(1) = \mathbf{x}$. By the Fundamental Theorem of Line Integrals, we identify:

$$\mathbb{M}_t^{-1}(\mathbf{x}, z_t^s, z_t^d) = c_0(z_t^s, z_t^d) + \int_\gamma \nabla_{\mathbf{x}} \mathbb{M}_t^{-1}(\mathbf{z}(\tau), z_t^s, z_t^d) \cdot d\mathbf{z}(\tau), \tag{F.27}$$

where $\nabla_{\mathbf{x}} \mathbb{M}_t^{-1}$ is the vector of identified partial derivatives from (F.26). This generalizes the previous Manhattan-path integral to any path within the connected support.

Similarly, we identify the gradient of $\bar{h}_t$ with respect to its continuous arguments and recover its level via path integration:

$$\bar{h}_t(m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h) = c_2(z_{t-1}^s, z_{t-1}^d, z_{t-1}^h) + \Lambda_{\bar{h}_t}(m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h), \tag{F.28}$$

where $\Lambda_{\bar{h}_t}$ is the identified line integral of $\nabla \bar{h}_t$ starting from the normalization point $(m_{t-1}^*, \dots)$.

Thus, we have identified $\mathbb{M}_t^{-1}(\cdot)$ and $\bar{h}_t(\cdot)$ up to the unknown constants $\{c_0(z_t^s, z_t^d), c_2(z_{t-1}^s, z_{t-1}^d, z_{t-1}^h)\}$.

Define $\widetilde{H}_t$ as the conditional expectation of the identified components:

$$\widetilde{H}_t(z_t^s, z_t^d, z_{t-1}^s, z_{t-1}^d, z_{t-1}^h) := E[\Lambda_m(\mathbf{x}_{it}, z_t^s, z_t^d) - \Lambda_{\bar{h}_t}(\mathbf{x}_{it-1}, z_{t-1}^h) | z_t^s, z_t^d, z_{t-1}^s, z_{t-1}^d, z_{t-1}^h].$$

Using the mean independence condition $E[\eta_{it} | z_t^s, \dots] = 0$, we obtain the equation linking the constants:

$$0 = \widetilde{H}_t(z_t^s, z_t^d, z_{t-1}^s, z_{t-1}^d, z_{t-1}^h) + c_0(z_t^s, z_t^d) - c_2(z_{t-1}^s, z_{t-1}^d, z_{t-1}^h). \tag{F.29}$$

Evaluating (F.29) at the normalization point (z_s^1, z_d^1) where $c_0(z_s^1, z_d^1) = 0$, we identify:

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

Next, evaluating (F.29) at the base lag (z_s^1, z_d^1, z_h^1), we identify $c_0(z_t^s, z_t^d)$:

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

With all constants identified, $\mathbb{M}_t^{-1}(\cdot)$, $\bar{h}_t(\cdot)$, and the distribution of η_{it} are fully identified. $\square$

Proposition F.3. *Suppose that Assumptions 3, 4, F.3, F.4, 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 $\partial \varphi_t^{-1}(\bar{r}_{it}, z_{it}) / \partial r_{it}$ up to scale.*

Proof. From (33) and (35), the markup $\partial \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it}) / \partial r_{it}$ is identified as (37). From ϕ_t and (37), the markup function $\mu_t(m_{it}, w_{it}, u_{it})$ is also identified as a function of (m_{it}, w_{it}, u_{it}) as (38). Substituting (38) into (33), we identify the partial derivatives $\partial f_t(x_{it}, z_{it}^s) / \partial q_{it}$ for $q_{it}^s \in \{m_{it}, k_{it}, l_{it}\}$ as (39).

Define $c_f(z_t^s) := f_t(m_{t0}^*, k_t^*, l_t^*, z_t^s)$ for the normalization point (m_{t0}^*, k_t^*, l_t^*) defined in Assumption 4. To recover the level of the production function, we invoke the connected support condition for the inputs (m_t, k_t, l_t) for any fixed z_t^s. Let $\mathbf{x}_{prod} := (m_t, k_t, l_t)$ and $\mathbf{x}^*_{prod} := (m^*_{t0}, k^*_t, l^*_t)$. For any target inputs $\mathbf{x}_{prod}$ in the support, there exists a path $\gamma : [0,1] \to \mathscr{X}_{prod}(z_t^s)$ connecting $\mathbf{x}^*_{prod}$ to $\mathbf{x}_{prod}$. By the Fundamental Theorem of Line Integrals, we identify f_t up to the constant $c_f(z_t^s)$:

$$f_t(\mathbf{x}_{prod}, z_t^s) = c_f(z_t^s) + \Lambda_f(\mathbf{x}_{prod}, z_t^s)$$

where Λ_f is the identified path integral of the marginal products:

$$\begin{aligned}
\Lambda_f(\mathbf{x}_{prod}, z_t^s) &= \int_\gamma \nabla_{\mathbf{x}} f_t(\mathbf{z}(\tau), z_t^s) \cdot d\mathbf{z}(\tau) \\
&= \int_0^1 \left[\sum_{q \in \{m,k,l\}} \frac{\partial f_t(\gamma(\tau), z_t^s)}{\partial q_{it}} \frac{d\gamma_q(\tau)}{d\tau} \right] d\tau.
\end{aligned}$$

To identify the unknown constants $c_f(z_t^s)$, we use the normalization in Assumption 4. Recall that at the global normalization point (z_t^{s*}, z_t^{d*}):

$$\begin{aligned}
&\varphi_t^{-1}\left(\phi_t\left(m^*_{t0}, k^*_t, l^*_t, z_t^{s*}, z_t^{d*}, u^*_t\right), z_t^{d*}, u^*_t\right) \\
=& f_t(m^*_{t0}, k^*_t, l^*_t, z_t^{s*}) + \mathbb{M}_t^{-1}(m^*_{t0}, k^*_t, l^*_t, z_t^{s*}, z_t^{d*}, u^*_t) = 0.
\end{aligned} \tag{F.30}$$

For an arbitrary z_t^s, the relationship holds as:

$$\begin{aligned}
&\varphi_t^{-1}\left(\phi_t\left(m^*_{t0}, k^*_t, l^*_t, z_t^s, z_t^{d*}, u^*_t\right), z_t^{d*}, u^*_t\right) \\
=& f_t(m^*_{t0}, k^*_t, l^*_t, z_t^s) + \mathbb{M}_t^{-1}(m^*_{t0}, k^*_t, l^*_t, z_t^s, z_t^{d*}, u^*_t) \\
=& c_f(z_t^s) + \mathbb{M}_t^{-1}(m^*_{t0}, k^*_t, l^*_t, z_t^s, z_t^{d*}, u^*_t).
\end{aligned}$$

The LHS can be expressed by integrating the identified markup term $\partial \varphi_t^{-1}/\partial r_{it}$. Let $r_{base} = \phi_t(m^*_{t0}, k^*_t, l^*_t, z_t^{s*}, z_t^{d*}, u^*_t)$ and $r_{target} = \phi_t(m^*_{t0}, k^*_t, l^*_t, z_t^s, z_t^{d*}, u^*_t)$. Then:

$$\begin{aligned}
&\int_{r_{base}}^{r_{target}} \frac{\partial \varphi_t^{-1}(s, z_t^{d*}, u^*_t)}{\partial r_{it}} ds \\
=& \varphi_t^{-1}(r_{target}, z_t^{d*}, u^*_t) - \varphi_t^{-1}(r_{base}, z_t^{d*}, u^*_t) \\
=& \varphi_t^{-1}(r_{target}, z_t^{d*}, u^*_t) \quad \text{(from F.30)} \\
=& c_f(z_t^s) + \mathbb{M}_t^{-1}(m^*_{t0}, k^*_t, l^*_t, z_t^s, z_t^{d*}, u^*_t).
\end{aligned}$$

Since $\phi_t(\cdot)$, the markup $\partial \varphi_t^{-1}(\cdot)/\partial r_{it}$, and the control function $\mathbb{M}_t^{-1}(\cdot)$ are already identified,

$c_f(z_t^s)$ is identified as:

$$c_f(z_t^s) = \int_{r_{base}}^{r_{target}} \frac{\partial \varphi_t^{-1}(s, z_t^{d*}, u_t^*)}{\partial r_{it}} ds - \mathbb{M}_t^{-1}(m_{t0}^*, k_t^*, l_t^*, z_t^s, z_t^{d*}, u_t^*).$$

Thus, $f_t(\cdot)$ is identified.

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

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

The output price for individual firms is then identified as $p_{it} = r_{it} - \varphi_t^{-1}(r_{it}, z_{it}^d, u_{it})$. □

F.3.2 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 (26). In step 2, we consider the control variable approach as in subsection F.2.

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

Proof. Following the same steps in the proof for Proposition F.2 by replacing $G_{m_t|\nu_t}(\cdot|\cdot)$ with $G_{m_t|\nu_t,\xi_t}(\cdot|\cdot, \xi_{it})$, we can identify $\mathbb{M}_t^{-1}(m_{it}, w_{it}, u_{it})$ and $\bar{h}_t\left(m_{t-1}, w_{t-1}, u_{t-1}, z_{t-1}^h\right)$ 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

$$\begin{aligned}
&\widetilde{H}_t(z_t^s, z_t^d, z_{t-1}^s, z_{t-1}^d, z_{t-1}^h) \\
&:= E[\Lambda_m(x_{it}, z_t^s, z_t^d, u_{it}) - \Lambda_{\bar{h}_t}(x_{it-1}, z_t^s, z_t^d, u_{it-1}, z_{t-1}^h)|z_t^s, z_t^d, z_{t-1}^s, z_{t-1}^d, z_{t-1}^h].
\end{aligned}$$

Applying the law of iterated expectations with $E\left[\eta_{it}|\nu_{it}, \xi_{it}\right] = 0$ from Assumption F.2(b), we have

$$E\left[\eta_{it}|z_t^s, z_t^d, z_{t-1}^s, z_{t-1}^d, z_{t-1}^h\right] = E\left[E\left[\eta_{it}|\nu_{it}, \xi_{it}\right]|z_t^s, z_t^d, z_{t-1}^s, z_{t-1}^d, z_{t-1}^h\right] = 0.$$

From this, we have (F.29) as follows:

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

Therefore, following the same steps in the proof for Proposition F.2, we can identify $\mathbb{M}_t^{-1}(\cdot)$ up to scale and location, and identify $G_{\eta_t}(\cdot)$ up to scale. □

In step 3, Proposition F.3 holds with the same proof.

F.4 Identification with Persistent Demand Shocks

In Section 2.1, identification relies on the limited persistence of ϵ_{it} (Assumption 1) to justify using lagged inputs as instruments. If ϵ_{it} is persistent (e.g., $\epsilon_{it} = \rho_\epsilon \epsilon_{it-1} + \zeta_{it}$), lagged inputs become endogenous as they depend on past demand shocks correlated with current demand.

This subsection establishes that the revenue function remains nonparametrically identified under persistent demand shocks, provided one observes a *lagged supply-side shifter* z_{it-1}^h (e.g., R&D investment or technology adoption). Let $w_{it} := (k_{it}, l_{it}, z_{it}^s, z_{it}^d)$ be the vector of controls. We replace the baseline restrictions with the following assumption:

Assumption F.5 (Supply-side instrument for persistent demand). *Suppose that the followings hold: (a) Relevance. Conditional on w_{it}, z_{it-1}^h shifts the distribution of current physical productivity ω_{it}, and consequently shifts the flexible input m_{it} via the input demand function. (b) Exclusion and Orthogonality. z_{it-1}^h is excluded from the inverse demand function (conditional on controls) and is independent of the* current *demand innovation ζ_{it} and the past demand shock ϵ_{it-1} conditional on w_{it}: $(\zeta_{it}, \epsilon_{it-1}) \perp\!\!\!\perp z_{it-1}^h \,\big|\, w_{it}$. (c) Completeness. The family of conditional distributions of m_{it} given (z_{it-1}^h, w_{it}) is complete. That is, for any measurable function $g(m, w)$, $E[g(m_{it}, w_{it}) \mid z_{it-1}^h, w_{it}] = 0$ almost surely implies $g(m_{it}, w_{it}) = 0$ almost surely.*

Assumption F.5(b) assumes that the instrument z_{it-1}^h is pre-determined relative to demand; alternatively, we may take the extra lag, i.e., replacing z_{it-1}^h with z_{it-2}^h.

Under Assumption F.5 and the strict monotonicity of the revenue function (Assumption 2), the function $\phi_t(\cdot)$ is uniquely identified by the IVQR moment condition using (z_{it-1}^h, w_{it}):

$$
\Pr\left[r_{it} \leq \phi_t(m_{it}, w_{it}, u) \,\middle|\, z_{it-1}^h, w_{it}\right] = u, \quad \forall u \in [0, 1].
$$

Intuitively, z_{it-1}^h isolates variation in input choice driven by supply/productivity shifts that are orthogonal to the current demand innovation ζ_{it}, even if ϵ_{it} itself is serially correlated.

This extension replaces the requirement of transitory demand shocks with the requirement of a valid supply-side instrument. The validity of Assumption F.5 requires that the firm's choice of z^h_{it-1} is not systematically driven by the persistent component of demand (ϵ_{it-1}). Therefore, this strategy is most credible when z^h_{it-1} reflects supply-side forces (e.g., technological opportunities or cost incentives) rather than responses to medium-run demand fluctuations.

F.5 Unobserved Quality Heterogeneity

This subsection introduces persistent latent product quality as a structural source of persistent demand heterogeneity. The goal is *not* to identify the baseline model under a persistent ϵ_{it} process. Rather, we maintain the baseline requirement that ϵ_{it} captures high-frequency demand fluctuations (and hence satisfies the limited-persistence logic used in Step 1), while allowing persistent movements in demand to operate through a persistent quality component that is explicitly modeled. Put differently, persistent demand heterogeneity is accommodated by δ_{it} below, while ϵ_{it} is interpreted as the remaining transitory demand shock.

Let Δ_{it} denote the unobserved quality of firm i's product at time t, and let $\delta_{it} := \ln \Delta_{it}$. Define quality-adjusted price and quantity by

$$\tilde{P}_{it} := \frac{P_{it}}{\Delta_{it}} \qquad \text{and} \qquad \tilde{Y}_{it} := Y_{it}\Delta_{it}.$$

In logs, $\tilde{p}_{it} = p_{it} - \delta_{it}$ and $\tilde{y}_{it} = y_{it} + \delta_{it}$. The representative consumer derives utility from quality-adjusted units, so inverse demand is written as

$$\tilde{P}_{it} = \Psi_t(\tilde{Y}_{it}, z^d_{it}, \epsilon_{it}),$$

with Ψ_t strictly decreasing in $\tilde{Y}_{it}$. In observed variables this implies

$$p_{it} = \psi_t(y_{it} + \delta_{it}, z^d_{it}, \epsilon_{it}) + \delta_{it},$$

where $\psi_t(\cdot) := \ln \Psi_t(\exp(\cdot))$. Hence revenue satisfies

$$r_{it} = p_{it} + y_{it} = \varphi_t(y_{it} + \delta_{it}, z^d_{it}, \epsilon_{it}) := \psi_t(y_{it} + \delta_{it}, z^d_{it}, \epsilon_{it}) + (y_{it} + \delta_{it}). \tag{F.31}$$

On the production side, physical output is

$$y_{it} = f_t(x_{it}, z^s_{it}) + \omega_{it}, \qquad x_{it} = (m_{it}, k_{it}, l_{it}).$$

Substituting into (F.31) yields

$$r_{it} = \varphi_t\Big(f_t(x_{it}, z^s_{it}) + \underbrace{(\omega_{it} + \delta_{it})}_{\nu_{it}}, z^d_{it}, \epsilon_{it}\Big), \tag{F.32}$$

where $\nu_{it} := \omega_{it} + \delta_{it}$ is *revenue productivity*. In this extension, the objects identified by the revenue-based strategy naturally correspond to revenue productivity ν_{it}, rather than physical productivity ω_{it} alone.

To preserve the control-function structure used in Section 3, we make explicit the information set underlying input choice. We assume the firm observes the composite revenue-productivity state ν_{it} (or equivalently observes ω_{it} and δ_{it}) when choosing the flexible input, so the intermediate input policy takes the form

$$m_{it} = \mathbb{M}_t(\nu_{it}, w_{it}, u_{it}), \qquad w_{it} := (k_{it}, l_{it}, z^s_{it}, z^d_{it}),$$

and ϵ_{it} (equivalently u_{it}) remains the demand-side shock entering the revenue function monotonically as in Assumption 2. With this timing, Step 1 continues to recover $\varphi_t(\cdot)$ and u_{it} (under the same orthogonality conditions as in the baseline for u_{it}), and Step 2–3 recover the law of motion for the *identified* productivity state (now ν_{it}).

Assume physical productivity and quality each follow AR(1) processes:

$$\omega_{it} = \rho_\omega \omega_{it-1} + \eta^\omega_{it}, \qquad \delta_{it} = \rho_\delta \delta_{it-1} + \eta^\delta_{it},$$

with innovations $(\eta^\omega_{it}, \eta^\delta_{it})$. If $\rho_\omega = \rho_\delta = \rho$, then $\nu_{it} = \omega_{it} + \delta_{it}$ also follows AR(1):

$$\nu_{it} = \rho\, \nu_{it-1} + (\eta^\omega_{it} + \eta^\delta_{it}). \tag{F.33}$$

In this case, the Markov structure exploited in Section 3 applies directly to ν_{it}, and the multi-step identification strategy carries through with ν_{it} interpreted as the productivity state.

If instead $\rho_\omega \neq \rho_\delta$, then ν_{it} need not be AR(1). A natural formulation is to treat $(\omega_{it}, \delta_{it})$ as a two-dimensional Markov state. The revenue equation (F.32) depends on these states only through ν_{it}, so revenue data alone generally do not allow one to separately identify ω_{it} and δ_{it} without additional structure (e.g., quality proxies, direct price/quantity information beyond revenue, or restrictions that link quality to observables). For expositional clarity and to remain close to Section 3, we focus on the empirically relevant case in which ν_{it} admits a parsimonious first-order representation such as (F.33).

This extension rationalizes persistent demand heterogeneity through a persistent quality component while keeping ϵ_{it} as the high-frequency demand shock required for Step 1. Under this interpretation, the baseline method identifies (i) the revenue function and demand shock u_{it} and (ii) the dynamics of revenue productivity ν_{it}. Separating physical productivity ω_{it} from quality δ_{it} is not identified from revenue alone and would require additional assumptions or measurements.

F.6 Identification with Heterogeneous Material Prices

In the main text, we treat deflated expenditures on materials as measures of inputs. This abstracts from unobserved heterogeneity in material prices. For instance, geographically segmented input markets may induce systematic price differences across regions. This Appendix shows that the identification strategy extends to such environments when material prices vary across firms as a function of observed characteristics.

Suppose that econometricians observe only material expenditure, $m^*_{it} = m_{it} + p^m_{it}$, but not m_{it} and p^m_{it} separately, and that the material price p^m_{it} varies across firms and is determined by exogenous observables z^m_{it}:

$$p^m_{it} = p^m_t \left(z^m_{it}\right). \tag{F.34}$$

One motivation for specification (F.34) is the presence of regional input markets. When input markets are geographically segmented, material prices may differ across regions. In this case, z^m_{it} can be specified as a region dummy indicating the location of firm i.

Substituting $m_{it} = m^*_{it} - p^m_t \left(z^m_{it}\right)$ into f_t, we define the transformed production function f^*_t as

$$\begin{aligned} f_t \left(m_{it}, k_{it}, l_{it}, z^s_{it}\right) &= f_t \left(m^*_{it} - p^m_t \left(z^m_{it}\right), k_{it}, l_{it}, z^s_{it}\right) \\ &:= f^*_t \left(m^*_{it}, k_{it}, l_{it}, z^s_{it}, z^m_{it}\right). \end{aligned}$$

Because this transformation is additive in m_{it}, the elasticities of f_t and f^*_t coincide. Hence, identification of f^*_t is sufficient for identification of f_t.

Since z^m_{it} is exogenously given for each firm, the firm's profit-maximization problem can be written as a choice of m^*_{it} rather than m_{it}:

$$\begin{aligned} &\max_{m_{it}} \exp\left(\varphi_t \left(f_t \left(m_{it}, k_{it}, l_{it}, z^s_{it}\right) + \omega_{it}, z^d_{it}, u_{it}\right)\right) - \exp\left(m_{it} + p^m_t \left(z^m_{it}\right)\right) \\ =&\max_{m^*_{it}} \exp\left(\varphi_t \left(f^*_t \left(m^*_{it}, k_{it}, l_{it}, z^s_{it}, z^m_{it}\right) + \omega_{it}, z^d_{it}, u_{it}\right)\right) - \exp\left(m^*_{it}\right). \end{aligned}$$

Since z^m_{it} is observed and exogenous, it can be treated in the same way as z^s_{it} throughout the identification argument. Conditional on (z^s_{it}, z^m_{it}), the model is isomorphic to the baseline case in the main text, and all steps of the three-step identification strategy apply without modification.

G Robustness to Decreasing Returns to Scale

In the main text, we impose constant returns to scale (CRS), $\theta_m + \theta_k + \theta_l = 1$, as a scale normalization for the Cobb–Douglas production function. To assess the sensitivity of our empirical results to this restriction, we estimate the model under decreasing returns to scale (DRS), setting $\theta_m + \theta_k + \theta_l = 0.9$ and $\theta_m + \theta_k + \theta_l = 0.95$. All other aspects of the estimation procedure remain identical to the baseline.

Industry	n	$\hat{\theta}_m$	$\hat{\theta}_k$	$\hat{\theta}_l$	$\hat{\bar{\mu}}$
31	736	0.766	0.011	0.122	1.252
		(0.032)	(0.008)	(0.031)	(0.053)
32	463	0.680	0.063	0.157	1.351
		(0.058)	(0.035)	(0.037)	(0.114)
38	391	0.617	0.039	0.245	1.495
		(0.057)	(0.032)	(0.051)	(0.139)

Table G.1: Chilean Manufacturing plant estimation under $\theta_m + \theta_k + \theta_l = 0.9$: 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	623	0.360	3.383	-8.667
		(0.005)	(0.205)	(0.205)
32	354	0.094	0.911	-6.912
		(0.018)	(0.265)	(0.702)
38	318	0.070	0.703	-6.009
		(0.016)	(0.291)	(0.437)

Table G.2: Chilean Manufacturing plant estimation under $\theta_m + \theta_k + \theta_l = 0.9$: Step 4 (Industries 31, 32, and 38 in 1996). Standard errors in parentheses with 100 non-parametric bootstrap iterations.

Industry	n	$\hat{\theta}_m$	$\hat{\theta}_k$	$\hat{\theta}_l$	$\hat{\bar{\mu}}$
31	736	0.809	0.012	0.129	1.322
		(0.034)	(0.009)	(0.032)	(0.056)
32	463	0.718	0.067	0.166	1.426
		(0.061)	(0.037)	(0.039)	(0.120)
38	391	0.651	0.041	0.258	1.578
		(0.060)	(0.034)	(0.053)	(0.147)

Table G.3: Chilean Manufacturing plant estimation under $\theta_m + \theta_k + \theta_l = 0.95$: Step 1, Step 2, and Step 3 (Industries 31, 32, and 38 in 1996). Standard errors in parentheses with 100 non-parametric bootstrap iterations.

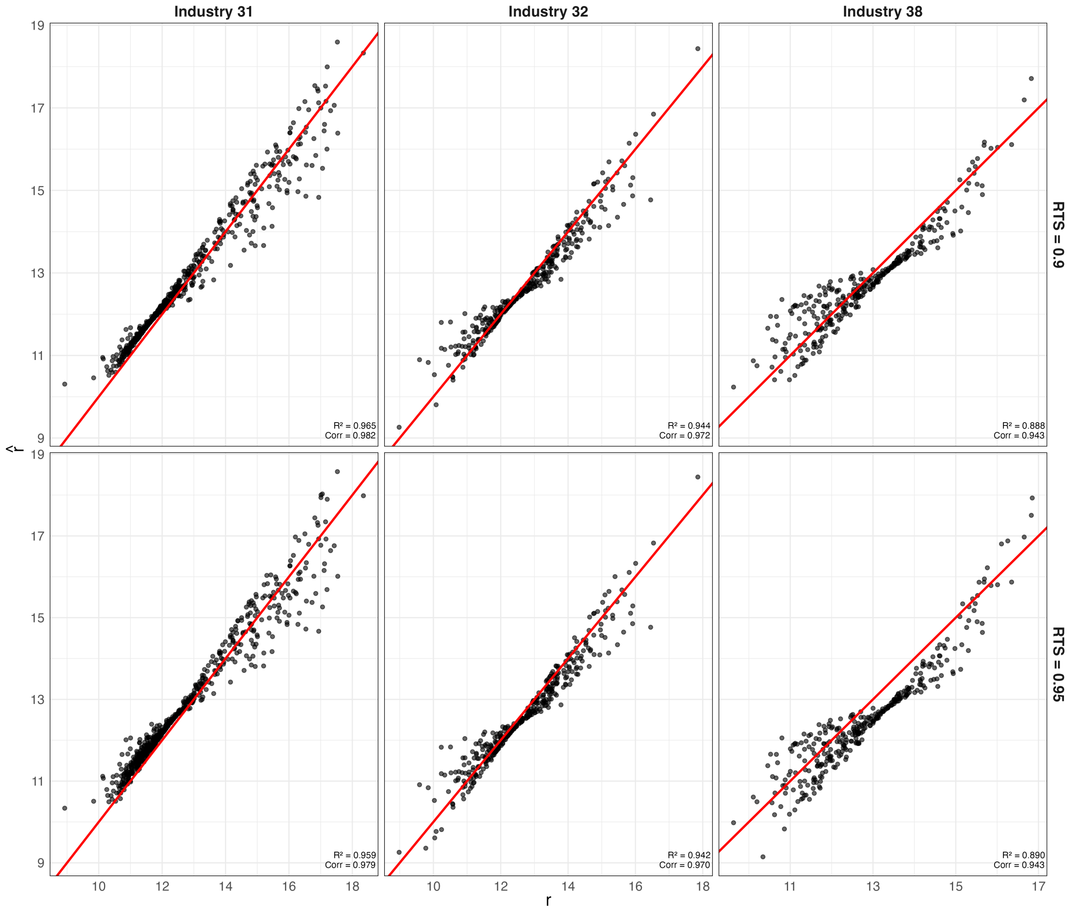
Industry	n	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$
31	668	0.183	1.976	-8.068
		(0.006)	(0.112)	(0.408)
32	386	0.088	0.916	-6.918
		(0.017)	(0.259)	(0.765)
38	336	0.066	0.695	-6.515
		(0.015)	(0.285)	(0.464)

Table G.4: Chilean Manufacturing plant estimation under $\theta_m + \theta_k + \theta_l = 0.95$: Step 4 (Industries 31, 32, and 38 in 1996). Standard errors in parentheses with 100 non-parametric bootstrap iterations.

The output elasticity and average markup estimates under DRS are proportionally scaled down relative to the CRS baseline. The HSA demand parameter $\hat{\beta}$ remains statistically significantly different from zero across all specifications and industries.

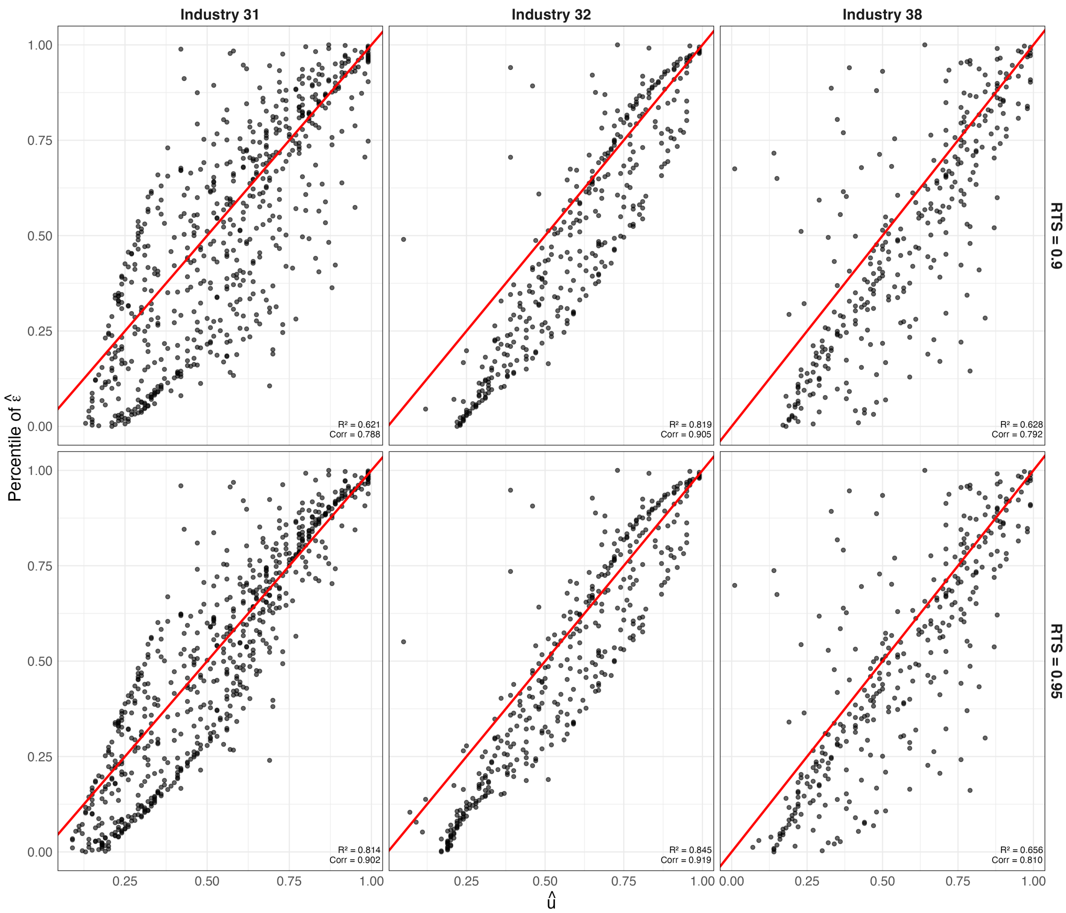
Figures G.1 and **??** present the analogues of Figures 4 and 5 from the main text under the concerning DRS specifications. As in the CRS case, observed revenue aligns closely with fitted revenue from the HSA demand system, and the quantile–quantile plots of demand shocks generally track the 45-degree line, confirming that the model fit is robust to the choice of returns to scale (RTS).

Figure G.1: Observed revenue vs. fitted revenue from Step 4 under DRS



Notes: The red line indicates the 45-degree line. Top panels: RTS = 0.9; bottom panels: RTS = 0.95.

Figure G.2: Rank of demand shock from Step 1 vs. Step 4 under DRS



Notes: The red line indicates the 45-degree line. Top panels: RTS = 0.9; bottom panels: RTS = 0.95.

Tables **??** and **??** report the counterfactual welfare results under DRS. The overall welfare losses from market power range from approximately 3%–10% of industry revenue, broadly consistent with the 3%–6% range found under CRS.

Industry	CV	$\Delta\Pi$	Overall
31	-9.97	-6.66	3.31
	(3.09)	(2.30)	(0.98)
32	-19.81	-12.32	7.49
	(4.44)	(3.92)	(1.40)
38	-20.00	-10.26	9.74
	(5.97)	(3.35)	(2.96)

Table G.5: 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 with $\theta_m + \theta_k + \theta_l = 0.9$. Standard errors in parentheses with 100 non-parametric bootstrap iterations.

Industry	CV	$\Delta\Pi$	Overall
31	-11.02	-7.93	3.10
	(3.31)	(2.76)	(0.87)
32	-22.58	-14.77	7.81
	(6.03)	(5.44)	(1.70)
38	-14.19	-7.58	6.61
	(5.49)	(3.64)	(2.38)

Table G.6: 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 with $\theta_m + \theta_k + \theta_l = 0.95$. Standard errors in parentheses with 100 non-parametric bootstrap iterations.