Prices, Profits, Proxies, and Production*

Victor H. Aguiar ⓡ Nail Kashaev ⓡ Roy Allen†

First version: October 10, 2018
This version: February, 2022

Abstract This paper studies nonparametric identification and counterfactual bounds for heterogeneous firms that can be ranked in terms of productivity. Our approach works when quantities and prices are latent, rendering standard approaches inapplicable. Instead, we require observation of profits or other optimizing-values such as costs or revenues, and either prices or price proxies of flexibly chosen variables. We extend classical duality results for price-taking firms to a setup with discrete heterogeneity, endogeneity, and limited variation in possibly latent prices. Finally, we show that convergence results for nonparametric estimators may be directly converted to convergence results for production sets.

JEL classification: C5, D24.

Keywords: Counterfactual bounds, cost minimization, nonseparable heterogeneity, partial identification, profit maximization, production set, revenue maximization, shape restrictions.

*An earlier version of this paper was circulated as "Prices, Profits, and Production: Identification and Counterfactuals." The "ⓡ" symbol indicates that the authors' names are in certified random order, as described by Ray ⓡ Robson (2018). We thank the editor, the associate editor, and three anonymous referees for their comments and suggestions. We are grateful to Paul Grieco, Lance Lochner, Rosa Matzkin, Salvador Navarro, David Rivers, Susanne Schennach, Holger Sieg, and Al Slivinsky for useful comments and encouragement. We also thank the ceminar participants at University of Montreal, McMaster University, and attendants of NASMES 2019, Empirical Microeconomics Workshop at University of Calgary, MEG 2019, CESG 2019, NAWMES 2020.

†Aguiar: Department of Economics, University of Western Ontario; vaguiar@uwo.ca. Kashaev: Department of Economics, University of Western Ontario; nkashaev@uwo.ca. Allen: Department of Economics, University of Western Ontario; rallen46@uwo.ca.

Introduction

This paper studies nonparametric identification of production sets and counterfactual bounds for firms, allowing multiple inputs and outputs, in an environment where both quantities and prices can be latent. We assume an analyst has data on the values of an optimization problem, such as profits, costs, or revenues, as well as prices or *price proxies*.

Identifying heterogeneous production sets is challenging in situations where the observability of some outputs/inputs or prices is problematic. For instance, in the housing market output quantities and output prices cannot be directly observed because houses provide different services that are hard to measure. However, housing values that can serve as price proxies may be observed (Epple, Gordon and Sieg, 2010). Other industries, such as health and banking, suffer from similar issues with unobservable inputs or outputs.[1] The latency of quantities makes standard approaches to estimate production functions not directly applicable. In addition, the latency of prices makes classical approaches using duality theory impossible to apply as well. In contrast, we require observability of values and prices or price proxies. While these variables are not always observed, they are available in many existing data sets.[2]

In order to obtain identification of firm-specific production possibility sets we exploit variation in prices or price proxies across markets and variation of optimization values across firms. Our framework extends classical duality theory by allowing (i) rich forms of complementarity and substitutability between outputs and inputs with discrete heterogeneity across firms, (ii) endogeneity between prices and productivity due to simultaneity and market entry decisions, and (iii) omitted prices of flexibly chosen variables. Classical duality theory focuses on either a nonstochastic or representative agent framework in which all prices are observed. Important contributions include Shephard (1953), Fuss and McFadden (1978), and Diewert (1982) among many others.

[1] In the health industry, it is difficult to measure inputs such as drugs since they vary widely in their physical characteristics. However, prices and total costs may be observable (Bilodeau, Cremieux and Ouellette, 2000). In the banking industry, outputs such as business loans and consumers loans are difficult to measure because a loan is a financial service that entails many unobservable goods and services. However, the price of a loan is observed as well as profits in some settings (Berger, Hancock and Humphrey, 1993).

[2] See Epple et al. (2010), Combes, Duranton and Gobillon (2021), and Albouy and Ehrlich (2018) in the context of housing; Burke, Bergquist and Miguel (2019) in the context of agriculture; Nerlove (1963) and Fabrizio, Rose and Wolfram (2007) in the context of electricity generation; Roberts and Supina (1996), Foster, Haltiwanger and Syverson (2008), and Doraszelski and Jaumandreu (2013) in the context of manufacturing.

We assume that firms can be ranked in terms of productivity that can take finitely many values. This assumption is key to unpack heterogeneity in multiple output/input production sets across firms from data such as prices or price proxies and scalar values of an optimization problem. We formalize this by assuming that a firm with higher productivity has access to all the production possibilities of a less productive firm, and more. Our framework covers Hicks-neutral heterogeneity in productivity as a special case.

Our approach exploits the rich shape constraints in our environment for identification and counterfactual analysis. Leveraging that firms can be ranked according to discrete productivity, we present a new method to identify the structural value function (e.g. profit function). This technique works with bounded measurement error, but allows rich forms of selection into market. We require a weak monotone presence assumption, so that if a firm is present in some market with certain observables, then each more productive firm must be present in some market with the same observables. This handles certain monotone selection rules, e.g. only firms that can make nonnegative profits enter, but is much more general.

We next tackle the important possibility that not all prices are observed. Instead, we use *price proxies*, which are unknown functions of the missing prices. As one example, we show that aggregate market-level quantities can serve as price proxies. We leverage homogeneity of the value function to recover these unknown functions. This technique is new, and is applicable to other settings with homogeneity of a structural function, and is therefore of independent interest.

Once the structural value function is identified, we turn to recoverability of the production sets. Here we leverage the classic insight that the value function serves as the support function of the production set. This allows us to characterize the most that can be said about heterogeneous production sets, even when price variation is limited. Building on this, we present a general framework for counterfactual questions such as sharp bounds on quantities or profits at a new price. Importantly, these bounds hold for each level of productivity, and thus characterize features of the distribution of firm behavior.

As mentioned previously, relative to classic work on duality we make several contributions by incorporating heterogeneity, endogeneity due to selection, and potential lack of prices.[3] Even when prices are observed but contain limited variation, we

[3]Outside of the firm problem, duality has been used in the presence of heterogeneity in discrete choice (McFadden, 1981), matching models (Galichon and Salanié, 2015), hedonic models (Chernozhukov, Galichon, Henry and Pass, 2017), dynamic discrete choice (Chiong, Galichon and Shum, 2016), and the additively separable framework of Allen and Rehbeck (2019).

contribute by providing new results using structural value functions to recover sets and conduct counterfactual analysis. This builds on Farrell (1957) and Afriat (1972), who study efficiency measurement and conditions under which producer datasets are consistent with the hypothesis of optimization. Relatedly, Hanoch and Rothschild (1972) focuses on finite deterministic datasets of individual firms' profits or costs, and prices. Hanoch and Rothschild (1972) does not study identification of the production set or the profit function, but focuses on providing necessary and sufficient conditions under which an observed production function is consistent with profit maximization or cost minimization.[4] Another paper studying limited price variation is Varian (1984), which works with quantities and prices and does not study unobservable heterogeneity.[5] While observation of prices and quantities implies observation of profits, the reverse is not true.

This paper contributes to the recent literature on identification and estimation of multi-output production with unobservable heterogeneity (e.g., Cunha, Heckman and Schennach, 2010, De Loecker, Goldberg, Khandelwal and Pavcnik, 2016, and Grieco and McDevitt, 2016). We differ since we do not observe quantities and we do not impose separability or parametric restrictions on the shape of production sets. Because we allow production of multiple outputs in flexible ways, use cross-sectional variation, and do not observe quantities, we also differ from an important recent literature studying single output production in dynamic panel settings using quantities data, including Griliches and Mairesse (1995), Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, Caves and Frazer (2015), and Gandhi, Navarro and Rivers (2020).[6]

We also contribute to the literature studying recoverability of sets. We build on the tight relationship between the structural value function and the production possibility sets of firms, by providing an equality relating estimation error of value functions and estimation error of production possibility sets. This result allows one to adapt consistency results for any nonparametric estimators of the value function for the purpose of set estimation. The result is related to a classical result in convex analysis linking the distance of support functions with the distance of the corresponding sets,

[4] Cherchye, Rock and Walheer (2016) studies the identification of profits and production sets with a finite deterministic dataset on prices and quantities.

[5] See also Cherchye, Demuynck, De Rock and De Witte (2014) and Cherchye, Demuynck, De Rock and Verschelde (2018). Cherchye et al. (2018) differs from this paper because they assume observed input quantities in the context of cost minimization.

[6] As noted in Ackerberg et al. (2015), some output and input data often come in the form of sales and expenditures that need to be transformed into quantities. We work directly with total values (e.g. profits, total costs, or revenues).

which has been exploited previously in the literature on partial identification.[7] We cannot apply the classical result since it would require seeing negative prices.

The rest of this paper proceeds as follows. In Section 1, we present a model of heterogeneous production in which firms are rankable in terms of productivity. Section 2 shows how to identify the structural value function. In Section 3, we extend our methodology to environments where one observes proxies that determine unobservable prices. Our main identification result for production possibility sets is in Section 4. Section 5 provides a general framework to conduct sharp counterfactual analysis in production environments. In Section 6, we show duality between estimation error in value functions and production sets. We conclude in Section 7. All proofs can be found in Appendix A. An estimator of the restricted profit function and an illustrative application are in Appendices B and C. Appendix D contains extensions and additional results.

1. Setup

This paper studies recoverability of the technology of heterogeneous firms given data on the value function of their maximization problems, as well as data on prices or price proxies that alter the maximization problems.

The technology of heterogeneous firms is described by a correspondence $Y : E \rightrightarrows \mathbb{R}^{d_y}$. Each set $Y(e)$ describes the possible input/output (or "netput") vectors that are feasible for a firm of type e. The variable e captures unobservable heterogeneity in productivity. Negative components of $Y(e)$ correspond to net demands by the firm and positive components correspond to net supply. This formulation allows us to treat single output and multi-output firms in a common framework.[8] We require the following conditions.

Definition 1. A correspondence $Y : E \rightrightarrows \mathbb{R}^{d_y}$ is a production correspondence if, for every $e \in E$,

(i) $Y(e)$ is closed and convex;

[7] See, for instance, Beresteanu and Molinari (2008), Beresteanu, Molchanov and Molinari (2011), Kaido and Santos (2014), Kaido (2016), and Kaido, Molinari and Stoye (2019).

[8] An alternative approach is to use transformation functions. See Grieco and McDevitt (2016) for a recent application.

(ii) $Y(e)$ satisfies *free disposal*: if y is in $Y(e)$, then any y^* such that $y_j^* \leq y_j$ for all $j \in \{1, \cdots, d_y\}$ is also in $Y(e)$;

(iii) $Y(e)$ satisfies *the recession cone property*: if $\{y^m\}$ is a sequence of points in $Y(e)$ satisfying $\|y^m\| \to \infty$ as $m \to \infty$, then accumulation points of the set $\{y^m / \|y^m\|\}_{m=1}^{\infty}$ lie in the negative orthant of $\mathbb{R}^{d_y}$.

These conditions rule out infinite profits and ensure that the maximization problems we consider have a solution.[9]

We study the general *restricted profit maximization problem*

$$\pi_r(y_{-z}, p_z, e) = \max_{y_z : (y_{-z}, y_z) \in Y(e)} p_z' y_z \, ,$$

where y_{-z} is a vector of restricted or fixed variables, y_z denotes the variables of choice, and p_z is a vector of prices of y_z. The variable of choice y_z is constrained to belong to the convex set $Y_r(y_{-z}, e)$ defined as

$$Y_r(y_{-z}, e) = \left\{ y_z \in \mathbb{R}^{d_{y_z}} \; : \; (y_{-z}, y_z) \in Y(e) \right\} \, .$$

We refer to $Y_r(y_{-z}, \cdot)$ as the restricted production correspondence.[10]

The behavioral restriction of this model is that given y_{-z}, the firm chooses y_z to maximize restricted profits, taking prices p_z as given. In the special case where y_{-z} is not present, this is the usual profit maximization setup. When y_{-z} consists of inputs, this covers revenue maximization. When y_{-z} consists of outputs, this is cost minimization once we interpret negative y_z as inputs and write

$$\max_{y_z :\, (y_{-z}, y_z) \in Y(e)} p_z' y_z = - \min_{y_z :\, (y_{-z}, y_z) \in Y(e)} p_z' (-y_z).$$

We emphasize that throughout, y_{-z} can be a vector, and so we cover cost minimization with multiple inputs, and revenue maximization with multiple outputs.

Overall, we consider firms that are price-taking in the variables of choice y_z, and study a static problem without uncertainty. We note though that in principle the production set $Y(e)$ is general enough to describe paths of production possibilities throughout time, as would arise if there is investment.

[9] See Kreps (2012), p. 199 for more details.

[10] More formally, it is only a multi-valued mapping because it can be empty for certain combinations of y_{-z} and e. We note that the results in this paper do not need the full strength of $Y(\cdot)$ being a production correspondence. Instead, we require that the set $Y_r(y_{-z}, e)$ be closed and convex, satisfy free disposal, and satisfy the recession cone property.

1.1. Setting and Data

We study identification in settings in which an analyst observes many realizations of certain values of the restricted profit maximization problem as prices vary. In the most general version, we observe noisy measurements of restricted profits, which are the values of the restricted problem. Specifically, we consider the setup

$$\boldsymbol{\pi}_r = \pi_r(\mathbf{y}_{-z}, \mathbf{p}_z, \mathbf{e}) + \boldsymbol{\eta} \text{ a.s.},$$

where $\boldsymbol{\pi}_r$ and $\mathbf{y}_{-z}$ are observed,[11] $\boldsymbol{\eta}$ is unobserved measurement error, and $\mathbf{e}$ is unobservable productivity level. For each component of $\mathbf{p}_z$, the analyst either observes the corresponding price, or more generally observes a price proxy $\mathbf{x}_j$ that is linked to the unobserved price by the relationship $\mathbf{p}_{z,j} = g_j(\mathbf{x}_j, \tilde{\mathbf{x}})$, where $\tilde{\mathbf{x}}$ consists of some control variables. We provide further examples and discussion of such proxies in Section 3.

As an example of observables for cost minimization of hospitals (Bilodeau et al., 2000), the analyst observes total cost on variable inputs $\mathbf{y}_z$ (labor, supplies, food for patients, drugs, and energy), input prices or input-price proxies, fixed outputs (inpatient care and outpatient visits), and the fixed inputs (number of physicians and capital). We emphasize that we do not need to observe the quantities $\mathbf{y}_z$ of the flexibly chosen variables.[12]

Now we turn to the description of the sources of variation in our setup. Although we do not fully flesh out an equilibrium model incorporating selection, we provide an informal discussion of these forces. First, prices can vary across markets due to variation in endowments or the income or tastes of consumers. Our results apply when an analyst observes a single firm from each market, and has observations from many markets. Our results also apply when an analyst observes multiple firms in each market. We focus on the former case to simplify presentation, so that we can avoid market-level subscripts.

[11] We use bold font for random variables and vectors and regular font for their realizations.

[12] As discussed in the introduction, for additional data sets, see Nerlove (1963), Roberts and Supina (1996), Fabrizio et al. (2007), Foster et al. (2008), Epple et al. (2010), Doraszelski and Jaumandreu (2013), Albouy and Ehrlich (2018), Burke et al. (2019), and Combes et al. (2021).

2. Recoverability of Restricted Profit Function

Our ultimate goal is to learn about the production correspondence. We proceed in three steps. In this section, we first identify the restricted profit function (or value function) for heterogeneous firms assuming that the prices are perfectly observed. In Section 3 we show how to apply our analysis to the general case with unobserved prices. In subsequent sections we show how to use information on the restricted profit function to recover features of the production correspondence and describe the most that can be learned concerning counterfactual questions.

Identifying the restricted profit function for heterogeneous firms is challenging. The value function is nonseparable in latent productivity. Both the restricted variables $\mathbf{y}_{-z}$ and prices $\mathbf{p}_z$ may be endogenous. This leads to simultaneity and selection biases. We consider a setting without panel data or instruments. We present a new technique to identify the restricted profit function that addresses these challenges. The key restrictions of the technique are that (i) heterogeneity is one dimensional and allows us to rank firms, and (ii) there are finitely many types of firms.

2.1. Production Monotonicity

It is well-known that the firm problem admits a representative agent, and in principle this observation can be used to recover a *representative agent* restricted profit function. Even a representative agent analysis here is nontrivial because of challenging selection/simultaneity issues discussed previously. Here, we wish to recover not only a representative agent restricted profit function, but also recover the heterogeneous structural restricted profit functions. Recovering heterogeneous structural functions allows us to a conduct rich counterfactual analysis concerning how different types of firms are differentially affected by a policy.

To get traction on this problem, we assume firms are rankable in terms of productivity. We think of heterogeneous productivity as an ability to produce more with a given level of inputs (or produce the same output using lower levels of inputs). In other words, the production set of a firm with lower value of productivity is a subset of the production set of a firm with a higher productivity (see Figure 1). Note that $Y_r(y_{-z}, e) \subseteq Y_r(y_{-z}, \tilde{e})$ if and only if $\pi_r(y_{-z}, p_z, e) \leq \pi_r(y_{-z}, p_z, \tilde{e})$ for all p_z. This means that more productive firms have access to a bigger set of production possibilities, and will make more profits or pay lower costs given prices. We formalize

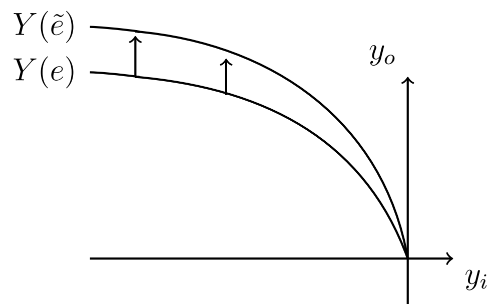

Figure 1 – Nested Production Sets. $\tilde{e} > e$.

this monotonicity by the following ranking assumption on the restricted profit function.

Assumption 1 (Strict Monotonicity)**.** *For every y_{-z}, p_z, e, and $\tilde{e}$ in the support, if $e < \tilde{e}$, then $\pi_r(y_{-z}, p_z, e) < \pi_r(y_{-z}, p_z, \tilde{e})$.*

Strict monotonicity of structural functions has been considered previously in e.g. Matzkin (2003). Assumption 1 is satisfied in many settings. For instance, it is satisfied in a standard single output production function setting with Hicks-neutral productivity. To be more specific, let the single output be y_o and let inputs be l and k, interpreted as labor and capital. Then the set $Y(e)$ is described by tuples $(y_o, -l, -k)$ that satisfy $y_o \leq f(l, k, e)$, where f is the production function. If $f(l, k, e) = A(e)\bar{f}(l, k)$ for some nonnegative, strictly increasing function A, and $\bar{f}$ is a nonnegative strictly convex function, then $f(l, k, e)$ is strictly increasing in e. In this case, $\pi(p, \cdot)$ satisfies Assumption 1.

More generally, the function $f(l, k, e) = A_o(e)\bar{f}(A_l(e)l, A_k(e)k)$ for strictly increasing functions A_o, A_l, and A_k fits into our setup.[13] A more general setup would allow a different shock to enter A_o, A_l, and A_k (e.g. Doraszelski and Jaumandreu, 2018) and would be outside of our framework. Overall, while Hicks-neutral heterogeneity is a special case of our framework when there is a *single* output, it is considerably more restrictive than needed for the monotonicity assumption to hold.

The assumption that production sets are nested in e is equivalent to the profit function being weakly increasing in e. Thus, value functions are the "right" structural function in which to impose monotonicity if we think of higher productivity as leading to more production possibilities. One may draw the intuition that in general other

[13] Li and Sasaki (2017) study a related setup with random coefficients Cobb-Douglas technology, imposing that the ratio of random coefficients is a monotone function of a single latent scalar random variable.

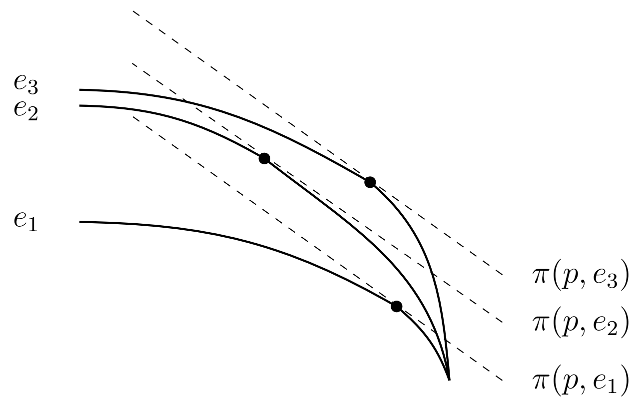

Figure 2 – Nonmonotonic supply.

structural functions are monotone in unobservable heterogeneity. This intuition is false without more structure.

Example 1 (Nonmonotonicity of Inputs/Outputs)**.** Consider the production sets depicted in Figure 2. Each production set is given by $Y(e_i) = \{(y_o, l)' \in \mathbb{R} \times \mathbb{R}_+ : y_o \leq f(l, e_i)\}$, where $f(l, e_1) < f(l, e_2) < f(l, e_3)$ for all $l > 0$. Here, $\pi(p, e_1) < \pi(p, e_2) < \pi(p, e_3)$ for all positive p and Assumption 1 is satisfied. Given the price vector $p = (p_o, p_k)'$ in Figure 2, the optimal levels of inputs and outputs are nonmonotone in productivity since $l^*(p, e_1) < l^*(p, e_3) < l^*(p, e_2)$ and $y_o^*(p, e_1) < y_o^*(p, e_3) < y_o^*(p, e_2)$. For a numerical example see Appendix D.2.

Failures of monotonicity in the optimal choice of input or output have been discussed as well in Pakes (1996, Section 4). Thus, rather than focus on the structural functions describing optimal input/output choices, this paper focuses instead on the restricted profit function, which *is* monotone in a scalar unobservable under the assumption that production sets are nested in e.

2.2. Discrete Heterogeneity and Monotone Selection

With this setup, we consider a new technique to identify the restricted profit function allowing endogeneity. The reason endogeneity is a central concern in such problems is that constraints may be endogenous. For example, in the cost minimization problem, output ($\mathbf{y}_{-z} = \mathbf{y}_o$) is typically a choice variable for the firm. An additional endogeneity concern is that firms may choose in which markets to operate. This can induce a selection issue, though we emphasize that once a market is chosen, the input/output vector is determined taking market prices as fixed. As discussed in Section 1.1, price variation in our setting arises because firms operate in different markets, which have different endowments or consumer tastes.

The key restriction we impose is that there are finitely many types of firms. We formalize this as follows.

Assumption 2 (Finite Heterogeneity)**.** *$E = \{1, 2, \ldots, d_e\}$, where d_e is finite and unknown to the researcher.*

This assumption allows us to identify structural functions without instruments. If instruments are available, continuous heterogeneity can be tackled by existing techniques *provided* there is no measurement error; see for example Appendix D.1. We emphasize that heterogeneity here is in terms of the production types, but due to measurement error in the data we may see continuous distributions of the restricted values, even when we condition on all other observables. In this modeling decision we are close to structural dynamic discrete choice literature that often assumes unobserved discrete heterogeneity that is smoothed out by some continuous idiosyncratic noise (e.g. extreme value distributed preference shock). See, for instance, Arcidiacono and Miller (2011).[14] We are not aware of any identification results that allow for both measurement error and continuous nonseparable structural unobserved heterogeneity in cross-sectional data.

We allow rich selection into markets, but impose a monotonicity restriction relating the types of firms that can be present, conditional on certain observables.

Assumption 3 (Monotone Presence)**.**

$$\mathbb{P}\left(\mathbf{e} = e|\mathbf{y}_{-z} = y_{-z}, \mathbf{p}_z = p_z\right) > 0 \implies \mathbb{P}\left(\mathbf{e} = \tilde{e}|\mathbf{y}_{-z} = y_{-z}, \mathbf{p}_z = p_z\right) > 0$$

for all y_{-z}, p_z, e, and $\tilde{e}$ in the support such that $e < \tilde{e}$.

This means that if we see a firm of type e active in *some* market and producing y_{-z}, conditional on $\mathbf{p}_z = p_z$, then for any higher productivity $\tilde{e}$, there is some market in which the higher type is active at the same value of conditioning variables. In principle, this other "market" could be the same market in which the firm with productivity e is present. The key restriction is that since we also condition on quantities, we need the higher type to *also* produce the same quantities.

[14] For applications of discrete unobserved heterogeneity, see Fox and Gandhi (2016) in multinomial choice and Bonhomme and Manresa (2015) with panel data.

As an example, consider the (unrestricted) profit function. Suppose entry depends on whether a firm obtains nonnegative profits. Specifically,

$$e \text{ enters} \iff \pi(p, e) \geq 0,$$

where there are no restricted variables. Since we assume monotonicity of π in e, this is a monotone threshold rule, and satisfies Assumption 3.

Assumption 3 is considerably more general than a one-sided selection rule. Importantly, it is only about the *support* of $\mathbf{e}$ conditional on some other variables. The reason we require this is that while reasonable selection rules into *markets* may result in a one-sided threshold rule, here we also need to allow selection into the quantities of the restricted variables y_{-z}. For example, as e increases the optimal quantity of the restricted variables may change. Assumption 3 allows this and is satisfied if, for example, there are other unobserved variables that shift the optimal choice of restricted variables y_{-z} (e.g. unobserved prices of the restricted variables).

2.3. Identification

We now turn to identification of the restricted profit function. First, recall that we observe potentially mismeasured restricted profits:

$$\boldsymbol{\pi}_r = \pi_r(\mathbf{y}_{-z}, \mathbf{p}_z, \mathbf{e}) + \boldsymbol{\eta}.$$

If $\boldsymbol{\eta}$ is independent of $(\mathbf{y}_{-z}, \mathbf{p}_z, \mathbf{e})$, then Assumption 2 implies that the conditional distribution of $\boldsymbol{\pi}_r$ can be written as a finite mixture of shifted distributions of $\boldsymbol{\eta}$:

$$F_{\boldsymbol{\pi}_r|\mathbf{y}_{-z},\mathbf{p}_z}(\cdot|y_{-z}, p_z) = \sum_{e\in E} F_{\boldsymbol{\eta}}(\cdot - \pi_r(y_{-z}, p_z, e))\mathbb{P}\left(\mathbf{e} = e|\mathbf{y}_{-z} = y_{-z}, \mathbf{p}_z = p_z\right),$$

where $F_{\boldsymbol{\pi}_r|\mathbf{y}_{-z},\mathbf{p}_z}(\cdot|y_{-z}, p_z)$ is the conditional cumulative distribution function (c.d.f.) of $\boldsymbol{\pi}_r$ conditional on $\mathbf{y}_{-z} = y_{-z}$ and $\mathbf{p}_z = p_z$, and $F_{\boldsymbol{\eta}}$ is the c.d.f. of $\boldsymbol{\eta}$. There are numerous ways to identify the above finite mixture model under different sets of assumptions that may be valid in different environments (see, for instance, Kitamura and Laage, 2018 and references therein). However, most of these results use either repeated measurements (i.e. panels) or use variation in conditioning variables, and require some form of exclusion restrictions (e.g., some conditioning variables affect $\pi_r(y_{-z}, p_z, e)$ but do not affect $\mathbb{P}\left(\mathbf{e} = e|\mathbf{y}_{-z} = y_{-z}, \mathbf{p}_z = p_z\right)$), or the presence of instruments. We propose a new set of assumptions to identify the above finite mixture in cross-sections, without instruments and exclusion restrictions. Moreover, our approach is constructive and the assumptions are easy to interpret.

Let $\Delta\pi_r(y_{-z}, p_z, e) = \pi_r(y_{-z}, p_z, e) - \pi_r(y_{-z}, p_z, e-1)$ denote the restricted profit difference between firms with adjacent productivity. We impose the following assumption on the measurement error $\boldsymbol{\eta}$.

Assumption 4. *(i) $\boldsymbol{\eta}$ is independent of $(\mathbf{y}_{-z}, \mathbf{p}_z, \mathbf{e})$, mean zero, has connected support, and satisfies $\mathbb{P}(|\boldsymbol{\eta}| \leq K/2) = 1$ for some $K < \infty$;*

(ii) (Separatedness) There exists (y^*_{-z}, p^*_z, e^*) in their support such that

$$K < \begin{cases} \Delta\pi_r(y^*_{-z}, p^*_z, e^*+1), & \text{if } e^* = 1, \\ \Delta\pi_r(y^*_{-z}, p^*_z, e^*), & \text{if } e^* = d_e, \\ \min\left\{\Delta\pi_r(y^*_{-z}, p^*_z, e^*+1),\ \Delta\pi_r(y^*_{-z}, p^*_z, e^*)\right\}, & \text{otherwise.} \end{cases}$$

We note that multiplicative measurement error can be handled by similar independence and separatedness assumptions.[15]

Assumption 4(i) means that the measurement error is classical. It also imposes a location normalization on the boundedly-supported measurement error.[16] The bounded support assumption is empirically relevant in many settings. For instance, revenues and costs cannot be negative, which provides a one-sided bound. Assumption 4(ii) is more substantial. This assumption imposes that the gap between the structural profits of the types adjacent to e^* must be sufficiently small compared with the support of measurement error. This can be restrictive in certain empirical settings but is essential for this method. We argue that boundedness and separatedness are appropriate in our empirical illustration in Appendix C.

Note that Assumption 4(ii) has to be imposed on one triplet (y^*_{-z}, p^*_z, e^*) only. Thus, in general, the measurement error may completely change the ranking of restricted profits. Moreover, this triplet does not need to be known. A simple sufficient condition for Assumption 4(ii) that uses shape restrictions of the restricted profit function is stated in the following result.

Lemma 1 (Rich Support)**.** *If Assumption 1 holds and there exist y^*_{-z} and p^*_z such that $\cup_{\lambda\geq 1}\{\lambda p^*_z\}$ is in the support of $\mathbf{p}_z$ conditional on $\mathbf{y}_{-z} = y^*_{-z}$, then Assumption 4(ii) is*

[15] The bounded support and separatedness conditions in Assumption 4 can be relaxed using results in Schennach (2016) if one has access to repeated cross-sections.

[16] For examples of papers studying boundedly-supported measurement errors see Hu and Ridder (2010), D'Haultfœuille and Février (2015), and Hu, Schennach and Shiu (2017).

satisfied.

This exploits homogeneity in prices, i.e. $\pi_r(y^*_{-z}, \lambda p^*_z, e) = \lambda\pi_r(y^*_{-z}, p^*_z, e)$ for all e and $\lambda > 0$. The idea behind Lemma 1 is that although the difference between profits evaluated at a particular price may not be big enough to offset the effect of the measurement error (e.g. $\Delta\pi_r(y^*_{-z}, p^*_z, e^* + 1) \leq K$), by exploiting homogeneity we always can find λ^* big enough such that

$$\Delta\pi_r(y^*_{-z}, \lambda^* p^*_z, e^* + 1) = \lambda^*\Delta\pi_r(y^*_{-z}, p^*_z, e^* + 1) > K.$$

The conditions of Lemma 1 guarantee that an extreme price $\lambda^* p^*_z$ can be found in the support for every finite K. Thus, the support of prices does not have to be unbounded, just sufficiently large relative to the initial difference.

Now we can state our main identification result for the restricted profit function.

Theorem 1. *Suppose Assumptions 1-4 hold. Then using $F_{\boldsymbol{\pi}_r|\mathbf{y}_{-z},\mathbf{p}_z}$, π_r is identified over the joint support of $\mathbf{y}_{-z}$, $\mathbf{p}_z$, and $\mathbf{e}$.*

Here, we may not be able to identify the structural restricted profit function for certain arguments outside of the support. This is particularly relevant for low types; there may be many combinations of prices and quantities such that low types do not produce either because it is infeasible for them or unprofitable.

Importantly, Theorem 1 only imposes a mild restriction on the stochastic dependence between unobservable heterogeneity $\mathbf{e}$ and observed $\mathbf{y}_{-z}$ and $\mathbf{p}_z$. In particular, in cost minimization settings, the output level and input prices can be related to the distribution of productivity in flexible ways. What is key is the monotonicity restriction on selection into markets described in Assumption 3.

The intuition behind Theorem 1 is that without restricting the dependence structure, monotonicity in the restricted profit function implies that firms always can be ranked. The assumption of the discrete heterogeneity allows us to match firms with the same ranking across different markets, and thereby construct the restricted profit function.

Theorem 1 can be used to weaken assumptions usually made in analysis of restricted profit maximizing behavior. For instance, with cost minimization, Bilodeau et al. (2000) focuses on a parametric setup with additively separable heterogeneity and assumes that fixed variables are exogenous. While working with the same observables, our methodology does not require parametric restrictions, and does not assume exogeneity.

Remark 1 (Testability). Theorem 1 identifies the restricted profit function π_r without using the shape restrictions that characterize such functions. Thus, the assumptions in this paper are testable. Specifically, for each e, the identified function $\pi_r(y_{-z}, p_z, e)$ must be convex, monotonically decreasing, and homogeneous of degree 1 in the prices of the flexible variables p_z. These implications can be tested with data on the values of the restricted problem $\boldsymbol{\pi}_r$, the restricted quantities $\mathbf{y}_{-z}$, and prices $\mathbf{p}_z$.

3. Unobservable Prices and Proxies

In Section 2, we showed how to identify the restricted profit function when the entire vector of prices of flexibly chosen variables, $\mathbf{p}_z$, is observed. In many empirical applications not all prices are observed. This may cause concern about *omitted price bias* (see Zellner, Kmenta and Dreze, 1966, Klette and Griliches, 1996, Katayama, Lu and Tybout, 2003, and Epple et al., 2010). However, the researcher may have access to some observable proxies that are informative about unobservable prices. For example, the rental rate of capital may be linked to market-specific characteristics such as short-term and long-term interest rates. Wages may be linked to the unemployment level or aggregate labor supply. De Loecker et al. (2016) uses output price, market shares, product dummies, firm location, and export status as proxies for unobservable input prices. In the housing market, an analyst may use location as a price proxy for a house as in Combes et al. (2021).[17]

This section studies how to identify the function linking prices proxies to unobserved prices through

$$\mathbf{p_j} = g_j(\mathbf{x}_j, \tilde{\mathbf{x}}) \text{ a.s.},$$

where g_j is an unknown function and $\mathbf{p}_j$ is a component of the vector of prices of the flexibly-chosen variables $\mathbf{p}_z$. We show how to identify g_j using the fact that the restricted profit function is homogeneous of degree 1, though as discussed in the Introduction, the technique we present is new and applies to any degree of homogeneity.[18] We assume that every price has its own *excluded* proxy $\mathbf{x}_j$, which is a proxy that affects its own price and does not affect any other prices. The vector of common proxies $\tilde{\mathbf{x}}$ may include common market characteristics such as size of the market or other macroeconomic characteristics. Importantly, since g_j is fully

[17] Hedonic pricing models also exhibit similar structure. However, in that literature it is assumed that both prices and proxies are observed. See, for instance, Ekeland, Heckman and Nesheim (2004).

[18] Homogeneity has been used for identification in Matzkin (1992), which differs in techniques and setting.

nonparametric, $\tilde{\mathbf{x}}$ can include categorical variables such as location (e.g. country or state) and time (e.g. month or year) identifiers. The special case in which price is observed corresponds to $g_j(x_j, \tilde{x}) = x_j$, where x_j is the price of y_j. To simplify the exposition we drop $\tilde{\mathbf{x}}$ from the notation, and analysis may be interpreted conditional on $\tilde{\mathbf{x}} = \tilde{x}$. For instance, we write $g_j(\mathbf{x}_j)$ instead of $g_j(\mathbf{x}_j, \tilde{\mathbf{x}})$.

Note that we assume prices are not a function of e or any other unobservables. Importantly, this rules out measurement error in prices. In our setup prices vary across markets, but are constant within a given market. Price-taking behavior implies that prices can be a function of the distribution of $\mathbf{e}$ in a market, but not the firm-specific productivity e.

We first present an informal outline how to identify g when one observes unrestricted profits, so that there are no restricted variables. We denote $x = (x_j)_{j=1,\ldots,d_{yz}} \in X$ and $g(x) = (g_j(x_j))_{j=1,\ldots,d_{yz}}$. Profits are given by $\pi(g(x), e)$. If the function g were known, we could identify π directly by previous arguments. What remains is to identify g. Recall that the profit function $\pi(\cdot, e)$ is homogeneous of degree 1, which from Euler's homogeneous function theorem yields the system of equations

$$\sum_{j=1}^{d_y} \partial_{p_j} \pi(p, e) p_j = \pi(p, e) \, .$$ [19]

Replacing prices with price proxies, we obtain

$$\sum_{j=1}^{d_y} \partial_{p_j} \pi(g(x), e) g_j(x_j) = \pi(g(x), e) \, . \tag{1}$$

Define $\tilde{\pi}(x, e) = \pi(g(x), e)$. Because x_j is exclusive to p_j, the cross-partial derivatives satisfy $\partial_{x_j} g_k(x_k) = 0$ for $j \neq k$. We thus have

$$\partial_{x_j} \tilde{\pi}(x, e) = \sum_k \partial_{p_k} \pi(g(x), e) \partial_{x_j} g_k(x_k) = \partial_{p_j} \pi(g(x), e) \partial_{x_j} g_j(x_j) \, .$$

Plugging this in to (1) we obtain

$$\sum_{j=1}^{d_y} \partial_{x_j} \tilde{\pi}(x, e) \frac{g_j(x_j)}{\partial_{x_j} g_j(x_j)} = \tilde{\pi}(x, e) \, . \tag{2}$$

Assume for now that $\tilde{\pi}(\cdot, e)$ is identified. Thus the only unknowns involve g. By

[19]Recall that we work with the unrestricted profit function for notational simplicity, but the restricted profit function is also homogeneous of degree 1 in prices.

varying x, holding everything else fixed, Equation 2 can be used to generate a system of equations. We show that when a certain rank condition is satisfied, it is possible to identify the entire function g using an appropriate scale/location normalization. We note that if all prices are observed except one, then we may directly apply Equation 2 to learn about g_j.

To formalize this, we impose location/scale conditions and some regularity conditions on g.

Assumption 5. *(i) $g_1(x_1) = x_1$ for all x_1, i.e. the price of the 1-st flexibly chosen variable is observed;*

(ii) The value of g is known at one point, i.e. there exist known x_0 and p_0 such that $g(x_0) = p_0$;

(iii) $X = \prod_{j=1}^{d_{yz}} X_j$ where each set $X_j \subseteq \mathbb{R}$ is an interval with nonempty interior;

(iv) $g_j(\cdot)$ is continuous everywhere and differentiable on the interior of X_j, and the set

$$\left\{x_j \in X_j \,:\, \partial_{x_j} g(x_j) = 0\right\}$$

has Lebesgue measure zero for every j.

Assumptions 5(i)-(ii) allow us to identify the scale and the location, respectively, of the multivariate function g. Since we can always relabel both outputs and inputs, Assumption 5(i) is equivalent to assuming that at least one price (not necessary p_1) is observed.

We now turn to our rank condition. This condition ensures that the system of equations generated from (2) has sufficient variation to recover terms such as $g_j(x_j)/\partial_{x_j} g_j(x_j)$.

Definition 2. We say that $h : \prod_{j=1}^{d_{yz}} X_j \to \mathbb{R}$ satisfies the rank condition at a point $x_{-1} \in \prod_{j=2}^{d_{yz}} X_j$ if there exists a collection $\{t_l\}_{l=1}^{d_{yz}-1} \subseteq X_1$ such that

(i) $x_l^* = (t_l, x_{-1}')' \in \prod_{j=1}^{d_{yz}} X_j$;

(ii) The square matrix

$$\begin{bmatrix} \partial_{x_2} h(x_1^*) & \dots & \partial_{x_{d_{yz}}} h(x_1^*) \\ \partial_{x_2} h(x_2^*) & \dots & \partial_{x_{d_{yz}}} h(x_2^*) \\ \dots & \dots & \dots \\ \partial_{x_2} h(x_{d_{yz}-1}^*) & \dots & \partial_{x_{d_{yz}}} h(x_{d_{yz}-1}^*) \end{bmatrix}$$

is nonsingular.

We will apply this rank condition to $\tilde{\pi}$ in place of h. It is helpful to recall that by Hotelling's lemma, partial derivatives of $\tilde{\pi}$ take the form

$$\partial_{x_j}\tilde{\pi}(x,e) = \partial_{p_j}\pi(p,e)|_{p=g(x)}\partial_{x_j}g_j(x_j) = y_j(g(x),e)\partial_{x_j}g_j(x_j)\,,$$

where $y_j(g(x),e)$ is the supply for good j. Thus, this rank condition applied to $\tilde{\pi}$ may equivalently be interpreted as a rank condition involving the supply function for the goods as well as certain derivatives of g. In words, variation in observed prices should induce enough variation in supply of goods with unobserved prices.

The following result provides conditions under which the price-proxy function g is identified. We note that while our exposition above covered the case of unrestricted profits, the following result holds for the more general setting of restricted profits. Thus, instead of the function $\tilde{\pi}$, we will use its restricted version defined via $\tilde{\pi}_r(x,e) = \pi_r(y^*_{-z}, g(x), e)$, where y^*_{-z} is fixed.

Theorem 2. *Suppose Assumption 5 holds. Then g is identified over the support of $\mathbf{x}$ if for some y^*_{-z}, the following conditions hold:*

(i) $\tilde{\pi}_r(x,e)$ is identified for each x and e in the support;

(ii) For every $x_{-1} \in \prod_{j=2}^{d_{y_z}} X_j$, there exists e^{**} in the support such that $\tilde{\pi}_r(\cdot, e^{**})$ satisfies the rank condition at x_{-1}.

To interpret (i), recall that Theorem 1 provides conditions under which $\tilde{\pi}_r$ is identified from the conditional distribution of $\pi_r(\mathbf{y}_{-z}, g(\mathbf{x}), \mathbf{e})$ conditional $\mathbf{x} = x$ and $\mathbf{y}_{-z} = y_{-z}$. To apply those results one just needs to replace $\mathbf{p}_z$ by $\mathbf{x}$. Here we highlight that given *some* way to identify a structural function of the form of $\tilde{\pi}_r$, we can identify g. Thus, if a researcher has another means of identifying the structural function $\tilde{\pi}_r$, then this theorem can be applied.

Part (ii) requires sufficiently rich variation in the reduced form profit function $\tilde{\pi}_r$ for *some* value of productivity e^{**}. To further interpret the rank condition, we study it in two parametric examples in Appendix D.3. There we show that the rank condition can be satisfied for the Diewert (1973) profit function, but can fail for every possible parameter value with Cobb-Douglas technology. The reason Cobb-Douglas fails is that its profit function is additively separable when logs are taken.

Remark 2 (Other Degrees of Homogeneity). It is straightforward to generalize our technique to a homogeneous function of any degree $\alpha \geq 0$ since the main identifying equation (2) can be rewritten as

$$\sum_{j=1}^{d_y} \partial_{x_j} \tilde{\pi}(x, e) \frac{g_j(x_j)}{\partial_{x_j} g_j(x_j)} = \alpha \tilde{\pi}(x, e) \,. \tag{3}$$

Here we study the restricted profit function, so $\alpha = 1$, but an analogous equation holds for other homogeneous structural functions. As one example, recall the supply function is homogeneous of degree 0 in prices for a price-taking, profit-maximizing firm.

Remark 3 (Aggregation)*.* The key shape restriction used for identification in this section is homogeneity of a structural function. Importantly, homogeneity is a shape restriction that is preserved under expectations. Note that while we use homogeneity of degree 1 here, this is true for any degree of homogeneity. See in particular Equation 3, which has structure that is preserved under expectations. For this reason, our results work as well with a representative agent analysis involving mean structural demand. We formalize this in Appendix D.6.

3.1. Value as Proxy

This section shows how to interpret Epple et al. (2010) through the lens of price proxies. Specifically, we show that average house values in a market can be used as a proxy for a missing output price. We use this setup as well in the empirical illustration in Appendix C.

Epple et al. (2010) consider the production of housing in which all goods and services provided by a house are treated as a single output. The analyst observes total revenue of selling a house, and the price of land. Variation in these observables is driven by market variation. Importantly, output and its price are *both* unobserved. Each source of unobservability is recognized as an important problem for the measurement of housing production. Building on Epple et al. (2010) we show how average values in a market serve as a price proxy for this missing price.

In contrast to Epple et al. (2010), who work with a representative firm, we study identification in the presence of heterogeneity. As in Epple et al. (2010) we assume constant returns to scale in land and materials, so we can write

$$y_o = f(m, e),$$

where f is the production function per-acre, and output y_o and materials m are in units per acre (land). The production set associated with this production function is $Y(e) = \{(y_o, -m) : y_o \leq f(m, e)\}$. Firms treat land as pre-determined and choose m and y_o. We work with the profit function per-acre, written as

$$\pi(p_o, p_m, p_l, e) = \max_{(y_o, -m) \in Y(e)} p_o y_o - p_m m - p_l,$$

where p_o, p_m, and p_l are prices of output, materials, and land, respectively. Since the price of materials is unobserved, Epple et al. (2010) assume that it is the same across markets and equals 1. We will make the same assumption and drop p_m from the notation.

Since land is pre-determined, its price $\mathbf{p}_l$ does not affect the optimal choice of output or materials. Thus, the value of housing $v(\mathbf{p}_o, \mathbf{e}) = \mathbf{p}_o y_o(\mathbf{p}_o, \mathbf{e})$ and the average value of housing in a market with price $\mathbf{p}_o = p_o$, denoted $\overline{v}(p_o) = \int v(p_o, e) dF_{\mathbf{e}}(e)$, do not depend on price of land $\mathbf{p}_l$. Since $y_o(p_o, e)$ is monotone in p_o, the average value $\overline{v}(p_o)$ is also monotone in p_o. Importantly, $\overline{\mathbf{v}}$ is identified when we observe total revenue $\mathbf{p}_o \mathbf{y}_o$.

Lemma 2. *Suppose the distribution of firm productivity $F_{\mathbf{e}}$ is the same across markets and the other assumptions of this section hold. If $\overline{v}(p_o)$ is strictly increasing in p_o, then average value of housing per market $\overline{\mathbf{v}}$ is a price proxy, i.e. there exists a function g such that*

$$\mathbf{p}_o = g(\overline{\mathbf{v}}) \text{ a.s.}$$

This equation is analogous to Equation 6 in Epple et al. (2010) if we interpret their results as a representative agent analysis.

We note here that by using value as a price proxy for output, if profits were observed and the price of materials ($\mathbf{p}_m$) varied, we could directly use the average value $\overline{\mathbf{v}}$ and identify g using Theorem 2. Here, we do not observe profits and the price of materials is assumed fixed at 1. We thus impose an addition zero-profit assumption as in Epple et al. (2010). While that paper assumes a single type of firm, which attains zero profits, we assume that profits are zero *on average* in a given market[20]:

$$\int \pi(p_o, p_l, e) dF_{\mathbf{e}}(e) = p_o \overline{y}_o(p_o) - \overline{m}(p_o) - p_l = 0,$$

where $\overline{y}_o$ and $\overline{m}$ are the realizations of the aggregate output per-acre and the aggregate

[20] Melitz and Redding (2014) show that free-entry and constant returns of scale imply that ex-ante expected profits are zero, net of entry cost. Here we can assume entry cost is zero. In equilibrium, firms will have zero-profits on average *just before* firms with negative profits leave the market.

demand for materials per-acre in a given market. Since $\mathbf{p}_l$ and $\overline{\mathbf{v}}$ are observed, the equilibrium assumption nonparametrically recovers a revenue function from production minus materials cost (recall that $\mathbf{p}_m = 1$ a.s.),

$$p_l = \tilde{\pi}(\overline{v}) := g(\overline{v})\overline{y}_o(g(\overline{v})) - \overline{m}(g(\overline{v})).$$

Moreover, since $g(\overline{v})\overline{y}_o(g(\overline{v})) = \overline{v}$ by definition, we also identify material costs

$$\tilde{r}(\overline{v}) = -\overline{m}(g(\overline{v})).$$

Similar to Equation 12 in Appendix D.5, we identify the function g since we identify $\tilde{\pi}(\overline{v})$ and $\tilde{\pi}(\overline{v}) - \tilde{r}(\overline{v})$. In particular, g will solve the following differential equation, which is implied by Equation 12:

$$\frac{\partial_{\overline{v}} g(\overline{v})}{g(\overline{v})} = \frac{\partial_{\overline{v}} \tilde{\pi}(\overline{v})}{\tilde{\pi}(\overline{v}) - \tilde{r}(\overline{v})} = \frac{\partial_{\overline{v}} \tilde{\pi}(\overline{v})}{\overline{v}}. \tag{4}$$

Knowing g we can identify $y_o(p_o, e)$ for different levels of heterogeneity since the observed $\mathbf{v}$ is equal to $g(\overline{\mathbf{v}})y_o(g(\overline{\mathbf{v}}), \mathbf{e})$. Thus, our approach generalizes Epple et al. (2010) to allow for unobserved heterogeneity in productivity.

4. Identification of the Production Correspondence

In Section 2, we showed how to identify the restricted profit function allowing endogenous entry and correlation between fixed quantities and productivity, without requiring instruments. Section 3 extends this result to settings when some prices are not observed but the analyst has price proxies, and provides examples of such proxies.

We now focus on how any of these identification results for the restricted profit function can be used to identify the primitive object of interest: the production correspondence. For the sake of notational simplicity from now on, we focus on the profit function though the results can be adapted to the restricted profit function by conditioning.

Recall that we start with identification of the profit function $\pi(p, \cdot)$ only over the support of prices. For notational simplicity, we work with prices and not price

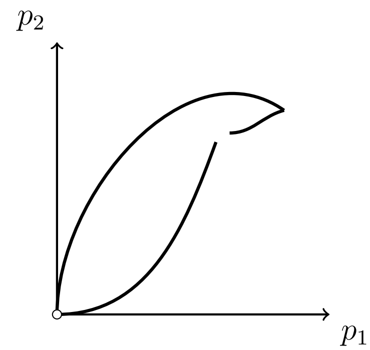

Figure 3 – The set $P(e)$ (depicted by black curve) satisfies Assumption 6 and has an empty interior. Dots represent "holes" in the support. Thus, $P(e)$ is not a connected set.

proxies.[21] The support of prices may consist of all nonnegative numbers, or may be much smaller, i.e. finite. We present a sharp identification result for the production correspondence that covers both cases.

First, we note that $\pi(\cdot, e)$ is homogeneous of degree 1 in prices. It is also convex in prices, hence continuous. These features lead to consideration of the following richness assumption, which ensures $Y(\cdot)$ may be recovered uniquely. Let $P(e)$ denote the conditional support of $\mathbf{p}$ conditional on $\mathbf{e} = e$ (if $\mathbf{p}$ and $\mathbf{e}$ are independent, then $P(e)$ does not vary with e).

Assumption 6.

$$\operatorname{int}\left(\operatorname{cl}\left(\bigcup_{\lambda>0}\{\lambda p \,:\, p \in P(e)\}\right)\right) = \mathbb{R}^{d_y}_{++}$$

for all e, where $\operatorname{cl}(A)$ and $\operatorname{int}(A)$ are the closure and the interior of A, respectively.

The set

$$\bigcup_{\lambda>0}\{\lambda p \,:\, p \in P(e)\}$$

consists of all prices where $\pi(\cdot, e)$ is known because of homogeneity. If that set has "holes," then we can fill them by taking the closure of the set since $\pi(\cdot, e)$ is convex, hence continuous.[22] Assumption 6 means that after we consider the implications of

[21] More generally we can identify the profit function over the support of $g(\mathbf{x})$, where $\mathbf{x}$ is the vector of price proxies.

[22] Beyond continuity, the manner in which convexity affects the data requirements that ensure point identification is subtle, and depends on the shape of $Y(\cdot)$. We provide an illustrative example in Appendix D.4.

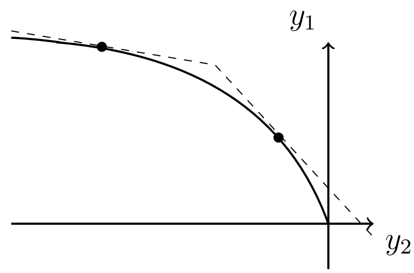

Figure 4 – $\tilde{Y}(e)$ and $Y'(e)$ for $d_y = 2$ and $P(e) = \{p^*, p^{**}\}$. $\tilde{Y}(e)$ is the area under the dashed lines. $Y'(e)$ is the area under the solid curve. Dashed lines correspond to two hyperplanes $p^{*\prime}y = \pi(p^*, e)$ and $p^{**\prime}y = \pi(p^{**}, e)$. They are tangential to the solid curve.

homogeneity and continuity, it is as if we have full variation in prices. Figure 3 is an example of a set satisfying this assumption. Another example is the Cartesian product of all natural numbers, $P(e) = \{1, 2, \ldots\}^{d_y}$. Thus, Assumption 6 does not impose that the support of $\mathbf{p}$ contains an open ball.

Theorem 3. *Let $\pi(p, e)$ be identified by some previous argument over the set $p \in P(e)$ for all e. Moreover, let $\tilde{Y}(\cdot)$ be defined via*

$$\tilde{Y}(e) = \Big\{ y \in \mathbb{R}^{d_y} \ : \ p'y \leq \pi(p, e), \ \forall p \in P(e) \Big\}$$

for all $e \in E$. Then

(i) $\tilde{Y}(\cdot)$ can generate the data and for each $e \in E$, $\tilde{Y}(e)$ is a closed, convex set that satisfies free disposal.[23]

(ii) A production correspondence $Y'(\cdot)$ can generate the data if and only if

$$\max_{y \in Y'(e)} p'y = \max_{y \in \tilde{Y}(e)} p'y$$

for every $e \in E$ and $p \in P(e)$. It follows that for any such $Y'(\cdot)$, $Y'(e) \subseteq \tilde{Y}(e)$, for each $e \in E$.

(iii) If Assumption 6 holds, then $\tilde{Y}(\cdot)$ is the only production correspondence that can generate the data.

[23] By generate the data we mean that the profit function induced by $\tilde{Y}$ agrees with the identified profit function $\pi(p, e)$ for all $e \in E$ and $p \in P(e)$.

Parts (i) and (ii) of Theorem 3 are a sharp identification result stating the most that can be said about the production correspondence under our assumptions. These results are related to Varian (1984), Theorem 15.[24] However, Varian (1984) works only with finite datasets, which are comparable to having a finite support of prices in our setting. In addition, Varian (1984) observes prices and quantities while we observe prices and profits. Recall that observing prices and quantities implies observation of profits. Finally, Varian (1984) does not consider unobservable heterogeneity.

Theorem 3(ii) establishes that $\tilde{Y}(\cdot)$ is the envelope of all production correspondences that can generate the data (see Figure 4). We note, however, that $\tilde{Y}(\cdot)$ may not be a production correspondence because it need not satisfy the recession cone property (recall Definition 1(iii)).[25]

Theorem 3(iii) is related to classic work on the identification of a deterministic production set from a deterministic profit function.[26] In this paper, however, we begin with the distribution of profits and prices. Part (iii) shows that with this distribution, it is possible to identify the distribution of features of $Y(\cdot)$, such as the distribution of possible profit-maximizing quantities. We emphasize that this is true even if quantities are unobservable. An additional manner in which (iii) differs from textbook analysis is that, in econometric settings, it is not always natural to assume that all prices are observed ($P(e) = \mathbb{R}^{d_y}_{++}$). Theorem 3 clarifies the variation in prices sufficient for nonparametric identification of production sets. We note that while Assumption 6 is sufficient for point identification of Y, it is not necessary as illustrated in Appendix D.4.

Remark 4. Our identification analysis does not impose any *a priori* restrictions that certain dimensions of $Y(e)$ correspond to inputs, i.e. weakly negative numbers. This additional restriction can be imposed by modifying the set constructed in Theorem 3. Specifically, the set $\tilde{Y}(e)$ constructed in this theorem may be intersected with an appropriate half-space that encodes that certain dimensions (corresponding to inputs) must be nonpositive. We note that an analogous restriction for outputs is not

[24]The set $\tilde{Y}(e)$ is related to the "outer" set considered in Varian (1984), Section 7. The set $\tilde{Y}(e)$ is constructed from price and profit information, however, rather than price and quantity information as in Varian (1984).

[25]To see this, suppose that a firm of type $e \in E$ has 2-dimensional output/input set, prices are a constant vector $P(e) = \{(1,1)'\}$, and profits at that price are given by $\pi((1,1)', e) = 0$. Then the set $\tilde{Y}(e)$ is $\{y \in \mathbb{R}^2 : y_1 + y_2 \leq 0\}$. This set induces infinite profits for a price-taking firm whenever $p_1 \neq p_2$. Hence, this set violates the recession cone property, which is necessary for the firm problem to have a maximizer since $\tilde{Y}(e)$ is closed and nonempty, e.g. Kreps (2012), Proposition 9.7. Note from part (iii), when Assumption 6 holds it follows that $\tilde{Y}$ is a production correspondence, and thus satisfies the recession cone property.

[26]See e.g. Kreps (2012), Corollary 9.18 for a textbook result.

informative because of the assumption of free disposal.

5. Sharp Counterfactual Bounds

Theorem 3 makes use of a shape restriction to characterize the identified set of the production correspondence for profit-maximizing, price-taking firms. This shape restriction may be used for a dual purpose of providing sharp counterfactual bounds. This follows a long tradition in revealed preference. Varian (1982, 1984) has exploited the close connections between empirical content, recoverability of structural functions, and counterfactuals. Recent work in demand analysis building on these connections includes Blundell, Browning and Crawford (2003), Blundell, Kristensen and Matzkin (2017), Allen and Rehbeck (2019), and Aguiar and Kashaev (2021). In this section we describe a method to bound objects of interest outside of the support of the data.

Since homogeneity and convexity of the heterogeneous profit function allow us to identify it over $\mathrm{cl}\left(\bigcup_{\lambda>0}\{\lambda p \,:\, p \in P(e)\}\right)$, we can associate the conditional support $P(e)$ (of prices condition on $\mathbf{e} = e$) with the set where $\pi(\cdot, e)$ is identified. That is why, for notational simplicity and in this section only, we assume that $P(e)$ is a closed subset of the unit sphere $\mathbb{S}^{d_y-1}$ for all e, and we consider counterfactual prices with norm normalized to 1.

We first present a result characterizing quantities consistent with profit maximization. Theorem 3(ii) is the basis for the following proposition.

Proposition 1. *Let $P(e)$ be a finite subset of the unit sphere $\mathbb{S}^{d_y-1}$. Given $P(e)$ and $\{\pi(p,\cdot)\}_{p\in P(\cdot)}$, the set of output/input functions $\{y_p(\cdot)\}_{p\in P(\cdot)}$ can generate $\{\pi(p,\cdot)\}_{p\in P(\cdot)}$ if and only if*

$$
\begin{aligned}
&p' y_p(e) = \pi(p,e)\,, \quad \forall p \in P(e), e \in E\,,\\
&p^{*\prime} y_{p^*}(e) \geq p^{*\prime} y_p(e)\,, \quad \forall p, p^* \in P(e), e \in E\,.
\end{aligned}
$$

The vector $y_p(e)$ is interpreted as a candidate supply vector given price p and productivity e; it need not be unique and thus may not be equivalent to the supply function. Recall that as discussed in Remark 4, we do not impose *a priori* restrictions that certain components of $Y(e)$ are inputs; this would correspond to imposing additional sign restrictions on the functions $y_p(\cdot)$ described in the proposition.

Proposition 1 essentially states that for each e there must exist output/input vectors such that the weak axiom of profit maximization holds (Varian, 1984). We note, however, that the primitive observables of our paper are the *distribution* of profits and prices.

We can adapt Proposition 1 to answer counterfactual questions by considering a hypothetical tuple (p^c, y_{p^c}) of prices and quantities. If Proposition 1 applies with these additional counterfactual values, then they are feasible given the theory. In more detail, we present bounds on counterfactual objects, potentially with additional restrictions. The counterfactual values involve a function C of interest. The restrictions involve a function s that depends on the counterfactual price p^c and quantity y_{p^c}. We encode the restrictions by the combinations such that $s(p^c, y_{p^c}) = 0$. For instance, if the counterfactual price is fixed to a given vector $\overline{p}^c$ and no restrictions are imposed on y_{p^c}, then $s(p^c, y_{p^c}) = p^c - \overline{p}^c$. The upper bound with heterogeneity level e is given by

$$\begin{aligned}
\overline{C}(e) = &\sup_{p^c, y_{p^c}, \{y_p\}_{p \in P(e)}} C(p^c, y_{p^c})\,, \\
\text{s.t. } & s(p^c, y_{p^c}) = 0\,, \\
& p' y_p = \pi(p, e)\,, \quad \forall p \in P(e)\,, \\
& p^{*\prime} y_{p^*} \geq p^{*\prime} y_p\,, \quad \forall p, p^* \in P(e) \cup \{p^c\}\,.
\end{aligned}$$

The lower bound is given by

$$\begin{aligned}
\underline{C}(e) = &\inf_{p^c, y_{p^c}, \{y_p\}_{p \in P(e)}} C(p^c, y_{p^c})\,, \\
\text{s.t. } & s(p^c, y_{p^c}) = 0\,, \\
& p' y_p = \pi(p, e)\,, \quad \forall p \in P(e)\,, \\
& p^{*\prime} y_{p^*} \geq p^{*\prime} y_p\,, \quad \forall p, p^* \in P(e) \cup \{p^c\}\,.
\end{aligned}$$

We provide some examples covered by this general setup. Note that these bounds hold for each e, and thus one may also bound the distribution of $\overline{C}(\mathbf{e})$ and $\underline{C}(\mathbf{e})$. We reiterate that these upper and lower bounds apply to prices on the unit sphere, though they may be adapted for prices off the unit sphere as illustrated in the following examples.

Example 2 (Profit bounds for a counterfactual price)**.** Suppose that we are interested in upper and lower bounds for profits at a given counterfactual price $\overline{p}^c$. When prices p^c are on the unit sphere, we may specify $C(p^c, y_{p^c}) = p^{c\prime} y_{p^c}$ and $s(p^c, y_{p^c}) = p^c - \overline{p}^c$.

Then the problem can be simplified to get

$$\overline{C}(e) = \sup_{y \in \tilde{Y}(e)} \overline{p}^{c\prime} y\,,$$
$$\underline{C}(e) = \max_{p \in P(e)} \inf_{y \in \tilde{Y}(e)\,:\, p' y = \pi(p,e)} \overline{p}^{c\prime} y\,,$$

where $\tilde{Y}(e)$ is the envelope of all production possibility sets consistent with the data defined in Theorem 3. The above bounds are sharp in the following sense: if $\overline{C}(e)$ is finite, then it is feasible, i.e. there exists a production set that can generate $\overline{C}(e)$. If $\overline{C}(e)$ is not finite, then for any finite level K there exists a production set that can generate $C(p^c, y_{p^c}) > K$. Analogous statements hold for the lower bounds $\underline{C}(e)$. Recall that we assume the support of prices $P(e)$ is a subset of the unit sphere. This may be imposed in empirical settings by replacing prices with normalized prices $\mathbf{p}/\left\|\mathbf{p}\right\|$. For counterfactual questions involving a price off the unit sphere $\overline{p}^c$, one can bound counterfactual profits at price $\overline{p}^c/\left\|\overline{p}^c\right\|$ and then multiply the upper and lower bounds by $\left\|\overline{p}^c\right\|$.

Example 3 (Quantity bounds for a counterfactual price)**.** Suppose that we are interested in the upper and lower bounds for $u' y_{p^c}$ for a given counterfactual price $\overline{p}^c$, where u is a vector. For example, with $u = (1, 0, \ldots, 0)'$ we are interested in bounds on the first component of y. Then $C(p^c, y_{p^c}) = u' y_{p^c}$ and $s(p^c, y_{p^c}) = p^c - \overline{p}^c$.

Example 4 (Profit bounds for a counterfactual quantity)**.** Suppose a regulator is considering imposing a new regulation that the first component of the output/input vector is fixed at $\overline{y}_1^c$. For example, in analysis of health care (Bilodeau et al., 2000) a hospital may be required to treat a certain number of patients. To bound profits we may write the objective function as $C(p^c, y_{p^c}) = p^{c\prime} y_{p^c}$. The constraint is given by $s(p^c, y_{p^c}) = y_{1,p^c} - \overline{y}_1^c$.[27] Bounds on profits with this quantity may be useful for a regulator wondering whether a hospital of type e would be profitable with the hypothetical regulation. If the upper bound on profits is negative, the answer is definitively no. If the lower bound on profits is positive, the answer is definitively yes.[28] An additional question a regulator might ask is which types of firms could still be profitable. This can be addressed by studying functions $\overline{C}(\cdot)$ and $\underline{C}(\cdot)$ as e varies. Note that the constraints s are general, and inequality constraints may be incorporated as well by using indicator functions.

[27] Note that the problem may not have a solution since the set of parameters that satisfy restrictions may be empty.

[28] This maintains the assumptions of price-taking, profit-maximizing behavior with a technology that is described by a production correspondence.

When $P(e)$ is finite, computing bounds in Examples 2 and 3 is straightforward since they are the values of linear programs. Example 4 is also a linear program if we add the additional constraint that the counterfactual price is fixed, $p^c = \overline{p}^c$. In general, the computational difficulty of the bounds $\overline{C}$ and $\underline{C}$ depends on the nature of the objective function and the constraint.

6. Estimation of Production Sets and Consistency

The previous identification results describe how to identify the profit or restricted profit function. Appendix B describes one estimator of the restricted profit function, but there are many depending on assumptions concerning exogeneity or whether productivity is discrete or continuous. This section links *any* estimator of the restricted profit function to an induced estimator of the corresponding production set. As in previous section, for notational convenience we work with the profit function, though the analysis applies to the restricted profit function by conditioning. In the restricted case, we would instead estimate the restricted production correspondence.

We now describe how an estimator $\hat{\pi}(\cdot, e)$ of the profit function may be used to construct an estimator $\hat{Y}(e)$ of the production possibility set for a firm with productivity level e. The main result in this section relates the estimation error of $\hat{\pi}$ (for π) and that of the constructed set $\hat{Y}$ (for Y). Consistency and rates of convergence results for $\hat{\pi}$ thus have analogous statements for $\hat{Y}$.

As setup, we now formalize our notions of distance both for functions and sets. We present our result for a fixed $e \in E$. We assume that $\pi(\cdot, e)$ is identified over $P(e) = P = \mathbb{R}^{d_y}_{++}$ (we assume Assumption 6). Given a fixed $e \in E$ and $\hat{\pi}(\cdot, e)$, a natural estimator for $Y(e)$ is

$$\hat{Y}(e) = \left\{ y \in \mathbb{R}^{d_y} \ : \ p'y \leq \hat{\pi}(p, e), \forall p \in P \right\} .$$

This set is a plug-in estimator motivated by Theorem 3. A commonly used notion of distance between convex sets is the Hausdorff distance. The Hausdorff distance between two convex sets $A, B \subseteq \mathbb{R}^{d_y}$ is given by

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} \|a - b\| , \sup_{b \in B} \inf_{a \in A} \|a - b\| \right\} .$$

Unfortunately, the Hausdorff distance between $Y(e)$ and $\hat{Y}(e)$ can be infinite. For this reason we will consider the Hausdorff distance between certain extensions of these sets. The following example illustrates why the original distance may be infinite.

Example 5. Suppose that $d_y = 2$ and for some $e \in E$,

$$\begin{aligned} Y(e) &= \left\{ y \in \mathbb{R} \times \mathbb{R}_- \ : \ y_1 \leq \sqrt{-y_2} \right\} , \\ \hat{Y}^m(e) &= \left\{ y \in \mathbb{R} \times \mathbb{R}_- \ : \ y_1 \leq (1 - 1/m)\sqrt{-y_2} \right\} , \quad m \in \mathbb{N}. \end{aligned}$$

Note that although $\lim_{m\to\infty}(1 - 1/m)\sqrt{-y_2} = \sqrt{-y_2}$ for every finite $y_2 \leq 0$, the Hausdorff distance between these sets is infinite for every finite $m \in \mathbb{N}$.

Example 5 illustrates a technical concern with the Hausdorff distance that arises because of the unboundedness of production possibility sets. However, in empirical applications one may be interested in production possibility sets in regions that correspond to prices that are bounded away from zero. Thus, instead of working with all possible prices we will work only with certain empirically relevant compact convex subsets of $\mathbb{R}^{d_y}_{++}$. We consider the Hausdorff distance between extensions such as

$$\begin{aligned} Y_{\bar{P}}(e) &= \left\{ y \in \mathbb{R}^{d_y} \ : \ p'y \leq \pi(p, e), \ \forall p \in \bar{P} \right\} \\ \hat{Y}_{\bar{P}}(e) &= \left\{ y \in \mathbb{R}^{d_y} \ : \ p'y \leq \hat{\pi}(p, e), \ \forall p \in \bar{P} \right\} , \end{aligned}$$

where $\bar{P} \subseteq P$ is convex and compact. These sets nest the original sets (e.g. $Y(e) \subseteq Y_{\bar{P}}(e)$) because the inequalities hold only for $p \in \bar{P}$, not for every $p \in P$. Moreover, the parts of the production possibility frontiers of the sets $Y(e)$ and $Y_{\bar{P}}(e)$ coincide at points that are tangential to price vectors from $\bar{P}$ (see Figure 5).

We now turn to the main result in this section, which establishes an equality relating the distance between $\hat{\pi}$ and π, and the distance between extensions of $\hat{Y}$ and Y. Our distance for these profit functions is given by

$$\tilde{d}_{\bar{P}}(e) = \sup_{p \in \bar{P}} \left\| \frac{\hat{\pi}(p, e) - \pi(p, e)}{\|p\|} \right\| .$$

To state the following result, let $\bar{\mathcal{P}}$ be a collection of all compact, convex, and nonempty subsets of P.

Theorem 4. *Maintain the assumption that $\pi(\cdot, e)$ is homogeneous of degree 1 and convex.*[29] *Suppose, moreover, that for every $e \in E$, $\hat{\pi}(\cdot, e)$ is an estimator of $\pi(\cdot, e)$*

[29]Recall that this is equivalent to price-taking, profit-maximizing behavior with technology described

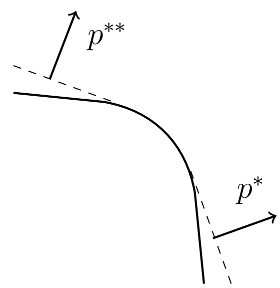

Figure 5 – $Y(e)$ and $Y_{\bar{P}}(e)$ for $d_y = 2$ and $\bar{P} = \{p \in P \,:\, \delta \le p_2/p_1 \le 1/\delta,\, \|p\| \le 1\}$, $0 < \delta < 1$. $Y(e)$ is the area under the solid curve. $Y_{\bar{P}}(e)$ is the area under the dashed lines. Dashed lines correspond to two hyperplanes $p^{*\prime}y = \pi(p^*, e)$ and $p^{**\prime}y = \pi(p^{**}, e)$. They are tangential to the solid curve. p^* is such that $p^*_2/p^*_1 = \delta$ and p^{**} is such that $p^{**}_2/p^{**}_1 = 1/\delta$.

that is homogeneous of degree 1 *and continuous. If* $\hat{\pi}(\cdot, e)$ *is convex, then*

$$d_H(Y_{\bar{P}}(e), \hat{Y}_{\bar{P}}(e)) = \tilde{d}_{\bar{P}}(e) \quad \text{a.s.}$$

for every $\bar{P} \in \bar{\mathcal{P}}$.

Theorem 4 is a nontrivial extension of a well-known relation between the Hausdorff distance and the support functions of convex *compact* sets to convex, closed, and *unbounded* sets.[30] Homogeneity of an estimator can be imposed by rescaling the data by dividing by one of the prices. Unfortunately, convexity can be more challenging to impose and so we turn to a related result that covers cases in which $\hat{\pi}$ is not convex. To formalize our result, we introduce two additional parameters:

$$R_{\bar{P}}(e) = \sup_{p \in \bar{P}} \frac{\pi(p, e)}{\|p\|}, \quad r_{\bar{P}}(e) = \inf_{p \in \bar{P}} \frac{\pi(p, e)}{\|p\|}.$$

Proposition 2. *Maintain the assumption that* $\pi(\cdot, e)$ *is homogeneous and convex. Suppose, moreover, that for every* $e \in E$, $\hat{\pi}(\cdot, e)$ *is an estimator of* $\pi(\cdot, e)$ *that is homogeneous of degree* 1 *and continuous. If* $\tilde{d}_{\bar{P}}(e) = o_p(1)$ *and* $0 < r_{\bar{P}}(e) < R_{\bar{P}}(e) < \infty$*, then*

$$d_H(Y_{\bar{P}}(e), \hat{Y}_{\bar{P}}(e)) \le \tilde{d}_{\bar{P}}(e) \frac{R_{\bar{P}}(e)}{r_{\bar{P}}(e)} \frac{1 + \tilde{d}_{\bar{P}}(e)/R_{\bar{P}}(e)}{1 - \tilde{d}_{\bar{P}}(e)/r_{\bar{P}}(e)}$$

by a production correspondence.

[30] See Kaido and Santos (2014) for a recent application of this result for convex compact sets.

with probability approaching 1*, for every* $\bar{P} \in \bar{\mathcal{P}}$*. In particular,*

$$d_H(Y_{\bar{P}}(e), \hat{Y}_{\bar{P}}(e)) = o_p(1)\,.$$

7. Conclusion

In this paper we provide an update to classical duality theory in order to identify heterogeneous production sets in the presence of endogeneity, measurement error, omitted prices, and unobservable quantities. Our framework's main strength is to unpack rich heterogeneity as well as rich substitution/complementarity patterns with market level variation, using values of optimization problems. We achieve this by exploiting all shape constraints imposed by the economic environment we consider. This includes a key restriction that firms can be ranked in terms of productivity, and there are finitely many types of firms. Our identification results are constructive and can be applied in many available data sets.

References

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer (2015) "Identification properties of recent production function estimators," *Econometrica*, 83 (6), 2411–2451.

Afriat, Sidney N (1972) "Efficiency estimation of production functions," *International economic review*, 568–598.

Aguiar, Victor H and Nail Kashaev (2021) "Stochastic revealed preferences with measurement error," *The Review of Economic Studies*, 88 (4), 2042–2093.

Albouy, David and Gabriel Ehrlich (2018) "Housing productivity and the social cost of land-use restrictions," *Journal of Urban Economics*.

Allen, Roy and John Rehbeck (2019) "Identification with additively separable heterogeneity," *Econometrica*, 87 (3), 1021–1054.

Andrews, Donald WK (2011) “Examples of L2-complete and boundedly-complete distributions,” *Mimeo.*

Arcidiacono, Peter and Robert A Miller (2011) “Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity,” *Econometrica*, 79 (6), 1823–1867.

Beresteanu, Arie, Ilya Molchanov, and Francesca Molinari (2011) “Sharp identification regions in models with convex moment predictions,” *Econometrica*, 79 (6), 1785–1821.

Beresteanu, Arie and Francesca Molinari (2008) “Asymptotic properties for a class of partially identified models,” *Econometrica*, 76 (4), 763–814.

Berger, Allen N, Diana Hancock, and David B Humphrey (1993) “Bank efficiency derived from the profit function,” *Journal of Banking & Finance*, 17 (2-3), 317–347.

Bilodeau, Daniel, Pierre-Yves Cremieux, and Pierre Ouellette (2000) “Hospital cost function in a non-market health care system,” *Review of Economics and Statistics*, 82 (3), 489–498.

Blundell, Richard W, Martin Browning, and Ian A Crawford (2003) “Nonparametric Engel curves and revealed preference,” *Econometrica*, 71 (1), 205–240.

Blundell, Richard W, Dennis Kristensen, and Rosa Liliana Matzkin (2017) “Individual counterfactuals with multidimensional unobserved heterogeneity,”Technical report, cemmap working paper.

Bonhomme, Stéphane and Elena Manresa (2015) “Grouped patterns of heterogeneity in panel data,” *Econometrica*, 83 (3), 1147–1184.

Brunel, Victor-Emmanuel (2016) “Concentration of the empirical level sets of Tukey’s halfspace depth,” *Probability Theory and Related Fields*, 1–32.

Burke, Marshall, Lauren Falcao Bergquist, and Edward Miguel (2019) “Sell low and buy high: arbitrage and local price effects in Kenyan markets,” *The Quarterly Journal of Economics*, 134 (2), 785–842.

Cherchye, Laurens, Thomas Demuynck, Bram De Rock, and Kristof De Witte (2014) “Non-parametric Analysis of Multi-output Production with Joint Inputs,” *The Economic Journal*, 124 (577), 735–775.

Cherchye, Laurens, Thomas Demuynck, Bram De Rock, and Marijn Verschelde (2018) "Nonparametric identification of unobserved technological heterogeneity in production," Working Paper Research 335, National Bank of Belgium, https://EconPapers.repec.org/RePEc:nbb:reswpp:201802-335.

Cherchye, Laurens, Bram De Rock, and Barnabe Walheer (2016) "Multi-output profit efficiency and directional distance functions," *Omega*, 61, 100 – 109, https://doi.org/10.1016/j.omega.2015.07.010.

Chernozhukov, Victor, Alfred Galichon, Marc Henry, and Brendan Pass (2017) "Single market nonparametric identification of multi-attribute hedonic equilibrium models," *arXiv preprint arXiv:1709.09570*.

Chernozhukov, Victor and Christian Hansen (2005) "An IV model of quantile treatment effects," *Econometrica*, 73 (1), 245–261.

Chiong, Khai Xiang, Alfred Galichon, and Matt Shum (2016) "Duality in dynamic discrete-choice models," *Quantitative Economics*, 7 (1), 83–115.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2021) "The production function for housing: Evidence from France," *Journal of Political Economy*, 129 (10), 2766–2816.

Cunha, Flavio, James J Heckman, and Susanne M Schennach (2010) "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, 78 (3), 883–931.

De Loecker, Jan, Pinelopi K Goldberg, Amit K Khandelwal, and Nina Pavcnik (2016) "Prices, markups, and trade reform," *Econometrica*, 84 (2), 445–510.

D'Haultfœuille, Xavier (2011) "On the completeness condition in nonparametric instrumental problems," *Econometric Theory*, 27 (3), 460–471.

D'Haultfœuille, Xavier and Philippe Février (2015) "Identification of mixture models using support variations," *Journal of Econometrics*, 189 (1), 70–82.

Diewert, W Erwin (1982) "Duality approaches to microeconomic theory," *Handbook of mathematical economics*, 2, 535–599.

Diewert, W.E (1973) "Functional forms for profit and transformation functions," *Journal of Economic Theory*, 6 (3), 284 – 316, https://doi.org/10.1016/0022-0531(73)90051-3.

Doraszelski, Ulrich and Jordi Jaumandreu (2013) "R&D and productivity: Estimating endogenous productivity," *Review of Economic Studies*, 80 (4), 1338–1383.

——— (2018) "Measuring the bias of technological change," *Journal of Political Economy*, 126 (3), 1027–1084.

Ekeland, Ivar, James J Heckman, and Lars Nesheim (2004) "Identification and estimation of hedonic models," *Journal of political economy*, 112 (S1), S60–S109.

Epple, Dennis, Brett Gordon, and Holger Sieg (2010) "A new approach to estimating the production function for housing," *American Economic Review*, 100 (3), 905–24.

Fabrizio, Kira R., Nancy L. Rose, and Catherine D. Wolfram (2007) "Do Markets Reduce Costs? Assessing the Impact of Regulatory Restructuring on US Electric Generation Efficiency," *American Economic Review*, 97 (4), 1250–1277, 10.1257/aer.97.4.1250.

Farrell, Michael James (1957) "The measurement of productive efficiency," *Journal of the Royal Statistical Society: Series A (General)*, 120 (3), 253–281.

Foster, Lucia, John Haltiwanger, and Chad Syverson (2008) "Reallocation, firm turnover, and efficiency: selection on productivity or profitability?" *American Economic Review*, 98 (1), 394–425.

Fox, Jeremy T and Amit Gandhi (2016) "Nonparametric identification and estimation of random coefficients in multinomial choice models," *The RAND Journal of Economics*, 47 (1), 118–139.

Fuss, Melvyn and Daniel McFadden eds. (1978) *Production Economics: A Dual Approach to Theory and Applications: Applications of the Theory of Production*: Elsevier.

Galichon, Alfred and Bernard Salanié (2015) "Cupid's invisible hand: Social surplus and identification in matching models."

Gandhi, Amit, Salvador Navarro, and David A Rivers (2020) "On the identification of gross output production functions," *Journal of Political Economy*, 128 (8), 2973–3016.

Glaeser, Edward L, Joseph Gyourko, and Raven E Saks (2005) "Why have housing prices gone up?" *American Economic Review*, 95 (2), 329–333.

Grieco, Paul LE and Ryan C McDevitt (2016) “Productivity and quality in health care: Evidence from the dialysis industry,” *The Review of Economic Studies*, 84 (3), 1071–1105.

Griliches, Zvi and Jacques Mairesse (1995) “Production functions: the search for identification,”Technical report, National Bureau of Economic Research.

Hanoch, Giora and Michael Rothschild (1972) “Testing the assumptions of production theory: a nonparametric approach,” *Journal of Political Economy*, 80 (2), 256–275.

Hu, Yingyao and Geert Ridder (2010) “On deconvolution as a first stage nonparametric estimator,” *Econometric Reviews*, 29 (4), 365–396.

Hu, Yingyao, Susanne M Schennach, and Ji-Liang Shiu (2017) “Injectivity of a class of integral operators with compactly supported kernels,” *Journal of Econometrics*, 200 (1), 48–58.

Kaido, Hiroaki (2016) “A dual approach to inference for partially identified econometric models,” *Journal of econometrics*, 192 (1), 269–290.

Kaido, Hiroaki, Francesca Molinari, and Jörg Stoye (2019) “Confidence intervals for projections of partially identified parameters,” *Econometrica*, 87 (4), 1397–1432.

Kaido, Hiroaki and Andres Santos (2014) “Asymptotically efficient estimation of models defined by convex moment inequalities,” *Econometrica*, 82 (1), 387–413.

Katayama, Haijime, Shihua Lu, and James R Tybout (2003) “Why plant-level productivity studies are often misleading, and an alternative approach to interference.”

Kim, Youngseok, Peter Carbonetto, Matthew Stephens, and Mihai Anitescu (2020) “A fast algorithm for maximum likelihood estimation of mixture proportions using sequential quadratic programming,” *Journal of Computational and Graphical Statistics*, 29 (2), 261–273.

Kitamura, Yuichi and Louise Laage (2018) “Nonparametric Analysis of Finite Mixtures.”

Klette, Tor Jakob and Zvi Griliches (1996) “The inconsistency of common scale estimators when output prices are unobserved and endogenous,” *Journal of applied econometrics*, 11 (4), 343–361.

Kreps, David M (2012) *Microeconomic foundations I: choice and competitive markets*, 1: Princeton university press.

Kriegel, Hans-Peter, Peer Kröger, Jörg Sander, and Arthur Zimek (2011) "Density-based clustering," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1 (3), 231–240.

Levinsohn, James and Amil Petrin (2003) "Estimating production functions using inputs to control for unobservables," *The Review of Economic Studies*, 70 (2), 317–341.

Li, Tong and Yuya Sasaki (2017) "Constructive Identification of Heterogeneous Elasticities in the Cobb-Douglas Production Function," *arXiv preprint arXiv:1711.10031*.

López, Marco and Georg Still (2007) "Semi-infinite programming," *European Journal of Operational Research*, 180 (2), 491–518.

Matzkin, Rosa L (1992) "Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models," *Econometrica: Journal of the Econometric Society*, 239–270.

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

McFadden, Daniel (1981) "Econometric models of probabilistic choice," *Structural analysis of discrete data with econometric applications*, 198272.

Melitz, Marc J and Stephen J Redding (2014) "Heterogeneous firms and trade," in *Handbook of international economics*, 4, 1–54: Elsevier.

Nerlove, Marc (1963) "Returns to Scale in Electricity Supply," *Measurement in Economics*.

Newey, Whitney K and James L Powell (2003) "Instrumental variable estimation of nonparametric models," *Econometrica*, 71 (5), 1565–1578.

Olley, G Steven and Ariel Pakes (1996) "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64 (6), 1263–1297.

Pakes, Ariel (1996) "Dynamic structural models, problems and prospects: mixed continuous discrete controls and market interaction," in *Advances in Econometrics, Sixth World Congress*, 2, 171–259, by C. Sims.

Ray, Debraj ⓡ Arthur Robson (2018) "Certified random: A new order for coauthorship," *American Economic Review*, 108 (2), 489–520.

Roberts, Mark J and Dylan Supina (1996) "Output price, markups, and producer size," *European Economic Review*, 40 (3-5), 909–921.

Rockafellar, R Tyrrell (1970) *Convex Analysis*: Princeton University Press.

Schennach, Susanne M (2016) "Recent advances in the measurement error literature," *Annual Review of Economics*, 8, 341–377.

Shephard, Ronald William (1953) *Cost and production functions*: Princeton University Press.

Varian, Hal R (1982) "The nonparametric approach to demand analysis," *Econometrica: Journal of the Econometric Society*, 945–973.

——— (1984) "The nonparametric approach to production analysis," *Econometrica: Journal of the Econometric Society*, 579–597.

Zellner, Arnold, Jan Kmenta, and Jacques Dreze (1966) "Specification and estimation of Cobb-Douglas production function models," *Econometrica: Journal of the Econometric Society*, 784–795.

A. Proofs of Main Results

A.1. Proof of Lemma 1

Fix y^*_{-z} and p^*_z. By homogeneity of degree 1 of the restricted profit function in prices and Assumption 1,

$$\Delta\pi_r(y^*_{-z}, \lambda p^*_z, e) = \lambda\Delta\pi_r(y^*_{-z}, p^*_z, e) > 0$$

for every e and $\lambda > 0$. Since $\cup_{\lambda>0}\{\lambda p^*_z\}$ in the conditional support, we always can find λ large enough and e^* such that Assumption 4(ii) is satisfied.

A.2. Proof of Theorem 1

First, note that since the support of $\boldsymbol{\eta}$ is a connected set (Assumption 4(i)) and $\mathbf{e}$ is discrete, the conditional support of $\boldsymbol{\pi}_r$ conditional on $\mathbf{y}_{-z} = y_{-z}$ and $\mathbf{p}_z = p_z$ is a union of connected sets for all y_{-z} and p_z in their joint support. Hence, we can find the shortest (with respect to Lebesgue measure) isolated connected segment of the support for every y_{-z} and p_z. Next, among those short segments we can find the shortest one. By construction this segment will correspond to (y^*_{-z}, p^*_z, e^*) from Assumption 4(ii). As a result, under Assumption 4, we can find an interval $[a, b]$ in the support of $\boldsymbol{\pi}_r$ conditional on $\mathbf{y}_{-z} = y^*_{-z}$, $\mathbf{e} = e^*$, and $\mathbf{p}_z = p^*_z$ such that

$$\mathbb{P}\left(a \le \pi_r(y^*_{-z}, p^*_z, e^*) + \boldsymbol{\eta} \le b\right) = 1$$

and

$$\mathbb{P}\left(a \le \pi_r(y^*_{-z}, p^*_z, e) + \boldsymbol{\eta} \le b\right) = 0$$

for any $e \neq e^*$. Hence, we identify

$$\pi_r(y^*_{-z}, p^*_z, e^*) = \mathbb{E}\left[\boldsymbol{\pi}_r | a \le \boldsymbol{\pi}_r \le b, \mathbf{y}_{-z} = y^*_{-z}, \mathbf{e} = e^*, \mathbf{p}_z = p^*_z\right],$$

where we leverage that $\boldsymbol{\eta}$ has mean zero even after conditioning.

Thus, we can also recover the distribution of $\boldsymbol{\eta}$ by subtracting the identified $\pi_r(y^*_{-z}, p^*_z, e^*)$ from the known distribution of $\boldsymbol{\pi}_r | a \le \boldsymbol{\pi}_r \le b, \mathbf{y}_{-z} = y^*_{-z}, \mathbf{e} = e^*, \mathbf{p}_z = p^*_z$. Since $\boldsymbol{\eta}$ and $\pi_r(\mathbf{y}_{-z}, \mathbf{p}_z, \mathbf{e})$ have bounded support and are independent conditional on $\mathbf{y}_{-z} = y_{-z}$ and $\mathbf{p}_z = p_z$, we can constructively identify the moment generating function of $\pi_r(\mathbf{y}_{-z}, \mathbf{p}_z, \mathbf{e})$ conditional on $\mathbf{y}_{-z} = y_{-z}$ and $\mathbf{p}_z = p_z$ as the ratio of the moment generating functions of $\boldsymbol{\pi}_r$ conditional on $\mathbf{y}_{-z} = y_{-z}$ and $\mathbf{p}_z = p_z$ and $\boldsymbol{\eta}$. Since the distribution of $\pi_r(\mathbf{y}_{-z}, \mathbf{p}_z, \mathbf{e})$ conditional on $\mathbf{y}_{-z} = y_{-z}$ and $\mathbf{p}_z = p_z$ is discrete, its moment generating function is sufficient for its identification. Note that the moment generating function of $\boldsymbol{\eta}$ is well-defined and is never equal to zero since $\boldsymbol{\eta}$ is a bounded random variable.

Assumption 3 implies that whenever a type e occurs with positive probability conditional on y_{-z} and p_z, then higher types also occur with positive probability. Assumption 1 then implies that the ranking over restricted profits is equivalent to the ranking over productivity e. As a result, if some firm of type e does not operate given y_{-z} and p_z, then it has to be a low type. Let $\Pi_r(y_{-z}, p_z)$ be the support of $\pi_r(\mathbf{y}_{-z}, \mathbf{p}_z, \mathbf{e})$ conditional on $\mathbf{y}_{-z} = y_{-z}$ and $\mathbf{p}_z = p_z$. Fix some y_{-z} and p_z. Since the support of $\mathbf{e}$ is finite, the set $\Pi_r(y_{-z}, p_z)$ will also be finite. As a result, Assumption 1 implies that

$$\pi_r(y_{-z}, p_z, d_e) = \max\left[\Pi_r(y_{-z}, p_z)\right].$$

That is, the most productive firm will make more profits than any other firm. Note that the firm with productivity $e = d_e - 1$, if it is present in the market, will be the second one in terms of restricted profits :

$$\pi_r(y_{-z}, p_z, d_e - 1) = \max\left[\Pi_r(y_{-z}, p_z, s) \setminus \{\pi_r(y_{-z}, p_z, d_e)\}\right].$$

In general, given y_{-z} and p_z, if the firm with productivity e operates ($|\Pi_r(y_{-z}, p_z)| > d_e - e$), then

$$\pi_r(y_{-z}, p_z, e) = \max\left[\Pi_r(y_{-z}, p_z) \setminus \bigcup_{e'>e}\{\pi_r(y_{-z}, p_z, e')\}\right].$$

Note that we may not be able to identify the structural restricted profit function for arguments in which e is too low.

A.3. Proof of Theorems 2

Fix some x_{-1}, and take y_z^* from the statement of the theorem and $e^{**} \in E$ from condition (ii). We abuse notation and drop e^{**} and y_{-z}^*. By homogeneity of degree 1 of $\pi_r(\cdot)$ we have that for every x

$$\sum_{j=1}^{d_{yz}} \partial_{g_j}\pi_r(g(x))g_j(x_j) = \pi_r(g(x)). \tag{5}$$

Moreover, since $\tilde{\pi}_r(x) = \pi_r(g(x))$ (recall that we dropped e^{**} and y_{-z}^* from the notation) and $\partial_{x_j}g_k(x_k) = 0$ for $j \neq k$, we have that

$$\partial_{x_j}\tilde{\pi}_r(x) = \sum_k \partial_{g_k}\pi_r(g(x))\partial_{x_j}g_k(x_k) = \partial_{g_j}\pi_r(g(x))\partial_{x_j}g_j(x_j), \tag{6}$$

for every $j = 1, \ldots, d_{y_z}$. Combining (5) and (6) we get that

$$\sum_{j=1}^{d_{yz}} \partial_{x_j}\tilde{\pi}_r(x)\frac{1}{\partial_{x_j}(\log(g_j(x_j)))} = \tilde{\pi}_r(x)$$

as long as $0 < \left|\dfrac{\partial_{x_j} g_j(x_j)}{g_j(x_j)}\right| < \infty$ for every $j = 1, \dots, d_{y_z}$. This latter condition is satisfied for almost every x_j with respect to Lebesgue measure by Assumption 5(iv). From Assumption 5(i), $g_1(x_1) = x_1$, so we obtain that

$$\sum_{j=2}^{d_{y_z}} \partial_{x_j} \tilde{\pi}_r(x) \frac{1}{\partial_{x_j}(\log(g_j(x_j)))} = \tilde{\pi}_r(x) - \partial_{x_1} \tilde{\pi}_r(x) x_1. \tag{7}$$

Let $\tilde{t} = \left(\dfrac{1}{\partial_{x_j}(\log(g_j(x_j)))}\right)_{j=2,\dots,d_{y_z}}$. Note that $\tilde{t}$ does not depend on x_1. Since $\tilde{\pi}_r$ satisfies the rank condition there exists a nonsingular $A(\tilde{\pi}_r(x^*))$ and b such that equation (7) can be rewritten as

$$A\tilde{t} = b\,, \tag{8}$$

where $b = (b_l)_{l=1,\dots,d_{y_z}-1}$ and $b_l = \tilde{\pi}_r(x_l^*) - \partial_{x_1} \tilde{\pi}_r(x_l^*) t_l$. Since $A(\tilde{\pi}_r(x^*))$ is of full rank and is identified, and b is identified, $\tilde{t}$ is identified. Since the choice of x_{-1} was arbitrary and we know the location (Assumption 5(ii)), we identify $g_j(\cdot)$ for every $j = 1, \dots, d_{y_z}$.

A.4. Proof of Theorem 3

It is immediate that $\tilde{Y}(e)$ is closed, convex, and satisfies free disposal for every $e \in E$. Moreover, $\max_{y \in \tilde{Y}(e)} p'y = \pi(p, e)$ for every $p \in P(e)$ and $e \in E$. Thus, conclusion (i) follows from the fact that $\pi(p, e)$ is identified for each $p \in P(e)$ and $e \in E$ by Theorem 1.

To establish conclusion (ii), recall that under the assumptions of Theorem 1, any given production set $Y'(e)$ can generate the data if and only if $\max_{y \in Y'(e)} p'y = \pi(p, e)$ for every $p \in P(e)$. The set $\tilde{Y}(e)$ is constructed as the largest set (not necessary production set) consistent with profit maximization. This set is closed, convex, and satisfies free disposal. Since a production correspondence *also* must satisfy the recession cone property, we obtain that $Y'(e) \subseteq \tilde{Y}(e)$.

To prove (iii), note that since $\pi(\cdot, e)$ is homogeneous of degree 1 for every $e \in E$ we can identify $\pi(\cdot, e)$ over

$$\bigcup_{\lambda > 0} \{\lambda p \,:\, p \in P(e)\}\,.$$

Next, since $\pi(\cdot, e)$ is convex it is continuous, hence it is identified over

$$\operatorname{int}\left(\operatorname{cl}\left(\bigcup_{\lambda>0}\{\lambda p \,:\, p \in P(e)\}\right)\right) .$$

When Assumption 6 holds, identification of $Y(\cdot)$ follows from Corollary 9.18 in Kreps (2012).

A.5. Proof of Proposition 1

Fix some $e \in E$. To simplify notation we drop e from the objects below (e.g. $\pi(p,e) = \pi(p)$ and $y_p(e) = y_p$). Suppose $\{y_p\}_{p\in P}$ can generate $\{\pi(p)\}_{p\in P}$. Since $\{y_p\}_{p\in P}$ are profit-maximizing output/input vectors we must have $p'y_p = \pi(p)$. To prove that $p^{*\prime}y_{p^{*\prime}} \geq p^{*\prime}y_p$ for all $p, p^* \in P$, assume the contrary. But then y_{p_*} is not maximizing profits at p^* since y_p is available. The contradiction proves necessity.

To prove sufficiency consider

$$Y^* = \operatorname{co}(\{y_p\}_{p\in P}) + \mathbb{R}_-^{d_y} ,$$

where $\operatorname{co}(A)$ denotes the convex hull of a set A, i.e. the smallest convex set containing A. The summation is the Minkowski sum. Y^* is sometimes referred to as the free-disposal convex hull of $\{y_p\}_{p\in P}$. In particular, note that Y^* is convex, closed, and satisfies free disposal.

We obtain that for every $p \in \mathbb{R}_{++}^{d_y} \cap \mathbb{S}^{d_y-1}$,

$$\sup_{y\in Y^*} p'y = \sup_{y\in\operatorname{co}(\{y_p\}_{p\in P})} p'y + \sup_{y\in\mathbb{R}_-^{d_y}} p'y = \sup_{y\in\operatorname{co}(\{y_p\}_{p\in P})} p'y .$$

Because P is finite, $\{y_p\}_{p\in P}$ is bounded. Thus, its convex hull $\operatorname{co}(\{y_p\}_{p\in P})$ is also bounded. This implies that $\sup_{y\in Y'} p'y$ is finite for every $p \in \mathbb{R}_{++}^{d_y} \cap \mathbb{S}^{d_y-1}$, hence the recession cone property is satisfied for the set Y^*.[31]

It is left to show that

$$\pi(p,e) = p'y_p = \sup_{y\in Y^*} p'y$$

for every $p \in P \cap \mathbb{S}^{d_y-1}$. The first equality is assumed. Suppose the second equality is

[31] We note that Varian (1984) studies a result related to this proposition, taking as primitives a deterministic dataset of prices and quantities. He does not verify the recession cone property.

not true for some p^*. Then there exists $\tilde{y} \in Y^*$ such that $p^{*\prime} y_{p^*} < p^{*\prime}\tilde{y}$. Since $\tilde{y} \in Y^*$ it can be represented as a finite convex combination of points from $\{y_p\}_{p\in P}$. But since

$$p^{*\prime} y_{p^*} \geq p^{*\prime} y_p \,,$$

for all $p, p^* \in P$ it has to be the case that

$$p^{*\prime} y_{p^*} \geq p^{*\prime}\tilde{y}.$$

The contradiction completes the proof. Since the choice of e was arbitrary the result holds for all $e \in E$.

A.6. Proof of Theorem 4 and Proposition 2

The Hausdorff distance between two convex sets $A, B \subseteq \mathbb{R}^{d_y}$ is given by

$$d_H(A, B) = \max\left\{\sup_{a\in A}\inf_{b\in B}\|a-b\|, \sup_{b\in B}\inf_{a\in A}\|a-b\|\right\}.$$

Alternatively, the Hausdorff distance can be defined as

$$d_H(A, B) = \inf\{\rho \geq 0 \;:\; A \subseteq B + \rho\mathbb{B}^{d_y-1}, B \subseteq A + \rho\mathbb{B}^{d_y-1}\}\,,$$

where $\mathbb{B}^{d_y-1} = \{y \in \mathbb{R}^{d_y} \;:\; \|y\| \leq 1\}$ is the unit ball and $\inf\{\emptyset\} = \infty$. The support function of a closed convex set A is defined for $u \in \mathbb{R}^{d_y}$ via $h_A(u) = \sup_{w\in A} u'w$. If A is unbounded in direction u, then $h_A(u) = \infty$.

As preparation, we need a technical lemma. This lemma involves a polar cone, which for a set C is defined by

$$\mathrm{PolCon}(C) = \{u \in \mathbb{R}^{d_y} \;:\; u'p \leq 0, \forall p \in C\}.$$

Lemma A.1. *Let $\bar{P} \subseteq \mathbb{S}^{d_y-1}$ be a closed set such that $\cup_{\lambda>0}\{\lambda p, p \in \bar{P}\}$ is a closed, convex cone, and let $a : \mathbb{R}^{d_y} \to \mathbb{R}$ be a convex, homogeneous of degree 1 function. Define*

$$A = \{y \in \mathbb{R}^{d_y} \;:\; p'y \leq a(p), \forall p \in \bar{P}\}.$$

If $\text{PolCon}(\bar{P})$ *is nonempty, then for any* $u \in \mathbb{S}^{d_y - 1}$,

$$h_A(u) = \begin{cases} a(u), & \textit{if } u \in \bar{P}, \\ +\infty, & \textit{otherwise.} \end{cases}$$

Proof. Case 1. Take $u \in \bar{P}$. Since $a(\cdot)$ is convex and homogeneous of degree 1 $h_A(u) = a(u)$.

Case 2. Take $u \in \mathbb{S}^{d_y-1} \setminus \bar{P}$. First, we establish that there always exists $u^* \in \text{PolCon}(\bar{P})$ such that $u'u^* > 0$. To prove this suppose to the contrary that for every $u^* \in \text{PolCon}(\bar{P})$, $u'u^* \leq 0$, it follows that $u \in \text{PolCon}(\text{PolCon}(\bar{P}))$. The latter is not possible, since $\text{PolCon}(\text{PolCon}(\bar{P}))$ is the smallest closed convex cone containing $\bar{P}$ (Rockafellar, 1970, Theorem 14.1), and $u \notin \bar{P}$ by assumption.

For some u^* that satisfies $u'u^* > 0$, consider $y^m = y^0 + mu^*$, $m = 1, 2, \ldots$, where y^0 is an arbitrary point from A. Since $u^* \in \text{PolCon}(\bar{P})$, by construction $u^{*\prime}p \leq 0$ for all $p \in \bar{P}$. Using this fact, note that $y^m \in A$ for all $m = 1, 2, \ldots$ since

$$p'y^m = p'y^0 + mu^{*\prime}p \leq a(p) + 0$$

for all $p \in \bar{P}$. Finally,

$$h_A(u) \geq u'y^m = u'y^0 + mu'u^*$$

diverges to $+\infty$, since $u'u^* > 0$. ■

We now provide a key lemma. This result generalizes a classical result that holds for $\bar{P} = \mathbb{S}^{d_y-1}$. To our knowledge this result is new, and it may be of independent interest.

Lemma A.2. *Let* $d_y \geq 2$ *and let the functions* $a, b : \mathbb{R}^{d_y}_{++} \to \mathbb{R}$ *be convex and homogeneous of degree* 1*. Define*

$$A = \left\{ y \in \mathbb{R}^{d_y} \ : \ p'y \leq a(p), \, \forall p \in \bar{P} \right\} ,$$
$$B = \left\{ y \in \mathbb{R}^{d_y} \ : \ p'y \leq b(p), \, \forall p \in \bar{P} \right\} ,$$

where $\bar{P} \subseteq \mathbb{R}^{d_y}_{++}$ *is convex and compact. Then*

$$d_H(A, B) = \sup_{p \in \bar{P}} \| a(p/\|p\|) - b(p/\|p\|) \| \, .$$

Proof. For closed convex sets $C, D \subseteq \mathbb{R}^{d_y}$ the following is true: $C \subseteq D$ if and only if

$h_C(u) \le h_D(u)$ for all $u \in \mathbb{S}^{d_y-1}$. Hence,

$$\begin{aligned}&\{\rho \in \mathbb{R}_+ \,:\, A \subseteq B + \rho\mathbb{B}^{d_y-1}, B \subseteq A + \rho\mathbb{B}^{d_y-1}\} \iff \\ &\{\rho \in \mathbb{R}_+ \,:\, h_A(u) \le h_{B+\rho\mathbb{B}^{d_y-1}}(u), h_B(u) \le h_{A+\rho\mathbb{B}^{d_y-1}}(u), \forall u \in \mathbb{S}^{d_y-1}\}\,.\end{aligned}$$

Because $\bar{P}$ is a subset of $\mathbb{R}^{d_y}_{++}$, its polar cone $\mathrm{PolCon}(\overline{P})$ is nonempty; in particular the polar cone contains the negative unit vector $(-1, \ldots, -1)'$. The set $\bar{P}$ satisfies the conditions of Lemma A.1, and so we obtain that $h_A(u) = h_{B+\rho\mathbb{B}^{d_y-1}}(u) = h_B(u) = h_{A+\rho\mathbb{B}^{d_y-1}}(u) = \infty$ for all $u \in \mathbb{S}^{d_y-1} \setminus \{p/\|p\|\,,\, p \in \bar{P}\}$. Hence,

$$\begin{aligned}&\{\rho \in \mathbb{R}_+ \,:\, A \subseteq B + \rho\mathbb{B}^{d_y-1}, B \subseteq A + \rho\mathbb{B}^{d_y-1}\} \\ &\qquad = \{\rho \in \mathbb{R}_+ \,:\, \ h_A(u) \le h_{B+\rho\mathbb{B}^{d_y-1}}(u), \\ &\qquad\qquad\quad h_B(u) \le h_{A+\rho\mathbb{B}^{d_y-1}}(u), \forall u \in \{p/\|p\| \,: p \in \bar{P}\}\} \\ &\qquad = \{\rho \in \mathbb{R}_+ \,:\, \ h_A(u) \le h_B(u) + h_{\rho\mathbb{B}^{d_y-1}}(u), \\ &\qquad\qquad\quad h_B(u) \le h_A(u) + h_{\rho\mathbb{B}^{d_y-1}}(u), \forall u \in \{p/\|p\| \,:\, p \in \bar{P}\}\} \\ &\qquad = \{\rho \in \mathbb{R}_+ \,:\, h_A(u) \le h_B(u) + \rho, h_B(u) \le h_A(u) + \rho, \forall u \in \{p/\|p\| \,: p \in \bar{P}\}\} \\ &\qquad = \{\rho \in \mathbb{R}_+ \,:\, \sup_{u \in \{p/\|p\| \,:\, p \in \bar{P}\}} \|h_A(u) - h_B(u)\| \le \rho\}\,.\end{aligned}$$

Now note that $a(p)$ and $b(p)$ are values of the support functions of A and B evaluated at $p \in \bar{P}$, respectively, since $a(\cdot)$ and $b(\cdot)$ are homogeneous of degree 1 and convex. Thus,

$$d_H(A, B) = \sup_{p \in \bar{P}} \|a(p/\|p\|) - b(p/\|p\|)\|\,.$$

■

To prove Theorem 4 note that since $\pi(\cdot, e)$ and $\hat{\pi}(\cdot, e)$ are homogeneous of degree 1, we have

$$\begin{aligned}\pi(p, e)/\|p\| &= \pi\left(p/\|p\|, e\right)\,, \\ \hat{\pi}(p, e)/\|p\| &= \hat{\pi}\left(p/\|p\|, e\right)\,,\end{aligned}$$

for all $p \in \bar{P}$ and $e \in E$. Thus, Theorem 4 is obtained as corollary.

We now turn to the proof of Proposition 2. We first present two lemmas, which are modifications of Lemmas 6 and 7 in Brunel (2016).

Lemma A.3. *Assume that $\bar{P} \subseteq \mathbb{S}^{d_y-1} \bigcap P$ is compact and $\cup_{\lambda>0}\{\lambda p : p \in \bar{P}\}$ is convex. Let $a : \bar{P} \to \mathbb{R}$ be a continuous function. Let $A = \{y \in \mathbb{R}^{d_y} \,:\, p'y \le a(p),\, p \in \bar{P}\}$ be* *nonempty. It follows that for all $p^* \in \bar{P}$ there exists $y^* \in A$ such that $h_A(p^*) = p^{*\prime} y^*$. Moreover, there exists $P^* \subseteq \bar{P}$ such that*

(i) The cardinality of P^* is less than or equal to d_y;

(ii) $p' y^* = a(p)$ for all $p \in P^*$;

(iii) $p^* = \sum_{p \in P^*} \lambda_p p$ for some nonnegative numbers λ_p.

Proof. Fix some $p^* \in \bar{P}$. Note that $h_A(p^*) \leq a(p^*) < \infty$. Since A is closed, by the supporting hyperplane theorem $h_A(p^*) = p^{*\prime} y^*$ for some $y^* \in A$.

The rest of the lemma follows from Theorem 2(b) in López and Still (2007) if we show that $P' = \{p \in \bar{P} \, : \, p' y^* = a(p)\}$ is nonempty. By way of contradiction assume that P' is empty. Hence, $p' y^* < a(p)$ for all $p \in \bar{P}$. Since the function $a(\cdot) - \cdot' y^*$ is strictly positive on a compact $\bar{P}$, there exists $\nu > 0$ that bounds $a(\cdot) - \cdot' y^*$ from below. Hence, for every $p \in \bar{P}$,

$$p'(y^* + \nu p^*) = p' y^* + \nu p' p^* \leq a(p) - \nu + \nu p' p^* \leq a(p) \, .$$

Thus, $(y^* + \nu p^*) \in A$. But the later is not possible since $p^*(y^* + \nu p^*) = a(p^*) + \nu > a(p^*)$ implies that y^* is not a maximizer. Thus, P' is nonempty. ■

Lemma A.4. *Assume that $\bar{P} \subseteq \mathbb{S}^{d_y - 1} \bigcap P$ is compact and $\cup_{\lambda > 0} \{\lambda p \, : \, p \in \bar{P}\}$ is convex. Let $a : \bar{P} \to \mathbb{R}$ be continuous convex homogeneous of degree 1 function and $\{b_n : \bar{P} \to \mathbb{R}\}$ be a sequence of continuous homogeneous of degree 1 functions such that*

$$A = \left\{ y \in \mathbb{R}^{d_y} \, : \, p' y \leq a(p), \, \forall p \in \bar{P} \right\} ,$$
$$B_n = \left\{ y \in \mathbb{R}^{d_y} \, : \, p' y \leq b_n(p), \, \forall p \in \bar{P} \right\} ,$$

are nonempty for all $n \in \mathbb{N}$. Assume that $\eta_n = \sup_{p \in \bar{P}} \|a(p) - b_n(p)\| = o(1)$ and $0 < r = \inf_{p \in \bar{P}} a(p) < R = \sup_{p \in \bar{P}} a(p) < \infty$. Then there exists $N > 0$ such that

$$\sup_{p \in \bar{P}} \|a(p) - h_{B_n}(p)\| \leq \tilde{d}_n \frac{R}{r} \frac{1 + \eta_n / R}{1 - \eta_n / r}$$

for all $n > N$.

Proof. Fix some $p^* \in \bar{P}$ and some n such that $\eta_n < r$. By Lemma A.3 there exists a finite set P_n^*, a collection of nonnegative numbers $\{\lambda_{p,n}\}_{p \in P_n^*}$ and $y_n^* \in B_n$ such that

$h_{B_n} = p^{*\prime} y_n^*$, $p^* = \sum_{p\in P_n^*} \lambda_{p,n} p$, and $p' y_n^* = b_n(p)$ for all $p \in P_n^*$. Note that for all $p \in p_n^*$ we have that $b_n(p) = h_{B_n}(p)$. Then

$$
\begin{aligned}
a(p^*) = h_A(p^*) = h_A\left(\sum_{p\in P_n^*} \lambda_{p,n} p\right) &\leq \sum_{p\in P_n^*} \lambda_{p,n} h_A(p) = \sum_{p\in P_n^*} \lambda_{p,n} a(p) \leq \sum_{p\in P_n^*} \lambda_{p,n}(b_n(p)+\eta_n) \\
&= \sum_{p\in P_n^*} \lambda_{p,n} p' y_n^* + \eta_n \sum_{p\in P_n^*} \lambda_{p,n} = p^{*\prime} y_n^* + \eta_n \sum_{p\in P_n^*} \lambda_{p,n} = h_{B_n}(p^*) + \eta_n \sum_{p\in P_n^*} \lambda_{p,n}\,.
\end{aligned}
\tag{9}
$$

Moreover,

$$
h_{B_n}(p^*) \leq b_n(p^*) \leq a(p^*) + \eta_n\,. \tag{10}
$$

Hence, $\|a(p^*) - h_{B_n}(p^*)\| \leq \eta_n \max\{1, \sum_{p\in P_n^*} \lambda_{p,n}\}$.

Next note that the inequality in (10) implies that

$$
\sum_{p\in P_n^*} \lambda_{p,n} p' y_n^* = p^{*\prime} y_n^* = h_{B_n}(p^*) \leq a(p^*) + \eta \leq R + \eta_n\,.
$$

In addition,

$$
\sum_{p\in P_n^*} \lambda_{p,n} p' y_n^* = \sum_{p\in P_n^*} \lambda_{p,n} b_n(p) \geq \sum_{p\in P_n^*} \lambda_{p,n}(a(p)-\eta_n) \geq \sum_{p\in P_n^*} \lambda_{p,n}(r-\eta_n)\,.
$$

Hence,

$$
\sum_{p\in P_n^*} \lambda_{p,n} \leq \frac{R+\eta_n}{r-\eta_n}\,.
$$

As a result,

$$
\|a(p^*) - h_{B_n}(p^*)\| \leq \eta_n \max\left\{1, \sum_{p\in P_n^*} \lambda_{p,n}\right\} = \eta_n \max\left\{1, \frac{R+\eta_n}{r-\eta_n}\right\} = \eta_n \frac{R}{r}\frac{1+\eta_n/R}{1-\eta_n/r}\,.
$$

■

To prove Theorem 4 note that since $\pi(\cdot, e)$ and $\hat{\pi}(\cdot, e)$ are homogeneous of degree 1, we have

$$
\begin{aligned}
\pi(p,e)/\|p\| &= \pi\left(p/\|p\|, e\right)\,, \\
\hat{\pi}(p,e)/\|p\| &= \hat{\pi}\left(p/\|p\|, e\right)\,.
\end{aligned}
$$

To prove Proposition 2, note that by Lemma A.2, with probability 1,

$$d_H(Y_{\bar{P}}(e), \hat{Y}_{\bar{P}}(e)) = \sup_{p\in\bar{P}} \left\| \pi(p/\left\|p\right\|, e) - h_{\hat{Y}_{\bar{P}}(e)}(p/\left\|p\right\|) \right\| .$$

The conclusion then follows by applying Lemma A.4 to the right hand side of the equality above.

B. Additional Details on Estimation

This section presents an estimator that is used in the illustrative empirical application in Appendix C. The estimator builds on the constructive identification result of Theorem 1 and applies with continuous measurement error and discrete heterogeneity in productivity. It proceeds in two steps. First, we find a "minimal-width" region of profits that is used to estimate the distribution of measurement error. This uses the well-separatedness structure of Theorem 1. Second, we use the estimate of the distribution of measurement error to estimate the distribution of structural profit.

B.1. Estimation of Restricted Profit Function

To simplify the exposition, in this section we assume the researcher observes data on (unrestricted) profits and prices from M markets $\{\boldsymbol{\pi}_{i,m}, \mathbf{p}_m\}_{i=1,\ldots,N;m=1,\ldots,M}$. Here, $\boldsymbol{\pi}_{i,m}$ is the observed profit of firm i in market m, which may be mismeasured. The index i can be market specific, so in particular firm 1 in market 1 may differ from firm 1 in market 2. For each market m, all firms face the same price vector $\mathbf{p}_m$. There are N firms in every market.[32] The general case with restricted profits can be handled similarly. We assume that $\mathbf{p}_m = \mathbf{p}_{m'}$ with probability 1 if and only if $m = m'$ (i.e., markets have different prices). We require the number of firms per market N to grow to infinity. The number of markets M can be fixed, grow to a finite constant, or diverge to infinity as long as it grows slower than N.

2.1.a. *Estimation of the Measurement Error Distribution.–* With this setup, we can estimate the distribution of measurement error. We do so by first finding a partition of

[32]We assume the same number of firms in every market only to simplify the exposition.

profits in which some region has "minimal width." To formalize this we first describe how we partition. For a finite set of distinct reals $\mathcal{T} = \{t_\ell\}_{\ell=1}^L$ and $\kappa > 0$, let $t^{(\ell)}$ be the ℓ-th smallest element of $\mathcal{T}$. Next, let $\{T_k^\kappa\}_{k=1}^{K_\kappa}$ be a smallest (in terms of cardinality) partition of $\mathcal{T}$ such that $\max T_k^\kappa < \min T_{k+1}^\kappa$ for all $k = 1, \ldots, K_\kappa - 1$, and $\left|t^{(\ell)} - t^{(j)}\right| \leq |\ell - j| \, \kappa$ for any T_k^κ and any $t^{(\ell)}, t^{(j)} \in T_k^\kappa$. Such partition always exists but may not be unique. For our purposes, any such partition works.[33] Let $d(T_k^\kappa) = (\max T_k^\kappa - \min T_k^\kappa)$ be the diameter of the set T_k^κ. Given the partition, let k^* be the smallest integer such that $d(T_{k^*}^\kappa) \leq d(T_k^\kappa)$ for each $k = 1, \ldots, K_\kappa$. That is, $T_{k^*}^\kappa$ is the *first-shortest* element of the partition. Finally, let

$$\mathcal{C}(\mathcal{T}, \kappa) = \left\{ t - \frac{1}{|T_{k^*}^\kappa|} \sum_{t' \in T_{k^*}^\kappa} t' \right\}_{t \in T_{k^*}^\kappa}.$$

That is, the operator $\mathcal{C}$ takes the set $\mathcal{T}$ and threshold $\kappa > 0$, computes the set $T_{k^*}^\kappa$, and then re-centers this set such that the sample average of elements of it is zero.

Given a sequence of positive reals κ_N that slowly converges to 0, let m_N^* be a market that has the smallest $d\left(\mathcal{C}\left(\{\boldsymbol{\pi}_{i,m}\}_{i=1}^N, \kappa_N\right)\right)$. Then, under the assumptions of Theorem 1, the elements of $\mathcal{C}\left(\{\boldsymbol{\pi}_{i,m_N^*}\}_{i=1}^N, \kappa_N\right)$ mimic the unobserved realizations of the measurement error. Thus, we can apply any consistent estimator (e.g., kernels or sieves) to $\mathcal{C}\left(\{\boldsymbol{\pi}_{i,m_N^*}\}_{i=1}^N, \kappa_N\right)$ to obtain a consistent estimator of the p.d.f. of the measurement error.

Proposition B.1. *Take κ_N such that $\kappa_N = o(1)$ and $\log(N)/(N\kappa_N) = o(1)$. Assume the assumptions of Theorem 1 are satisfied. Assume $\boldsymbol{\eta}$ admits a continuous p.d.f. $f_{\boldsymbol{\eta}}$. Suppose there is an estimator $\hat{f}_{\boldsymbol{\eta}}(\cdot, \{\boldsymbol{\eta}_i\})$ that is consistent for $f_{\boldsymbol{\eta}}$, based on an i.i.d. sample from $f_{\boldsymbol{\eta}}$, denoted $\{\boldsymbol{\eta}_i\}$. Let m_N^* be such that $d\left(\mathcal{C}(\{\boldsymbol{\pi}_{i,m_N^*}\}, \kappa_N)\right) \leq d\left(\mathcal{C}(\{\boldsymbol{\pi}_{i,m}\}, \kappa)\right)$ for all m with probability 1. It follows that $\hat{f}_{\boldsymbol{\eta}}\left(\cdot, \mathcal{C}(\{\boldsymbol{\pi}_{i,m_N^*}\}, \kappa_N)\right)$ is a consistent estimator of $f_{\boldsymbol{\eta}}$.*

Proof. First, note that for any $\kappa > 0$ and two random variables $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ that are independently and identically distributed according to $f_{\boldsymbol{\eta}}$,

$$p(\kappa) = \mathbb{P}\left(|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2| \leq \kappa\right) = \int_{-K_1}^{-K_1+\kappa} [F_{\boldsymbol{\eta}}(x+\kappa) - F_{\boldsymbol{\eta}}(-K_1)] f_{\boldsymbol{\eta}}(x) dx +$$
$$\int_{K_2-\kappa}^{K_2} [F_{\boldsymbol{\eta}}(K_2) - F_{\boldsymbol{\eta}}(x-\kappa)] f_{\boldsymbol{\eta}}(x) dx + \int_{-K_1+\kappa}^{K_2-\kappa} [F_{\boldsymbol{\eta}}(x+\kappa) - F_{\boldsymbol{\eta}}(x-\kappa)] f_{\boldsymbol{\eta}}(x) dx,$$

[33] This partition is related to so-called density-based clustering. See Kriegel, Kröger, Sander and Zimek (2011) for a review.

where $F_{\boldsymbol{\eta}}$ is the c.d.f. of $\boldsymbol{\eta}$ supported on $[-K_1, K_2]$, where $K_1, K_2 > 0$. The first term in the above equation can be bounded by

$$\begin{aligned}\int_{-K_1}^{-K_1+\kappa}[F_{\boldsymbol{\eta}}(x+\kappa)-F_{\boldsymbol{\eta}}(-K_1)]f_{\boldsymbol{\eta}}(x)dx &\le \max_x f_{\boldsymbol{\eta}}(x)\int_{-K_1}^{-K_1+\kappa}[F_{\boldsymbol{\eta}}(-K_1+2\kappa)-F_{\boldsymbol{\eta}}(-K_1)]dx\\ &\le \max_x f_{\boldsymbol{\eta}}(x)\frac{F_{\boldsymbol{\eta}}(-K_1+2\kappa)-F_{\boldsymbol{\eta}}(-K_1)}{2\kappa}2\kappa^2.\end{aligned}$$

Similarly, the second term is bounded above by

$$\max_x f_{\boldsymbol{\eta}}(x)\frac{F_{\boldsymbol{\eta}}(K_2)-F_{\boldsymbol{\eta}}(K_2-2\kappa)}{2\kappa}2\kappa^2.$$

As a result, since $\max_x f_{\boldsymbol{\eta}}(x) < \infty$ ($f_{\boldsymbol{\eta}}$ is continuous on a compact support) and $F_{\boldsymbol{\eta}}$ has a bounded and continuous derivative on a compact set, as $\kappa \to 0$,

$$p(\kappa) = 2\kappa\int_{-K_1+\kappa}^{K_2-\kappa}[f_{\boldsymbol{\eta}}(x)+O(\kappa)]f_{\boldsymbol{\eta}}(x)dx + O(\kappa^2)$$

and

$$\lim_{\kappa\to 0}\frac{p(\kappa)}{\kappa} = C = 2\int_{-K_1}^{K_2} f_{\boldsymbol{\eta}}^2(x)dx > 0.$$

Second, note that given an i.i.d. sample $\{\boldsymbol{\eta}_i\}_{i=1}^n$ from $f_{\boldsymbol{\eta}}$

$$\begin{aligned}&\mathbb{P}\left(\max_i\min_{j\neq i}\left|\boldsymbol{\eta}_i-\boldsymbol{\eta}_j\right|\le\kappa\right) = \mathbb{P}\left(\bigcap_{i=1}^n\left\{\min_{j\neq i}\left|\boldsymbol{\eta}_i-\boldsymbol{\eta}_j\right|\le\kappa\right\}\right)\\ &\ge\sum_{i=1}^n\mathbb{P}\left(\min_{j\neq i}\left|\boldsymbol{\eta}_i-\boldsymbol{\eta}_j\right|\le\kappa\right)-(n-1) = 1-\sum_{i=1}^n\mathbb{P}\left(\min_{j\neq i}\left|\boldsymbol{\eta}_i-\boldsymbol{\eta}_j\right|>\kappa\right) =\\ &= 1-n\mathbb{P}\left(\left|\boldsymbol{\eta}_1-\boldsymbol{\eta}_2\right|>\kappa\right)^{n-1} = 1-n(1-p(\kappa))^{n-1} = 1-n(1-C\kappa+o(\kappa))^{n-1}.\end{aligned}$$

Hence,

$$\lim_{N\to\infty}\mathbb{P}\left(\max_i\min_{j\neq i}\left|\boldsymbol{\eta}_i-\boldsymbol{\eta}_j\right|\le\kappa_N\right)\ge 1-\lim_{N\to\infty}N\exp(-C(N-1)\kappa_N) = 1,$$

where the last equality follows from the fact that κ_N converges to 0 slower that $\log(N)/N$.

This bound on the measurement error distribution implies that the largest distance between neighboring observations that are coming from the same productivity level becomes less than κ_N with probability approaching 1 as N increases. Hence, since κ_N converges to zero and we pick the shortest element of the partition, $\mathcal{C}(\{\boldsymbol{\pi}_{i,m_N^*}\}, \kappa_N)$ will contain i.i.d observations that correspond to the same productivity level with probability approaching 1. Thus, any consistent estimator that is based on an i.i.d. sample will be consistent. ■

2.1.b. *Estimation of the Structural Profit Function.–* Given a consistent estimator of $f_{\boldsymbol{\eta}}$, we can estimate the distribution of $\pi(\mathbf{p}_m, \mathbf{e})$ conditional on a given market m. Note that since $\mathbf{e}$ has finite support, the observed distribution of profits in market m is a finite mixture

$$f_{\boldsymbol{\pi}|\mathbf{p}}(\cdot|p) = \sum_{e=1}^{d_e} f_{\boldsymbol{\eta}}(\cdot - \pi(p, e))\mathbb{P}\left(\mathbf{e} = e|\mathbf{p} = p\right).$$

Hence, for any consistent estimator of $f_{\boldsymbol{\eta}}$, $\hat{f}_{\boldsymbol{\eta}}$, and a known number of types d_e, we can define a parametric loglikelihood

$$\hat{L}(\theta) = \sum_{i=1}^{N} \log\left(\sum_{e=1}^{d_e} \hat{f}_{\boldsymbol{\eta}}(\boldsymbol{\pi}_{i,m} - \pi_e)\rho_e\right),$$

where $\theta = \left((\pi_e)'_{e=1,\dots,d_e}, (\rho_e)'_{e=1,\dots,d_e}\right)'$ is a vector of parameters of interest, and then find a maximum-likelihood estimator (MLE) as a solution to

$$\begin{aligned} &\max_{\theta} \hat{L}(\theta) \\ \text{s.t.}\ &\sum_{e=1}^{d_e} \rho_e = 1, \\ &\rho_e \geq 0, \quad e = 1, \dots, d_e, \\ &\pi_e \leq \pi_{e+1}, \quad e = 1, \dots, d_e - 1. \end{aligned}$$

B.2. Monte Carlo Simulations

In this section, we evaluate the performance of our estimator of the profit function in Monte Carlo simulations. In particular, we consider the setting of a single market. Thus, prices p are fixed and we drop them from the notation. We work with the following data generating process: there are four level of productivity that correspond to four levels of profits $\pi_e \in \{2, 3, 3.5, 6\}$. The probability of each profit level is $\rho_e \in \{0.2, 0.25, 0.25, 0.3\}$. The profits are contaminated with measurement error $\boldsymbol{\eta}$

that is independent of everything and is distributed on $[-1, 1]$ with p.d.f.

$$f_{\boldsymbol{\eta}}(\eta) = \mathbb{1}\,(\,|\eta| \leq 1\,)\,0.75(1-\eta^2).$$

As a result, observed profits (including measurement error) are distributed with p.d.f.

$$f_{\boldsymbol{\pi}}(\tilde{\pi}) = \sum_{e=1}^{4} f_{\boldsymbol{\eta}}(\tilde{\pi} - \pi_e)\rho_e.$$

Note that this data generating process satisfies Assumption 4 with $K = 2$ and $e^* = 4$ (i.e., $5 = \pi_4 - 1 > \pi_3 + 1 = 4.5$). To estimate the measurement error distribution, we used a kernel density estimator with the least-squares cross-validated bandwidth.

We estimated π_e, $e = 1, 2, 3, 4$, for four different sample sizes $N \in \{500, 1000, 1500\}$ and three different kernels (Epanechnikov, Biweight, and Triweight) using the procedure described in Section B.1. In our estimation, we assume that the number of productivity levels d_e is known. We replicate the experiment for every sample size and kernel 1000 times. The bias and the root-mean-square error (RMSE) of the estimator of π_e, $e = 1, 2, 3, 4$, are presented in Tables 1-3. The RMSE decreases with the sample size. The bias decreases for all productivity levels except π_2 for Epanechnikov kernel. All three kernels perform similarly, while the Epanechnikov kernel performs slightly better.

Table 1 – Bias and RMSE, Epanechnikov Kernel ($\times 10^{-2}$)

	Bias			RMSE		
	$N = 500$	$N = 1000$	$N = 1500$	$N = 500$	$N = 1000$	$N = 1500$
π_1	6.67	5.47	4.73	10.67	8.04	6.84
π_2	2.13	0.24	-1.37	13.67	11.41	9.6
π_3	-10.34	-8.51	-7.52	13.87	11.46	9.98
π_4	0.06	0.08	0.04	3.31	2.38	1.94

Notes: 1000 replications.

C. Illustrative Empirical Application

In this section, we analyze the production of houses using data from Epple et al. (2010) in line with the model described in Section 3.1.

Table 2 – Bias and RMSE, Biweight Kernel ($\times 10^{-2}$)

	Bias			RMSE		
	$N = 500$	$N = 1000$	$N = 1500$	$N = 500$	$N = 1000$	$N = 1500$
π_1	6.69	5.57	4.65	11.05	8.33	6.91
π_2	7.61	3.45	-0.07	22.49	18.52	15.11
π_3	-11.27	-8.84	-7.72	15.3	12	10.3
π_4	0.06	0.07	0.04	3.28	2.37	1.92

Notes: 1000 replications.

Table 3 – Bias and RMSE, Triweight Kernel ($\times 10^{-2}$)

	Bias			RMSE		
	$N = 500$	$N = 1000$	$N = 1500$	$N = 500$	$N = 1000$	$N = 1500$
π_1	6.78	5.59	4.68	11.08	8.32	6.89
π_2	7.8	3.36	-0.19	22.46	18.13	14.48
π_3	-11.38	-8.95	-7.68	15.39	12.07	10.26
π_4	0.05	0.07	0.04	3.27	2.36	1.92

Notes: 1000 replications.

Data. The data contains information on new housing construction in Allegheny County in Pennsylvania. For every dwelling i we have information about total revenue from selling the house $\mathbf{v}_i$, the price of land $\mathbf{p}_{l,i}$, materials per-acre $\mathbf{m}_i$, and the geographic location of the house that we use to identify the zip-code for each house. From the original sample constructed in Epple et al. (2010), we exclude houses with the value per unit of land and the price of land above 55 and 7, respectively. There are 5,641 houses in our sample. Table 4 provides summary statistics of our sample. Figure 6 displays a distribution of the price of land and the value per unit of land. For more details on the original data, see Epple et al. (2010).

Table 4 – Summary Statistics

Variable	Mean	Median	Std	Min	Max
Value per unit of land	14.43	13.24	8.6	0.15	54.03
Price of land	2.26	2.11	1.28	0.05.4	6.99

Notes: These summary statistics illustrate the heterogeneity of the value per unit of land and price of land.

Market definition. We highlight that our method takes markets as known, but in practice we have to define them. We assume that within Allegheny County, local

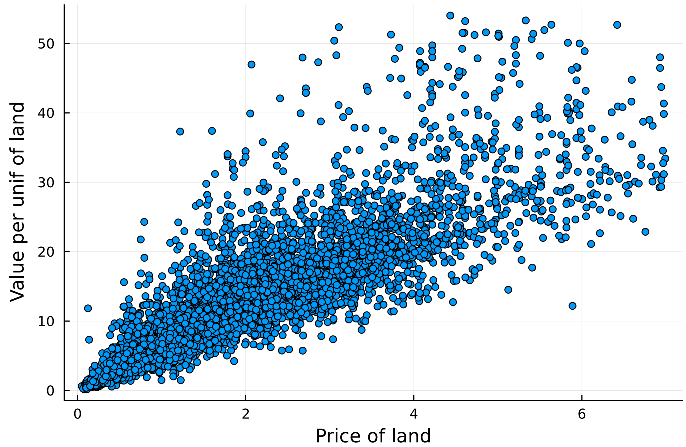

Figure 6 – Distribution of Price of Land and Value per Unit of Land in Sample

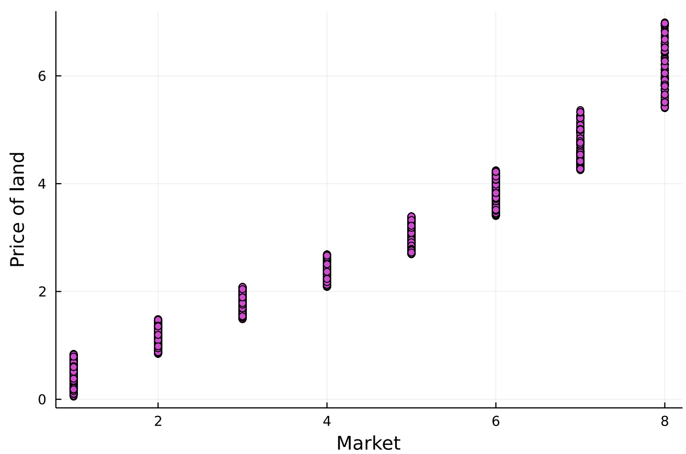

Figure 7 – Distribution of Price of Land across Markets

markets are determined by the location of the house (coordinates) and the price of land. To construct the markets, we use K-means clustering using location and the price of land. We select the number of cluster using the heuristic *elbow* method. We end up with 8 markets. After clustering the observations, we average the price of land within the market to obtain the market level price of land $\mathbf{p}_m$. The distribution of the price of land and the value per unit of land across the local markets are depicted on Figures 7 and 8. There is not much variation in the price of land within most of the markets, but substantial variation across markets.[34] This is evidence of the validity of our assumption that firms within the same market face the same prices. At the

[34]The largest across markets standard deviation is about 0.46, which corresponds to market 8.

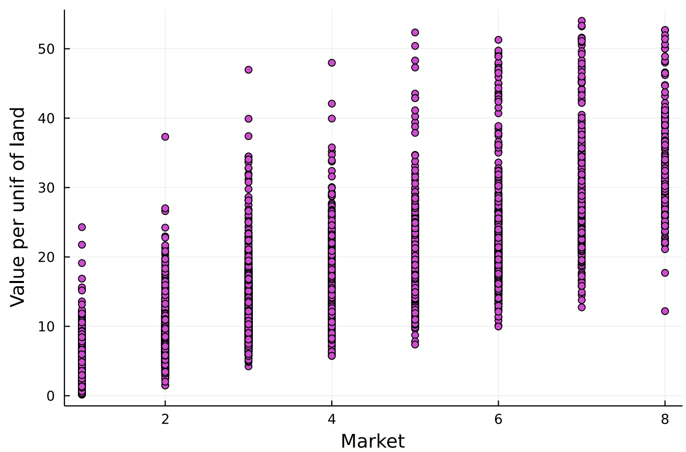

Figure 8 – Distribution of Value per Unit of Land across Markets

same time, there is a lot of variation in valuation per unit of land across and within markets. Moreover, the valuations are clearly bounded from below and some markets display clear separated sets of points that line up with our assumptions of discrete heterogeneity and bounded support of measurement error.

Estimation of the measurement error distribution. To estimate the distribution of the measurement error, we note that markets 6 and 7 exhibit two sets of observations with high value per unit of land, that are clearly separated from the rest of observations.[35] Moreover, these sets cover intervals of very similar length. This is consistent with the assumption that at least in one market at least one type is clearly separated from other types. We use the Epanechnikov kernel with the least-squares cross-validated bandwidth to estimate the measurement error p.d.f using these sets of observations.

Estimation of values per unit of land for different productivity levels. For every market, given the estimated density of the measurement error, we apply the procedure for estimation of finite mixtures described in Kim, Carbonetto, Stephens and Anitescu (2020) to get a good starting point to obtain the MLE of values per unit of land for firms with different productivity $\{\hat{v}_m(e)\}_{m=1,e=1}^{M,d_e}$. We assume for simplicity that the number of types of firms is the same across markets and is equal to $d_e = 4$, which is the minimal number of mixtures that is able to cover the observed support of mismeasured values per unit of land.[36]

Proxy function. To estimate the proxy function that maps average value per unit of land to the price of output, we follow Epple et al. (2010). In particular, we use a 3rd

[35]Formally, we use a density-based clustering technique in this step.

[36]The results for $d_e = 5$ and $d_e = 6$ are qualitatively the same and available upon request.

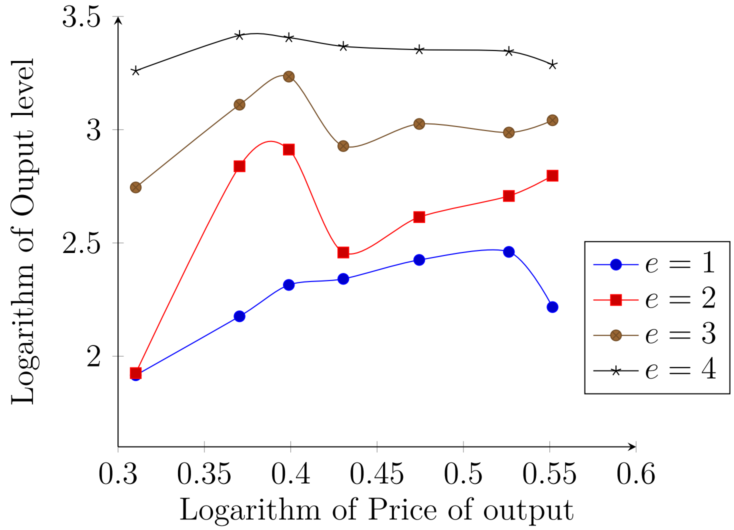

Figure 9 – Supply curves for different productivity levels for 7 markets with the highest price of output.

order degree polynomial to estimate $\mathbb{E}\left[\mathbf{p}|\bar{\mathbf{v}}\right]$ and then solved the ordinary differential equation (4). As a result, we can estimate the output price at every market $\{\hat{p}_{o,m}\}_{m=1}^{M}$.

Supply. To estimate the output level of firms with productivity e, we use $\{\hat{v}_m(e)/\hat{p}_{o,m}\}_{m=1,e=1}^{M,E}$. The resulting logarithm of supply as a function of the logarithm of the output price is depicted on Figure 9. The supply curves are close to be monotonically increasing. We attribute nonmonotonicity to estimation error. Next we enforce monotonicity by finding the output levels that preserve monotonicity in the output price and minimize the Euclidean distance to the estimated output level. The resulting logarithm of supply, for firms with different productivity, as a function of the logarithm of the output price is depicted on Figure 10.

Discussion. Our results indicate that there is substantial heterogeneity in the supply of housing. Recall that the results in Epple et al. (2010) focus on a representative firm. In contrast, our results suggest that we cannot ignore heterogeneity. For instance, factor-reallocation is total productivity enhancing when resources shift from the less productive to the more productive firms (Melitz and Redding, 2014).

In the production of housing, heterogeneity can be interpreted as *curb appeal* (Epple et al., 2010). The most productive firms ($e = 4$) produce the houses with the highest curb appeal. In that sense, it is important to disentangle this heterogeneity when estimating the housing supply elasticity. Housing supply elasticity has been seen as a key parameter (Glaeser, Gyourko and Saks, 2005) to understand the relationship between urban growth and new residential construction. An inelastic supply means

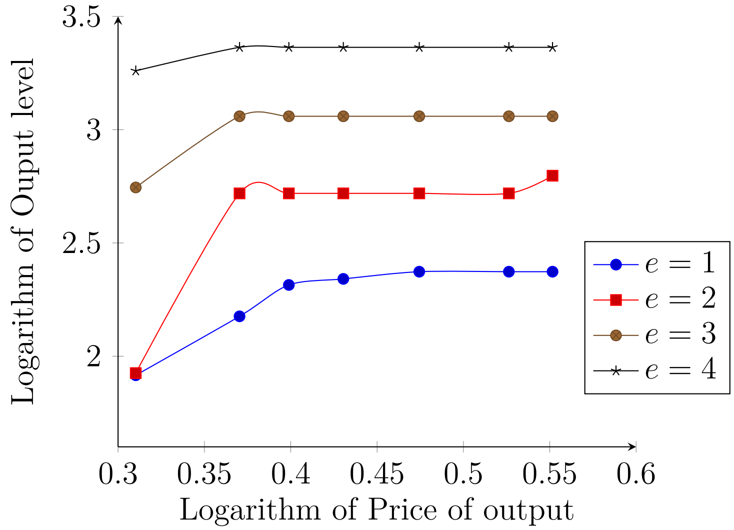

Figure 10 – Monotone supply curves for different productivity levels for 7 markets with the highest price of output.

that a positive regional shock will lead to higher paid workers and more expensive houses. If the housing supply is elastic, then we can expect smaller price changes and expansion of the size of the city. Using Figure 10, we computed the average elasticity for all types of firms (Table 5).

Table 5 – Average Elasticities of Firms with Different Productivity

	$e = 1$	$e = 2$	$e = 3$	$e = 4$
Average Elasticity	1.79	2.71	0.87	0.29

We observe from Table 5 that the most productive firm type ($e = 4$) has a very inelastic supply. This means that houses with the highest curb appeal will mainly see an increase of prices (without a large expansion) due to a positive regional shock. In contrast, we observe that the least productive firm types ($e = 1$ and $e = 2$) have elastic supplies. This means that there will be an expansion in the construction (with an smaller increase of prices) of houses with the lowest curb appeal as a result of the same positive regional shock.

D. Supplemental Results

D.1. Continuous Heterogeneity

In this section we consider the possibility of continuous heterogeneity. Unfortunately, in contrast to our main result in Theorem 1, when facing continuous heterogeneity we will need more observables (instruments) and more assumptions. Moreover, we will need to assume that there is no measurement error η.

For simplicity we state the results for the profit function. The analysis of the general restricted profit function is similar. Assume that the analyst observes $(\boldsymbol{\pi}, \mathbf{p}', \mathbf{w}')'$, where the instrumental variable $\mathbf{w}$ is supported on W and $\boldsymbol{\pi} = \pi(\mathbf{p}, \mathbf{e})$ are perfectly measured profits. We normalize $\mathbf{e}$ to be uniformly distributed.

Assumption D.1. *The distribution of* $\mathbf{e}$ *is uniform over* $[0, 1]$.

The following assumption is an independence condition that requires the instrumental variable to be independent of the unobservable heterogeneity $\mathbf{e}$.

Assumption D.2. $F_{\mathbf{e}|\mathbf{w}}(\cdot|w) = F_{\mathbf{e}}(\cdot)$ *for all* $w \in W$.

Assumption D.2 together with the requirement that the profit function $\pi(p, \cdot)$ is monotone (Assumption 1) imply the following integral equation familiar from the literature on nonparametric quantile instrumental variable models.

Lemma D.1. *If Assumptions 1, D.1 and D.2 are satisfied, then the following holds:*

$$\mathbb{P}\left(\boldsymbol{\pi} \le \pi(\mathbf{p}, e) | \mathbf{w} = w\right) = e \tag{11}$$

for all $e \in E$ *and* $w \in W$.

Proof. Fix some $w \in W$ and $e \in E$. First, note that by the law of iterated expectations

$$\mathbb{P}\left(\boldsymbol{\pi} - \pi(\mathbf{p}, e) \le 0 | \mathbf{w} = w\right) = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}\left(\, \pi(p, \mathbf{e}) - \pi(p, e) \le 0 \,\right) | \mathbf{p} = p, \mathbf{w} = w\right] | \mathbf{w} = w\right] .$$

By strict monotonicity of $\pi(p, \cdot)$ it follows that

$$\mathbb{E}\left[\mathbb{1}\left(\, \pi(p, \mathbf{e}) - \pi(p, e) \le 0 \,\right) | \mathbf{p} = p, \mathbf{w} = w\right] = \mathbb{E}\left[\mathbb{1}\left(\, \mathbf{e} \le e \,\right) | \mathbf{p} = p, \mathbf{w} = w\right] .$$

The law of iterated expectations together with Assumptions D.1 and D.2 then imply that

$$\mathbb{P}\left(\boldsymbol{\pi} - \pi(\mathbf{p}, e) \le 0 | \mathbf{w} = w\right) = e \, .$$

$\blacksquare$

This lemma says that in the presence of endogeneity, we can still rank firms conditional on the instrumental variable. Note that Equation 11 is an integral equation that connects the unknown profit function, the distribution of observables, and productivity e. Indeed, Equation 11 can be rewritten as

$$\int_{P_w} F_{\boldsymbol{\pi}|\mathbf{p},\mathbf{w}}(\pi(p,e)|p,w) f_{\boldsymbol{p}|\mathbf{w}}(p|w)dp = e\,,$$

for all $w \in W$ and $e \in E$, where P_w denotes the support of $\mathbf{p}$ conditional on $\mathbf{w} = w$ and we assume the conditional p.d.f. of $\mathbf{p}$ conditional $\mathbf{w} = w$ exists for all w. The above integral equation has a unique solution in

$$\mathcal{L}^2(P) = \left\{ m(\cdot) \ : \ \int_P |m(x)|^2 dx < \infty \right\} ,$$

for every $e \in E$, if the operator $T_e : \mathcal{L}^2(P) \to \mathcal{L}^2(W)$ defined by

$$(T_e m)(w) = \int_{P_w} F_{\boldsymbol{\pi}|\mathbf{p},\mathbf{w}}(m(p)|p,w) f_{\boldsymbol{p}|\mathbf{w}}(p|w)dp,$$

is injective for every $e \in E$. Injectivity of integral operators is closely related to the notion of completeness. Numerous sufficient conditions for injectivity of integral operators are available in the literature.[37] Next we establish identification of $\pi(\cdot)$ based on the results of Chernozhukov and Hansen (2005).

Note that Equation 11 is equivalent to the IV model of quantile treatment effects of Chernozhukov and Hansen (2005). Thus we can directly invoke their identification result. For some fixed $\delta, \underline{f} > 0$, define the relevant parameter space $\mathcal{P}$ as the convex hull of functions $\pi'(\cdot, e)$ that satisfy: (i) for every $w \in W$, $\mathbb{P}\left(\boldsymbol{\pi} \leq \pi'(\mathbf{p}, e)|\mathbf{w} = w\right) \in [e-\delta, e+\delta]$, and (ii) for each $p \in P$,

$$\pi'(p,e) \in s_p = \left\{ \pi \ : \ f_{\boldsymbol{\pi}|\mathbf{p},\mathbf{w}}(\pi|p,w) \geq \underline{f} \text{ for all } w \text{ with } f_{\mathbf{w}|\mathbf{p}}(w|p) > 0 \right\} .$$

Moreover, let $f_{\boldsymbol{\epsilon}|\mathbf{p},\mathbf{w}}(\cdot|p,w;e)$ denote the density of $\boldsymbol{\epsilon} = \boldsymbol{\pi} - \pi(\mathbf{p}, e)$ conditional on $\mathbf{p}$ and $\mathbf{w}$. The following theorem follows from Theorem 4 in Chernozhukov and Hansen (2005).

Theorem D.1. *Suppose that*

[37] See for example Newey and Powell (2003), Chernozhukov and Hansen (2005), D'Haultfœuille and Février (2015), Andrews (2011), D'Haultfœuille (2011), and Hu et al. (2017).

(i) $\pi(p,\cdot)$ *is strictly increasing for every* $p \in P$*;*

(ii) *Assumptions D.1 and D.2 hold;*

(iii) $\boldsymbol{\pi}$ *and* $\mathbf{w}$ *have bounded support;*

(iv) $f_{\boldsymbol{\epsilon}|\mathbf{p},\mathbf{w}}(\cdot|p,w;e)$ *is continuous and bounded over* $\mathbb{R}$ *for all* $p \in P$*,* $w \in W$*, and* $e \in E$*;*

(v) $\pi(p,e) \in s_p$ *for all* $p \in P$ *and* $e \in E$*;*

(vi) *For every* $e \in E$*, if* $\pi', \pi^* \in \mathcal{P}$ *and* $\mathbb{E}\left[(\pi'(\mathbf{p},e) - \pi^*(\mathbf{p},e))\omega(\mathbf{p},\mathbf{w};e)|\mathbf{w}\right] = 0$ a.s.*, then* $\pi'(\mathbf{p},e) = \pi^*(\mathbf{p},e)$a.s.*, for* $\omega(p,w;e) = \int_0^1 f_{\boldsymbol{\epsilon}|\mathbf{p},\mathbf{w}}(\delta(\pi'(p,e) - \pi^*(p,e))|p,w;e)d\delta > 0$*;*

Then for any $\pi'(\cdot,e) \in \mathcal{P}$ *such that*

$$\mathbb{P}\left(\mathbb{1}\left(\,\boldsymbol{\pi} \leq \pi'(\mathbf{p},e)\,\right)|\mathbf{w} = w\right) = e$$

for all $w \in W$*, it follows that* $\pi'(\mathbf{p},e) = \pi(\mathbf{p},e)$ a.s..

D.2. Nonmonotonicity of Supply

Consider the following production sets that correspond to three different levels of productivity. $Y(e_i) = \{(y_o, l)' \in \mathbb{R} \times \mathbb{R}_+ \,:\, y_o \leq f_i(l)\}$, where

$$f_1(l) = l^{0.4}, \qquad f_2(l) = 2 \cdot l^{0.4}$$

and

$$f_3(l) = \begin{cases} l^{0.2} & 0.01 \geq l \geq 0 \\ 7 \cdot (l - 0.01) + 0.01^{0.2} & 0.03 \geq l \geq 0.01 \\ 2 \cdot l^{0.4} + 7 \cdot 0.02 + 0.01^{0.2} - 2 \cdot 0.03^{0.4} & 0.03 \leq l. \end{cases}$$

Note that by construction $f_1(l) < f_2(l) < f_3(l)$ for all $l > 0$. Hence, $Y(e_1) \subseteq Y(e_2) \subseteq Y(e_3)$ and $\pi(p,e_1) < \pi(p,e_2) < \pi(p,e_3)$ for all positive p. If one takes $p = (p_o, p_l)'$ such that $p_o/p_l = 0.12$, then the optimal levels of inputs and outputs are

$$0.007 > l_1^* = 0.048^{5/3} > 0.006, \quad 0.2 > y_{o,1}^* = 0.048^{2/3} > 0.1$$
$$0.03 > l_2^* = 0.096^{5/3} > 0.02, \quad 0.5 > y_{o,2}^* = 2 \cdot 0.096^{2/3} > 0.41$$

$$0.01 > l_3^* = 0.024^{5/4} > 0.009, \quad 0.40 > y_{o,3}^* = 0.024^{1/4} > 0.39.$$

Thus, for this price vector neither the optimal level of the input nor the optimal level of the output are monotone in productivity since $l_1^* < l_3^* < l_2^*$ and $y_{o,1}^* < y_{o,3}^* < y_{o,2}^*$.

D.3. Parametric Examples and Price Proxies

Section 3 shows that if prices are not observed but price proxies are, then it is possible to reproduce price variation from such proxies. The technique requires a high level rank condition. We present two examples to better understand this rank condition.

Example 6 (Diewert function, $d_y = 3$)**.** Let

$$\pi(p,e) = \sum_{s=1}^{3}\sum_{j=1}^{3} b_{s,j}(e) p_s^{1/2} p_j^{1/2}\,.$$

Suppose that p_3 is observed, and $p_1 = g_1(x_1)$ and $p_2 = g_2(x_2)$. Assume, moreover, that $\partial_{x_s} g_s(x_s) \neq 0$, for all x_s and $s = 1, 2$. Fix any x_1 and x_2. Then the rank condition is satisfied if and only if there exists e^{**} such that

$$\frac{b_{1,1}(e^{**})\sqrt{g_1(x_1)} + b_{1,2}(e^{**})\sqrt{g_2(x_2)}}{b_{2,2}(e^{**})\sqrt{g_2(x_2)} + b_{1,2}(e^{**})\sqrt{g_1(x_1)}} \neq \frac{b_{1,3}(e^{**})}{b_{2,3}(e^{**})}\,.$$

In particular, if $g_1(\cdot) = g_2(\cdot)$, then the rank condition is satisfied if and only if

$$\frac{b_{1,1}(e^{**}) + b_{1,2}(e^{**})}{b_{2,2}(e^{**}) + b_{1,2}(e^{**})} \neq \frac{b_{1,3}(e^{**})}{b_{2,3}(e^{**})}\,.$$

In Example 6 the rank condition is satisfied except for a set of parameter values with Lebesgue measure zero. However, as the following example demonstrates, the rank condition may fail to hold for all possible values of parameters.

Example 7 (Cobb-Douglas)**.** For a fixed e, let $y_o \leq k^\alpha l^\beta$ be such that $\alpha + \beta < 1$ and $\alpha, \beta > 0$. Then

$$\pi(p,e) = (1-\alpha-\beta)\left[\frac{p_k}{\alpha}\right]^{\dfrac{\alpha}{\alpha+\beta-1}} \left[\frac{p_l}{\beta}\right]^{\dfrac{\beta}{\alpha+\beta-1}} (p_o)^{-\dfrac{1}{\alpha+\beta-1}}\,,$$

where $p = (p_o, p_k, p_l)'$. Suppose that only p_o is perfectly observed. Suppose $p_k = g_k(x_k)$ and $p_l = g_l(x_l)$. Then for any two p_o^* and p_o^{**} let $p^* = (p_o^*, p_k, p_l)'$ and $p^{**} = (p_o^{**}, p_k, p_l)'$. The matrix $A(\tilde{\pi}, x^*)$ is singular since it is equal to

$$\begin{bmatrix} \frac{\alpha\pi(p^*, e)}{(\alpha+\beta-1)g_k(x_k)}\partial_{x_k} g_k(x_k) & \frac{\beta\pi(p^*, e)}{(\alpha+\beta-1)g_l(x_l)}\partial_{x_l} g_l(x_l) \\ \frac{\alpha\pi(p^{**}, e)}{(\alpha+\beta-1)g_k(x_k)}\partial_{x_k} g_k(x_k) & \frac{\beta\pi(p^{**}, e)}{(\alpha+\beta-1)g_l(x_l)}\partial_{x_l} g_l(x_l) \end{bmatrix}.$$

It can be shown that the rank condition is never satisfied for Cobb-Douglas production function if only one of the prices is perfectly observed.

The rank condition is not satisfied for the Cobb-Douglas production function because the ratios of any two different quantities chosen (e.g. l/k, or y_o/l) do not depend on the price of the quantity not described in the ratio. Indeed, recall that

$$\partial_{x_j}\tilde{\pi}(x, e) = y_j(g(x), e)\partial_{x_j} g_j(x_j)\,.$$

Thus, if $y_j(g(x), e)/y_s(g(x), e)$ does not depend on observed price p_{d_y}, then the s-th column of $A(\tilde{\pi}, x^*)$ is a scaled version of the j-th column of $A(\tilde{\pi}, x^*)$. Hence, $A(\tilde{\pi}, x^*)$ is singular.

Note that, alternatively, the rank condition can be stated in terms of $\log(\tilde{\pi}(x, e))$. Indeed, the main identifying Equation (7) can be equivalently rewritten as

$$\sum_{j=1}^{d_y} \partial_{x_j}\log(\tilde{\pi}(x, e))\frac{1}{\partial_{x_j}\log(g_j(x_j))} = 1.$$

Hence, if $\log(\tilde{\pi}(x, e))$ is additively separable in the observed price (as in Example 7), then the variation in the observed price cannot be used to identify g, since $\partial_{x_j}\log(\tilde{\pi}(x, e))$ does not depend on it. For the Cobb-Douglas production function, however, an alternative identification result can be established. Note that for the profit function in Example 7, the following partial differential equation holds

$$\partial_{p_k}\log(\pi(p, e)) = \frac{\alpha}{\alpha+\beta-1}\partial_{p_k}\log(p_k)\,.$$

Hence,

$$\partial_{x_k}\log(\tilde{\pi}(x, e)) = \frac{\alpha}{\alpha+\beta-1}\partial_{x_k}\log(g_k(x_k))$$

and g_k is identified from $\partial_{x_k}\tilde{\pi}(x, e)$, if $g_k(x_{k,0})$ and $\partial_{x_k}\log\left(g_k(x'_{k,0})\right)$ are known for some $x_{k,0}$ and $x'_{k,0}$. Thus, if it is known that the production function is Cobb-Douglas, then, even if we do not observe any price, we still can identify g up to a location and scale normalization.

D.4. Point Identification and Assumption 6

It is natural to wonder when Assumption 6 is necessary and sufficient for point identification of $Y(\cdot)$. Unfortunately, this question is technical. It is essentially equivalent to asking when the function π_P, defined as π restricted to $P \times E$, has a *unique* extension $\tilde{\pi} : \mathbb{R}^{d_y}_{++} \times E \to \mathbb{R}^{d_y}$ such that $\tilde{\pi}$ is homogeneous of degree 1, convex, and satisfies $\tilde{\pi}(p, e) = \pi(p, e)$ for every $(p', e)' \in P \times E$.

First, we note that by exploiting continuity and homogeneity of degree 1, we know that there is a unique extension of π_P to the set

$$\operatorname{int}\left(\operatorname{cl}\left(\bigcup_{\lambda>0}\{\lambda p \,:\, p \in P\}\right)\right) \times E$$

that satisfies the properties described above. It is, however, possible that this set is strictly nested in $\mathbb{R}^{d_y}_{++} \times E$, and yet there is a unique extension of π_P to all of $\mathbb{R}^{d_y}_{++} \times E$.

Example 8 (Unique Extension without Assumption 6)**.** Consider $\pi(p, e) = e \sum_{j=1}^{d_y} |p_j|$ with $E = [0, M]$, $0 < M < \infty$. This functions is homogeneous of degree 1 and convex in p, and hence the profit function for price-taking firms, indexed by e (Kreps, 2012, Proposition 9.14). Let $\Delta^{d_y-1} = \{p \in \mathbb{R}^{d_y}_{++} \,:\, \sum_{j=1}^{d_y} p_j = 1\}$ denote the relative interior of the probability simplex, and let $S = \{p \in \Delta^{d_y-1} \,:\, |y_j - 1/d_y| \leq 1/d_y \text{ for each } j\}$ denote a convex set centered at the midpoint of the simplex. Let P be the probability simplex with the region S removed, i.e. $P = \Delta^{d_y-1} \setminus S$. Note that P is a subset of the affine space $\{p \in \mathbb{R}^{d_y} : \sum_{j=1}^{d_y} y_j = 1\}$, and $\pi_P(\cdot, e)$ is equal to e over P. Any convex extension of $\pi_P(\cdot, e)$ to the convex hull of P, Δ^{d_y-1}, must also be equal to e. In more detail, there is a unique such extension because Δ^{d_y-1} has dimension $d_y - 1$ (i.e. the smallest affine space containing this set has dimension $d_y - 1$). Because there is a unique convex extension of $\pi_P(\cdot, e)$ to all of Δ^{d_y-1}, there is a unique convex and homogeneous extension to all of $\mathbb{R}^{d_y}_{++}$. By Corollary 9.18 in Kreps (2012) the production correspondence is identified even though Assumption 6 fails to hold.

For additional geometric intuition behind this example, consider a line segment from $(0, 0)$ to $(1, 0)$ in $\mathbb{R}^2$. If one deletes a chunk out of the middle of this line segment, but maintains each endpoint, then the convex hull of this modified set is actually the original set.

This example also shows that it is possible to uniquely determine $\pi(p, e)$ at values p that are not in the set $\operatorname{int}\left(\operatorname{cl}\left(\bigcup_{\lambda>0}\{\lambda p \,:\, p \in P\}\right)\right)$. We are only able to construct "knife edge" examples in which the support restriction of Assumption 6 is *not* equivalent to point identification of $Y(\cdot)$. We note that strict convexity of $\pi(\cdot, e)$ rules out this sort of example.

D.5. Proxies with Other Observables

Theorem 1 applies when (only) values of the restricted profit function, restricted variables, and price proxies are observed. When other variables are observed, it can be adapted to handle other settings.

To illustrate this suppose that p_1 is not observed, does not have a proxy, and does not vary across markets. Suppose further that for each good $j \geq 3$, x_j and p_2 determine the price of good j. That is, $p_j = g_j(x_j, p_2)$ for $j \geq 3$. The fact that p_2 is in this function means it violates our previous exclusion restriction and so Theorem 1 cannot directly be applied. Nonetheless, we can adapt the technique to cover this case.

To see this, recall Euler's homogeneous function theorem states

$$\sum_{j=1}^{d_y} \partial_{p_j} \pi(p, e) p_j = \pi(p, e),$$

while Hotelling's lemma reads

$$y_j(p, e) = \partial_{p_j} \pi(p, e).$$

These imply

$$\sum_{j=3}^{d_y} \partial_{p_j} \pi(p, e) p_j = \pi(p, e) - p_1 y_1(p, e) - p_2 y_2(p, e)\,.$$

Moreover, for $j \neq 2$

$$\partial_{p_j} \pi(g(x, p_2), e) \partial_{x_j} g_j(x_j, p_2) = \partial_{x_j} \tilde{\pi}(x, p_2, e)\,.$$

Hence we obtain

$$\sum_{j=3}^{d_y} \partial_{x_j} \tilde{\pi}(x, p_2, e) \frac{g_j(x_j, p_2)}{\partial_{x_j} g_j(x_j, p_2)} = \tilde{\pi}(x, p_2, e) - \tilde{r}(x, p_2, e)\,, \tag{12}$$

where

$$\tilde{r}(x, p_2, e) = p_1 y_1(p_1, p_2, g(x, p_2), e) + p_2 y_2(p_1, p_2, g(x, p_2), e)$$

is the contribution of goods 1 and 2 to profits. The difference from Equation 2 is that in order to build a system of ordinary differential equations that identifies g we need to identify

$$\tilde{\pi}(x, p_2, e) - \tilde{r}(x, p_2, e)$$

as well. If $\tilde{\pi}(x, p_2, e)$ and $\tilde{r}(x, p_2, e)$ can be identified, we are done. More generally, it is not necessary to identify the heterogeneous structural functions separately. Instead, it is enough to identify their aggregate versions, because homogeneity is preserved under expectations. See Appendix D.6 for formal details. In sum, it is possible to identify prices that vary across markets even if there are prices that are unobserved but are fixed across markets (like p_1) and there are observed prices that do not satisfy the exclusion restriction (like p_2). We use this insight about prices that are unobserved but fixed across markets in Section 3.1 to show how our approach can be used to generalize Epple et al. (2010).

D.6. Proxies with Representative Agent

This section studies identification given price proxies, as in Section 3. Here we present a representative agent counterpart of Theorem 2 that is valid regardless of whether productivity is discrete or not. This is because it does not require identification of the type-specific structural mapping $\tilde{\pi}_r$, but instead only the average of this mapping.

Theorem D.2. *Suppose Assumption 5 holds. Then g is identified over the support of $\mathbf{x}$ if for some y^*_{-z}, the following conditions hold:*

(i) $\boldsymbol{\pi}_r = \pi_r(\mathbf{y}_{-z}, g(\mathbf{x}), \mathbf{e}) + \boldsymbol{\eta}$ a.s., *where $\boldsymbol{\pi}_r$ is observed.*

(ii) $F_{\mathbf{e}|\mathbf{p}_z,\mathbf{y}_{-z}}(e|p_z, y^*_{-z})$ *is homogeneous of degree 0 in p_z.*

(iii) $\mathbb{E}\left[\boldsymbol{\eta} \mid \mathbf{x}, \mathbf{y}_{-z}\right] = 0$ *a.s.*

(iv) $\mathbb{E}\left[\boldsymbol{\pi}_r \mid \mathbf{x} = \cdot, \mathbf{y}_{-z} = y^*_{-z}\right]$ *satisfies the rank condition of Definition 2, at every $x_{-1} \in \prod_{j=2}^{d_{y_z}} X_j$.*

Proof. Part (ii) means that for all e and p_z in the support and $\lambda > 0$,

$$F_{\mathbf{e}|\mathbf{p},\mathbf{y}_{-z}}(e|\lambda p_z, y^*_{-z}) = F_{\mathbf{e}|\mathbf{p},\mathbf{y}_{-z}}(e|p_z, y^*_{-z}).$$

Thus, the function

$$\mathbb{E}\left[\pi_r(\mathbf{y}_{-z}, \mathbf{p}_z, \mathbf{e})|\mathbf{p}_z = \cdot, \mathbf{y}_{-z} = y^*_{-z}\right]$$

inherits homogeneity of degree 1 in p_{-z} from π_r. From parts (i) and (iii),

$$\mathbb{E}\left[\boldsymbol{\pi}_r|\mathbf{x} = \cdot, \mathbf{y}_{-z} = y^*_{-z}\right] = \mathbb{E}\left[\pi_r(\mathbf{y}_{-z}, \mathbf{p}_z, \mathbf{e})|\mathbf{p}_z = g(\cdot), \mathbf{y}_{-z} = y^*_{-z}\right].$$

Thus, the result follows from applying the same arguments as in the proof Theorem 2, except to the function $\mathbb{E}\left[\boldsymbol{\pi}_r|\mathbf{x} = \cdot, \mathbf{y}_{-z} = y^*_{-z}\right]$ instead of $\tilde{\pi}(\cdot)$. ■

Part (i) formalizes that we modify the baseline setup by replacing prices with price proxies. Note that here we do not restrict $\mathbf{e}$ to be discrete, or the structural mapping π_r to be monotone in $\mathbf{e}$. Part (ii) is implied by conditional independence between $\mathbf{e}$ and $\mathbf{p}_z$, but is weaker. Instead it requires the conditional distribution of productivity in the market depends only on relative prices. This trivially happens if productivity is independent from flexible prices conditional on fixed quantities. It also may naturally hold in profit maximizing environments where entry decisions are driven by the threshold rule where firms with nonnegative profits enter. Since the profit function is homogeneous of degree 1 in prices, we can deduce that the entry decision is only determined by the direction of the price vector, not by its norm: $\pi(p, e) \geq 0$ if and only if $\pi(p/\left\|p\right\|, e) \geq 0$. Part (iii) requires that η is not systematically related to other variables. Part (iv) is a relevance condition.