// NO.24-049 | 07/2024

DISCUSSION PAPER

// NADINE HAHN

Product Differentiation and Quality in Production Function Estimation

ZEW

Product Differentiation and Quality in Production Function Estimation

Nadine Hahn*

Tuesday 16th July, 2024

Abstract

Production functions provide a mapping from the firms' input quantity and productivity to output quantity. This mapping only generates unbiased estimates if input and output quality variation within and between observation units is accounted for. I review and classify state-of-the-art methods to address quality and price variation in production function estimation. Even if inputs and outputs are observed in quantities, unobserved quality variation might bias production function estimates for industries with differentiated products. To account for quality variation, I introduce product characteristics to the estimation procedure and provide an application to the European car industry.

Keywords: *Production Functions, Product Differentiation, Quality Control Functions*

JEL Codes: D22, D24, C38, B41, L62

*KU Leuven, ZEW Mannheim & MaCCI, nadine.hahn@kuleuven.be.
This paper constitutes the second chapter of my dissertation at KU Leuven. I am very thankful to Jan De Loecker and Jo Van Biesebroeck for their advice. I particularly want to thank Bernhard Ganglmair, Pinelopi Goldberg, and Frank Verboven for helpful comments and feedback.

1 Introduction

The methodology of production function estimation gained considerable traction in recent years. One of the reasons for the popularity of this approach is that it mainly relies on data from the firms' balance sheets, which are widely available across many industries. Nonetheless, if the estimation procedure is applied to industries with differentiated products, unbiased estimation hinges on the inclusion of additional controls for latent price and quality variation.

In this paper, I provide an overview of the main biases that occur when price and quality variation are not fully accounted for in production function estimation. Building on earlier work on price biases by De Loecker and Goldberg (2014), I extend the analysis to the quality bias. This bias received comparatively little attention from previous literature and refers to latent variation in input and/or output quality. As pointed out by De Roux et al. (2021), even if inputs and outputs were observed in quantities, the quality bias might occur for industries with differentiated products. I systematically categorize the state-of-the-art approaches to address unobserved price and quality variation and discuss the inclusion of product characteristics as alternative control for unobserved quality variation in production functions. I provide an application to the European car industry in a Leontief production setting.

Because of their interdependence, I jointly discuss and qualify both the price and quality biases. In purely vertically differentiated industries, controlling for price variation might fully control for quality variation under certain conditions. For highly differentiated industries, however, controlling for both biases becomes challenging and requires additional assumptions and/or controls depending on data availability.

Input and output price biases might occur when establishment-level prices are unobserved, leading researchers to estimate production functions based on expenditures rather than physical units. This introduces latent price variation, which might (1) confound the productivity measure with market power and demand shocks, and (2) be correlated with the firms' input demand. I provide two examples of how these biases might lead to wrong conclusions on firm performance. Firstly, if the latent price variation stems from market power, it influences the firms' productivity evolution. Consequently, firms that seem highly productive produce relatively little output and appear technologically inefficient (Foster, Haltiwanger, and Syverson, 2008). Secondly, latent price variation might stem from demand responses to trade frictions. This suggests an inaccurate relationship between productivity and trade, potentially generating a negative correlation between the firms' export status and productivity (De Loecker, 2011).

The quality bias might occur for differentiated product industries even if inputs and outputs are observed in physical quantities. In this case, unobserved quality variation enters the production function and biases the estimates when correlated with input demand. Consider the automotive industry: two car manufacturers produce the same quantity of cars, but require different quantities of inputs depending on the car model that they produce. If the required input quantity increases with product quality (e.g., additional sound system, navigation system, etc.), the production function coefficients turn negative. In this case, variation in input quantity does not correspond to a proportionate variation in output quantity.

Previous literature developed various approaches to tackle these biases, each tailored to different industry contexts and data availability. I classify the approaches into six categories. The first category requires a one-to-one mapping from quality variation to price variation for vertically differentiated products. The second category relies on additional assumptions on firm conduct, which lead to the situation where the input price bias and output price bias cancel out each other in a sales-generating production function.[1] The third category controls for plant-level price variation by including industry-level price indices. This approach implies that firms produce homogeneous products in perfectly competitive markets. The fourth category includes plant-level price variation in the estimation procedure directly. Controlling for price variation also fully controls for quality variation if products are only vertically differentiated. The fifth category augments the production function with a demand system and assumptions on price setting as in Klette and Griliches (1996). The sixth category expresses unobserved quality variation with observables without parametrically committing to a demand system, based on insights from Berry (1994).

I introduce product characteristics to the production function to control for quality variation in two ways. First, adding characteristics allows to control for quality variation in industries with changing product characteristics over time. Consider again the car industry: The average fuel efficiency, horsepower, and physical size of SUVs and Minis vary over time, both within and between the car segments. For industries with vertically and horizontally differentiated products, variation in product quality is not necessarily reflected in a proportionate increase in product or input prices and might be correlated with input demand. As a result, unobserved quality variation might affect the production function estimates. Second, in a production framework that requires a fixed proportion between inputs as in a Leontief setting, including product characteristics allows to adjust the fixed proportion of inputs according to the output that the firms produce. For the car industry, this allows for varying fixed shares of automotive parts that are required to produce one car, depending on the individual cars' characteristics. For instance, a manufacturer of hybrid SUVs with convertible roofs might require a higher share of material inputs to produce one car compared to a manufacturer of gasoline Minis.

[1] These are: (i) monopolistic competition, (ii) vertical product differentiation with firms facing the same CES demand, (iii) input neutral input price variation, and (iv) constant returns to scale.

Production function estimation is an increasingly popular methodology, which is applied in a substantial body of literature. It is typically conducted for two main objectives (1) to retrieve the firms' output elasticities of variable inputs, which are used to construct markup and markdown estimates as measures for market power, and (2) to estimate total factor productivity. Recent contributions employing this methodology relate markups to common ownership (Gibbon and Schain, 2023), export status (De Loecker, 2013) or trade-barriers (De Loecker et al., 2016). The rising interest in input market power led to a surge of contributions that measure markdowns relying on production function estimates, either to retrieve measures for labor market power (Mertens, 2023), buyer power towards input suppliers (Avignon and Guigue, 2022), or both (Treuren, 2022). Contributions to the productivity literature that rely on production function estimation evaluate productivity drivers, for instance, R&D expenditure (Doraszelski and Jaumandreu, 2013), firm ownership (Braguinsky et al., 2015), managerial practices (Rubens, 2022), or learning by experimenting (Hendel and Spiegel, 2014).

In the classification of current approaches to control for price and quality variation, I abstract from technology specifications and identification strategies of production function coefficients. For the empirical application, I provide an illustration based on a Leontief production setting.[2] The provided literature review does not claim completeness of all contributions, but for ease of exposition rather comprises a selected overview of key contributions. I also abstract from the so-called input allocation bias that arises for multi-product firms. The increasing literature on this matter can be divided into three approaches: (1) defining input-to-output allocation rules (e.g., De Loecker et al., 2016; Orr, 2022; Valmari, 2023; Itoga, 2019), (2) estimating transformation functions between inputs and outputs (e.g., Dhyne et al., 2020; Maican and Orth, 2021), or (3) employing observed information on joint production inputs across outputs from data (Raval, 2024). I refer the reader to these contributions for a comprehensive overview of multi-product production function estimation.

The remainder of the paper is structured as follows. In section 2, I set the baseline notation for the production function estimation. In section 3, I discuss the price and quality biases. In section 4, I review and classify the state-of-the-art solutions to account for the biases. In section 5, I present the alternative solution of including product characteristics. In section 6, I provide the empirical framework for the production function estimation procedure including product characteristics in a Leontief production setting, which is followed by a description of the dataset in section 7. I present the results in section 8. Section 9 provides a discussion of the results and outlook for future applications in highly differentiated product industries. I conclude in section 10.

[2]For the detailed explanations of the structural control function approach, I refer to the seminal contributions by Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves, and Frazer (2015), and I refer to Blundell and Bond (1998) for the dynamic panel approach.

2 Production Function: Baseline and Notation

In this section, I provide the framework and notation that is employed throughout the paper and discuss the production function specifications that are taken to the data based on varying data availability and product differentiation. I follow the notation by De Loecker and Goldberg (2014).

The production function of firm i at time t relates quantities produced Q_{it} to a vector of physical inputs X_{it}. Both physical inputs and outputs are comparable units within and between firms. For simplicity, I abstract from the technology specifications of the production function. Lower letters indicate logarithmic transformations. I define the standard Hicks-neutral production function as:

$$q_{it} = x_{it}\alpha + \omega_{it} + \varepsilon_{it}, \tag{1}$$

with α being the vector of production function coefficients, ω_{it} representing the unobserved productivity of firms, and the error term ε_{it}.

Most datasets do not contain physical quantities of inputs and outputs. For this reason, instead of estimating equation (1), researchers typically employ sales as dependent variable instead of q_{it} and input expenditure instead of x_{it}. Sales are denoted by S_{it}, which represents the output price P_{it} multiplied with output quantity $S_{it} = P_{it}Q_{it}$. Input expenditures are denoted by E_{it}, which represents the input price Z_{it} multiplied with input quantity $E_{it} = Z_{it}X_{it}$. Inserting s_{it} and e_{it} into (1) yields the following equation that is typically taken to the data:

$$s_{it} = e_{it}\alpha + \omega_{it} + p_{it} - z_{it}\alpha. \tag{2}$$

Variation in product quality might enter the production function not only through input and output price variation but also through input quantities X_{it} and output quantities Q_{it}, which for many industries are not comparable within and between firms. To provide an example, consider again the car industry: even if input and output quantities were observed, substantial quality variation of cars within and between segments adds unobserved variation to the production function.

Drawing on insights from the literature on hedonic prices and quantities (e.g., Triplett, 1969; Griliches, 1964; Griliches, 1971), I define the plants' comparable output quantity as $q_{it} = \hat{q}_{it} + \hat{q}^H_{it}$, where $\hat{q}_{it}$ is the observed physical quantity and $\hat{q}^H_{it}$ is a hedonic quantity index, which rescales physical quantities to comparable units between plants. Similarly, I define comparable input quantity as $x_{it} = \hat{x}_{it} + \hat{x}^H_{it}$. This allows to reformulate equation (2) as:

$$s_{it} = e_{it}\alpha + \omega_{it} + p_{it} - z_{it}\alpha + \hat{q}^H_{it} - \hat{x}^H_{it}\alpha. \tag{3}$$

Equation (3) shows that in cases where input and output price variation is fully correlated with product quality variation, controlling for price variation p_{it} and z_{it} suffices as a control for quality variation. In cases where products are homogeneous and no input and output price variation between plants exists, equation (3) collapses to equation (1). If products are highly differentiated and input and output prices vary within and between plants, endogenous variation in p_{it}, z_{it}, $\hat{q}_{it}^{H}$, and $\hat{x}_{it}^{H}$ might bias the estimated coefficients α and productivity ω_{it}.

3 Output/Input Prices and Product Quality Bias

In this section, I introduce the biases that occur when the endogenous components of equation (3) are not accounted for in the estimation procedure.

Output Price Bias. First, I focus on unobserved output price variation in p_{it} only. This situation might give rise to the so-called output price bias. Instead of mapping input to output quantities, the production function represents a mapping of input quantities to sales:

$$s_{it} = x_{it}\alpha + p_{it} + \omega_{it} + \varepsilon_{it}, \tag{4}$$

where the unobserved structural error term $\omega_{it} + \varepsilon_{it}$ additionally contains variation in output prices p_{it}. Variation in p_{it} is most likely correlated with input demand x_{it}, which leads to biased production function coefficients α. I provide an example based on a monopoly setting: Firms that charge higher prices p_{it} sell lower quantities q_{it}, which implies a negative correlation between prices p_{it} and input demand x_{it}. Since p_{it} is unobserved, it constitutes an omitted variable that is correlated with x_{it}. A detailed discussion on the output price bias is provided in De Loecker (2011), Klette and Griliches (1996), and Katayama, Lu, and Tybout (2009).

Input Price Bias. Second, I focus on unobserved input price variation in z_{it} only. This situation might give rise to the so-called input price bias. Instead of mapping input to output quantities, the estimation procedure provides a mapping of input expenditures to output quantities:

$$q_{it} = e_{it}\alpha + \omega_{it} - z_{it}\alpha + \varepsilon_{it}, \tag{5}$$

where the unobserved structural error term additionally contains variation in input prices z_{it}. Estimating equation (5) leads to a negative bias and possibly wrong sign of the estimated production function coefficients α. Keeping the demanded input quantity fixed, a firm that faces higher input prices z_{it} faces higher input expenditure $z_{it}x_{it}$. This, however, does not necessarily imply an increase of physical output q_{it}.

To illustrate this notion, consider two car manufacturers, Company A and Company B. Both produce the same car, but Company A uses an underbody that is made of higher quality metal. The higher quality underbody assembled by Company A is more expensive than the lower quality underbody assembled by Company B. This creates higher input expenditure per car for Company A. Let us assume further that both companies produce the same amount of cars. Relating input expenditure to output quantity (as in equation (5)) for this example links higher input expenditure to constant output quantity. The estimated production function coefficient on material input expenditure would be negative because increases in input prices z_{it} are correlated with constant output quantity q_{it}.

Quality Bias. Third, I focus on unobserved quality variation. Let us assume both input and outputs are observed in physical quantities for differentiated products. In this case, quality variation enters the production function as follows:

$$\hat{q}_{it} = \hat{x}_{it}\alpha - \hat{q}^H_{it} + \hat{x}^H_{it}\alpha_{it} + \omega_{it} + \varepsilon_{it}. \tag{6}$$

Similarly to the notion of the input and output price biases, unobserved quality variation in output $\hat{q}^H_{it}$ and input quantities $\hat{x}^H_{it}$ might lead to biased production function coefficients α, even if inputs and outputs were measured in physical quantities. This is particularly the case when products are vertically and horizontally differentiated, and a one-to-one mapping between quality and sold quantity does not exist. De Roux et al. (2021) provide the following example for the quality bias. Consider a firm that produces higher-quality products, which differ from the products of other firms only in the sense that they require more labor hours per physical unit of output. In this case, production function estimates that rely on the work hours as inputs and output quantity would result in a downward biased labor coefficient. It would underestimate the role of labor in the production of the quality-adjusted output. De Loecker and Syverson (2021) discuss the associated challenges with measuring inputs and outputs in comparable quantities. As pointed out by the authors, constructing comparable units is particularly challenging when output is abstract (e.g., output of the financial sector), immeasurable (e.g., production of intangible capital), or affected by unobserved delays between input utilization and output generation (e.g., health improvement after medical treatment).

Both: Price and Quality Biases. Because they are highly interlinked, none of the presented biases typically occurs in isolation. This is particularly the case for industries with highly differentiated products where both outputs and inputs are observed in expenditures rather than comparable units. As pointed out by Kugler and Verhoogen (2012), producers of more expensive products typically use more expensive inputs. In this case, unobserved variation in input and output prices might offset each other. Depending on the degree of product differentiation, controlling for price variation might also fully control for quality variation. If input and output quality are not perfectly correlated, however, the quality bias is not automatically accounted for.

Going back to the car example, this implies that if input and output quality are perfectly correlated and the cars of Company A yield higher revenues, increasing revenues generated from sales might offset the increasing material input expenditure.

4 State-of-the-Art: Classification

In this section, I discuss the state-of-the-art solutions to control for the biases presented in the previous section. I classify them into six categories: (1) relying on sales-generating production functions for horizontally differentiated product industries in perfectly competitive markets, (2) relying on sales-generating production functions, input expenditure information, and setting additional assumptions on pass-through, (3) employing industry-level price deflators, (4) constructing plant-level prices, (5) adding demand systems, and (6) expressing price and quality variation based on observables.

Some of the mentioned categories might be jointly applied to estimate production functions in particular settings. For instance, combining sales-generating production functions and industry-level input price deflators in settings with perfectly competitive product and inputs markets and vertically differentiated products yields unbiased production function estimates.[3]

4.1 Sales-Generating Production Functions

As pointed out by De Loecker and Ackerberg (2024), there are instances where estimating production functions solely based on physical quantities instead of revenue information leads to biased estimates, especially when physical quantities are not comparable. In these situations, including prices in the form of revenues controls for product quality variation if quality variation can be fully mapped into output price variation. In these settings, including revenue information, essentially introducing output prices to the production function, helps to account for quality variation.

To provide a hypothetical example for the car industry, consider again Company A and B. In this scenario, they produce a similar model, but the cars produced by Company A are 10 percent bigger than the cars produced by Company B. The sold quantity of cars masks this variation and thus covers unobserved variation in product quality. The relationship between input quantity and output quantity is not a direct one-to-one mapping conditional on productivity but is instead

[3] De Loecker and Syverson (2021) discuss production function estimation with differentiated products and present the categories (2), (3), (4), and (5).

contingent upon productivity and car size. Including revenue information instead of output quantity controls for variation in car size under the assumption of perfectly competitive markets.[4]

4.2 Assumptions on Pass-Through of Price Variation

If inputs and outputs are observed as expenditures and sales, identification of the production function coefficients is possible even if firm-level input and output price variation cannot be controlled for in the estimation procedure directly. De Loecker and Goldberg (2014) provide a detailed discussion of this matter. Let us assume a sales-generating production function with input expenditure data:

$$s_{it} = e_{it}\alpha + \omega_{it} + p_{it} - z_{it}\alpha + \varepsilon_{it}. \tag{7}$$

The production function coefficients are identified if the output price bias fully offsets the input price bias with $p_{it} - z_{it}\alpha = 0$. As discussed by De Loecker and Goldberg (2014), this situation arises when additional assumptions hold, that is (1) monopolistic competition on product markets, (2) firms produce horizontally differentiated products, facing the same constant elasticity of substitution (CES) demand system, (3) input neutral input price variation, with $z_{it}^{h} = \lambda_{it} \forall h = 1, 2...., H$. H represents the total number of inputs, and (4) constant returns to scale (CRS). I provide further background on the relevance and intuition of assumptions (1)-(4) in Appendix A.

4.3 Employing Industry-Level Price Deflators

If input and/or outputs are observed as expenditures and sales, latent input and output price variation could be controlled for by approximating plant-level price variation in p_{it} with industry-level price deflators $\bar{p}_t$. These deflators are widely available and constructed by many data providers as price indices. In logarithmic transformation, the comparable output quantity q_{it} is constructed with deflated output prices $p_{it}^{*} = p_{it} - \bar{p}_t$, such that deflated sales is represented by $s_{it}^{*} = s_{it} - \bar{p}_t$. In this case, equation (7) can be rewritten as the following:

$$s_{it}^{*} = x_{it}\alpha + \omega_{it} + p_{it}^{*} + \varepsilon_{it}. \tag{8}$$

Similarly, input price variation in z_{it} might be approximated with industry-level price deflators $\bar{z}_t$. The deflated input price is then represented by $z_{it}^{*} = z_{it} - \bar{z}_t$. If both inputs and outputs are constructed based on expenditure information deflated with industry-level price deflators, the following equation is taken to the data:

[4]Another example are cookies that are sold in two package sizes: small packs and family-size packs. The sold quantity of cookie packs covers unobserved quality information (size of the packaging, here small vs. family-sized).

$$q_{it} = e_{it}\alpha + \omega_{it} + p^*_{it} - z^*_{it}\alpha + \varepsilon_{it}. \tag{9}$$

Klette and Griliches (1996) show that estimating equation (9) might lead to omitted variable biases if plant-level prices deviate from the deflators. Including price indices fully controls for latent price and quality variation only if (i) firm-level prices equal the price deflators with $\bar{p}_t = p_{it}$ and $\bar{z}_t = z_{it}$, or (ii) deviations of plant-level prices from the price deflators are i.i.d, so that they add to the error term ε_{it}, or (iii) remaining input and output price variation cancel out each other, and (iv) there is no latent quality variation in inputs and outputs that is correlated with the estimated production function coefficients or productivity.

Gonzales, Lach, and Miles (2021) discuss production function estimation relying on price indices instead of prices in levels. Similar to the notion in Klette and Griliches (1996), the authors argue that price indices introduce omitted variables which are the unobserved output price levels in the base years of the price indices.[5]

4.4 Employing Plant-Level Prices

If prices deviate from industry-level price indices (e.g., settings with imperfect competition), researchers typically either (i) construct plant-level prices in levels or indices from sold quantities and product-level prices, (ii) observe plant-level prices or price indices in the data directly, or (iii) construct plant-level prices from sold output quantities and sales. The price variable is either constructed for output prices (to control for the output price bias), for input prices (to control for the input price bias), or both.

Contributions where plant-level prices are either constructed in levels or as indices are for instance Collard-Wexler and De Loecker (2015), Smeets and Warzynski (2013), and Foster, Haltiwanger, and Syverson (2008). To control for output price variation, Collard-Wexler and De Loecker (2015) employ sold output quantities and product-level prices. To control for input price variation, the authors construct a plant-level input price index that relies on a weighted average of intermediate input-specific prices. This approach is based on the assumption that products within a product category are homogeneous, thus the main quality difference exists among product categories. It implies that quality variation within product categories is driven by the cost side, for instance through productivity differences.[6]

[5]The authors propose two solutions to this issue: (i) including fixed effects to account for unobserved base year prices in the notion of Gandhi, Navarro, and Rivers (2020), (ii) incorporating robustness checks to show that the results are independent of the chosen base year of the price indices.

[6]Similarly, Smeets and Warzynski (2013) construct a firm-level output price index based on market-level price data. They rely on a Tornqvist price index as a weighted average of the growth in prices for individual products. Foster, Haltiwanger, and Syverson (2008) employ output quantities and product-level prices to construct plant-level prices for homogenous product industries.

Contributions where firm-specific price indices are observed in the data are for instance Pozzi and Schivardi (2016) and Doraszelski and Jaumandreu (2013). Pozzi and Schivardi (2016) observe survey data on firm-level output price indices, where firms disclose the "average percentage change in the prices of goods sold". The authors assume homogeneity of the final goods produced within sectors.[7] Doraszelski and Jaumandreu (2013) observe firm-specific Paasche price indices for outputs and material inputs.

Contributions where plant-level prices are constructed from physical output quantities and sales are for instance Forlani et al. (2016), Dhyne et al. (2020), and Braguinsky et al. (2015). This approach is typically employed for homogenous product industries. Forlani et al. (2016) employ information on the output quantity of specific products[8] and the Euro-value of the products sold. Dhyne et al. (2020) use monthly product-specific revenues and quantities.

4.5 Specifying Demand Systems

Klette and Griliches (1996) emphasize that production function estimates are reduced form parameters, which comprise variation driven by both the production function and the demand system. They provide a general framework that augments the production function with a demand system and assumptions on conduct, which allows for separation between the supply-side and demand-side drivers of the estimates. De Loecker (2011) applies the notion of Klette and Griliches (1996) and estimates production functions without information on plant-level prices or quantities.[9] The author controls for unobserved prices and demand shocks in the production function by combining the production approach with a CES demand system of the following form:

$$Q_{it} = Q_{st} \left(\frac{P_{it}}{P_{st}}^{\eta_s} \right) exp\left(\xi_{it}\right). \tag{10}$$

The demand system relates the firms' output quantity Q_{it} to its product-level prices (P_{it}), the average segment-level product prices (P_{st}), the aggregate total demand in a given segment (Q_{st}), and unobserved demand shocks (ε_{it}). The parameter η_s represents the demand elasticity for segment s.[10]

De Loecker (2011) reformulates the demand system for the product-level prices P_{it}, inserts

[7] In their setting, this refers to three major sectors: Textile and Leather, Metal, and Mechanical and Electronic Machinery.

[8] They consider a variety of industries from the C and D NACE classification, which comprise 4500 distinct 8-digit product categories within the PRODCOM classification.

[9] In his setting, the author diverts from a single-parameter demand model as in Klette and Griliches (1996) to a more flexible demand system, where producers of textiles might face different demand elasticities depending on the product segment in which they produce.

[10] As pointed out by De Loecker (2011), the CES demand system in combination with monopolistic competition implicitly assumes constant markups over marginal cost for every segment ($\frac{\eta_s}{\eta_s}$+1).

the expression into the firms' revenues $R_{it} = P_{it}Q_{it}$, and combines it with a revenue-generating production function. These steps allow to separate the demand and supply-driven components of the output elasticities in the production function.

Other contributions that combine the production approach with demand systems focus on controlling for multiproduct firms in the estimation procedure, for instance Valmari (2023) and Orr (2022).[11]

4.6 Expressing Price Variation Based on Observables

De Loecker et al. (2016) express unobserved input price and quality variation based on output price variation and other observables, without parametrically committing to a demand system or market structure. This is particularly advantageous in settings where demand and firm conduct are endogenous to the research question.[12]

Relying on insights from Berry (1994), the authors show that for a variety of demand models and conduct assumptions, input price variation can be proxied by output price variation, market shares, and product dummies. The general notion is that input prices are an increasing function of input quality, which in turn is an increasing function of output quality. The underlying assumptions are that (i) the production of high-quality outputs requires high-quality inputs, (ii) higher-quality inputs are more expensive than lower-quality inputs, and (iii) the quality of different inputs is complementary. The last assumption implies that for instance producing a high quality car requires the combination of high-quality materials, assembled by highly skilled employees on high-quality production lines.[13] This assumption is necessary to ensure that the prices of different inputs can be expressed as a single index of product quality, and therefore it can be linked to output price variation.

A crucial assumption for their framework is that the input prices do not depend on the input quantity demanded; this excludes, for instance, the possibility of monopsony power. Input price variation only arises if firms (i) face exogenous price variation, for instance because of geographical price differences, or (ii) input quality variation.

[11] Valmari (2023) estimates production functions for multiproduct firms and accounts for price variation in product markets in monopolistic competition. Orr (2022) combines production function estimation with a demand system to identify unobserved input allocation and total factor productivity of multiproduct firms.

[12] Like Valmari (2023) and Orr (2022), De Loecker et al. (2016) control for the input allocation across products of multi-product firms. They employ data on single-product firms with a sample selection correction. This approach relies on the assumption that single-product firms use the same production technology compared to multi-product firms.

[13] As mentioned by De Loecker et al. (2016), complementarity in input quality is an assumption particularly common in the literature of O-ring type theories of production, see Kremer (1993) or Verhoogen (2008).

The framework in De Loecker et al. (2016) allows for vertical and horizontal product differentiation, where horizontal differentiation is assumed to be costless. The authors employ product dummies instead of product characteristics. This implies the assumption that product characteristics do not change over time.

5 Product Characteristics as Quality Controls

For industries where products are vertically and horizontally differentiated and product characteristics vary over time, product category dummies might not fully control for the quality bias in the product function. Going back to the car industry, including product dummies to control for quality variation in an estimation procedure spanning the years 2005-2024 requires that SUVs manufactured in the early two thousands are of the same quality as SUVs today. The car industry, however, experienced fundamental advancements in technology, ranging from navigation systems and automated parking to increased fuel efficiency and varying design preferences.

In this section, I first provide the baseline Leontief production framework relying on homogenous input and output quantities. I then describe the notion of including product characteristics to the production function for differentiated products industries.

5.1 Production Function Specification

5.1.1 Leontief Technology - Baseline

I assume that the car manufacturers produce with a Leontief technology:

$$Q_{it} = min\left\{\kappa_{it}M_{it}, \Omega_{it}F\left(L_{it}, K_{it}; \beta\right)\right\} exp(\varepsilon_{it}), \tag{11}$$

where the output Q_{it}, material input M_{it}, number of employees L_{it}, and capital K_{it} are measured in comparable units. The relationship between labor and capital substitutability is parameterized by the function $F(;\beta)$, while material inputs are considered to be perfect complements to the combination of labor and capital. This implies that the production location can produce the same quantity of cars with more production workers and less machinery (manual assembly) or less production workers and more machinery (automated assembly). The production locations, however, always require a fixed set of car parts to produce one car. The term κ_{it}^{M} represents the inverse of the required per-unit material inputs, which is specific to each plant. This factor allows for accommodating technological disparities and variations in product quality among car manufacturing plants. From the researcher's perspective, equation (11) contains two unobservables. The first is the productivity term Ω_{it}, which is observed or predictable by the firms when making input decisions. The second is the term ε_{it}, which represents potential measurement error and could be interpreted as unpredicted shocks to production, such as machine breakdowns.

Because of the Leontief property, the car manufacturer chooses the inputs to satisfy the following equation:

$$Q_{it} = \kappa_{it} M_{it} exp(\varepsilon_{it}) = \Omega_{it} F\left(L_{it}, K_{it}; \beta\right) exp(\varepsilon_{it}). \tag{12}$$

It implies that in any given year t, the production location equalizes the material input quantity with the required set of labor and capital to produce the quantity Q_{it}. The production location therefore acquires exactly the required amount of inputs to produce the desired output quantity. This rules out strategic investment into inventories.

5.1.2 Leontief Technology - Differentiated Products

Even if inputs and outputs were observed in physical quantities, taking equation (11) to the data results in biased estimates if variation in product characteristics is not fully accounted for. In this case, unobserved quality variation in outputs $\hat{q}_{it}$ and inputs $\hat{m}_{it}$ enters the production function, which might be correlated with input demand. For applications where inputs and outputs are observed in expenditures and sales and products are differentiated, equation (12) can be rewritten as the following:

$$\begin{aligned} q_{it} &= f(l_{it}, k_{it}; \beta) + \omega_{it} + a\left(p_{it}, \hat{q}_{it}; \alpha\right) + \varepsilon_{it}, \\ &= log(\kappa_{it}) + m_{it} + h\left(p_{it}, z^M_{it}, \hat{q}_{it}, \hat{m}_{it}; \gamma\right) + \varepsilon_{it}, \end{aligned} \tag{13}$$

where lowercase letters indicate logarithmic transformation. The functions $a(., \alpha)$ and $h(;\gamma)$ contain the additional variation that is introduced to the production function because of output price variation p_{it}, output quality variation $\hat{q}_{it}$, material input price variation z^M_{it} and material input quality variation $\hat{m}_{it}$. To allow for flexible specification, I define the functions $a(., \alpha)$ and $h(;\gamma)$ non-parametrically, following the tradition of Katayama, Lu, and Tybout (2009) and De Loecker et al. (2016).[14] Other applications such as Hallak and Sivadasan (2013) provide a parametric specification on the optimal production of quality. I assume that the quality and price of workforce and capital is similar across car manufacturers and instead focus on quality variation of material parts. The framework, however, could be extended to include plant-level quality variation in all inputs. The goal of the estimation procedure is to estimate unbiased production function coefficients β and productivity ω_{it}. Standard regressions of equation (13) lead to biased estimates because both the coefficients in β and unobserved productivity ω_{it} might be correlated with the unobservables.

[14] Katayama, Lu, and Tybout (2009) define quality as a VAR process, conditional on performance determinants. These are, for instance, average wages, lagged R&D expenditures, multinational ownership etc. Similarly, Melitz and Ottaviano (2000) introduce a quality adjusted productivity index to the estimation procedure.

6 Empirical Approach

In this section, I first describe the estimation procedure including product characteristics as demand-based quality controls and sales-weighted plant-level prices. I employ the control function approach in the tradition of Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves, and Frazer (2015) and combine it with the notion of the quality control function as in De Loecker et al. (2016).

I then describe the procedure that I employ to construct the sales-weighted plant-level prices and characteristics, and the approximation for the plant-level share of material inputs to produce one unit of output κ_{it}.

6.1 Estimation Procedure

I follow the control function approach as in Ackerberg, Caves, and Frazer (2015). In the first stage, it separates unobserved productivity ω_{it} from the measurement error ε_{it} with an input demand specification. In the second stage, the production function coefficients are estimated with a general method of moments procedure, leveraging assumptions about the timing of input demand and the productivity evolution.

First Stage of the Estimation Procedure. To separate unobserved productivity from the measurement error, I employ a control function in the spirit of De Loecker and Scott (2022) that is specific to the Leontief production technology. The authors show that the underlying fixed proportion rule of the Leontief technology allows to control for unobserved productivity without taking a stance on competition on input or product markets. Additionally, this control function is not bound to the functional dependency problem that occurs with technologies that rely on substitutable inputs only, for instance, Cobb-Douglas production technologies.[15] Reformulating equation (13) for ω_{it} yields the following equation:

$$\omega_{it} = log(\kappa_{it}) + m_{it} + h(.;\gamma) - f(.;\beta) - a(.;\alpha). \tag{14}$$

Inserting (14) into (13) yields an expression for the predicted output Φ_{it}, which relies on the plant-specific inverse amount of material inputs required for the construction of cars κ_{it}, material input quantity m_{it}, and the function $h(.)$. This step separates the two unobserved terms ω_{it} and ε_{it}:

[15] I refer the reader to Ackerberg, Caves, and Frazer (2015) and Gandhi, Navarro, and Rivers (2020) for a detailed discussion of this issue.

$$
\begin{aligned}
q_{it} &= \Phi_{it} + \varepsilon_{it}, \\
&= f(.;\beta) + a(.;\boldsymbol{\alpha}) + log(\kappa_{it}) + m_{it} + h(.;\gamma) - f(.;\beta) - a(.;\boldsymbol{\alpha}) + \varepsilon_{it}, \qquad (15) \\
&= log(\kappa_{it}) + m_{it} + h(.;\gamma) + \varepsilon_{it}.
\end{aligned}
$$

To control for $h(.;\gamma)$, I construct a quality control function in the spirit of De Loecker et al. (2016):

$$z_{it}^{M} = z_{t}^{M}\left(\rho_{it}, \chi_{it}, G_i, Y_t\right). \qquad (16)$$

Instead of product-level prices, market shares, product dummies, and geographic dummies as in De Loecker et al. (2016), I employ sales-weighted plant-level prices ρ_{it}, sales-weighted plant-level product characteristics ξ_{it}, country dummies G_i and time dummies Y_t. Including characteristics instead of the product category dummies allows for a higher level of product differentiation and varying product characteristics over time. The notion of adding product characteristics is based on insights from the hedonic pricing literature (e.g., Rosen, 1974; Feenstra and Levinsohn, 1995; Triplett, 1969).

Following the standard control function approach, I specify productivity $\omega_{it}(\beta;\boldsymbol{\alpha})$ based on a vector of parameters β and $\boldsymbol{\alpha}$:

$$\omega_{it}(\beta;\boldsymbol{\alpha}) = \Phi_{it} - f(l_{it}, k_{it}, \beta) - a(.,\boldsymbol{\alpha}). \qquad (17)$$

Second Stage of the Estimation Procedure. The second stage of the estimation procedure requires specifying the law of motion of productivity and timing assumptions on input choices. I assume that the law of motion of productivity is represented by a non-parametric function $g(.)$ that depends on lagged productivity ω_{it-1} and a dummy variable for lagged acquisitions acq_{it-1}. As shown by previous literature, for instance, Braguinsky et al. (2015), changes in ownership might affect the production locations' productivity. This might be driven by several factors, for instance, changes in managerial practices.

$$\omega_{it} = g_t\left(\omega_{it-1}, acq_{it-1}\right) + \xi_{it}. \qquad (18)$$

The term ξ_{it} represents the so-called innovation in the productivity process. This idiosyncratic innovation combined with timing assumptions about input choices serves for the construction of the moment conditions. I assume that capital is chosen at time t and labor is subsequently chosen at time $t+1$. This renders the capital demand fixed and dynamic; capital is chosen in the previous period and cannot be adjusted at time t. It thus has dynamic implications for the production process. Following the same notion, labor demand is rendered variable and static. Based on these assumptions, I employ the following moment conditions for the identification of the production function parameters in $f(.)$:

$$E\left(\xi_{it}(\beta;\alpha)\begin{pmatrix} l_{it-1} \\ k_{it} \end{pmatrix}\right) = 0. \tag{19}$$

6.2 Constructing Plant-Level Prices and Characteristics

Constructing Plant-Level Prices. Based on the available data, I face two challenges when constructing plant-level prices. These are (i) unobserved destination countries of manufactured cars, and (ii) unobserved relative quantities of car models in the product mix of car manufacturers.

I solve the first challenge by assuming that each production location sells its cars mainly to European countries. Volkswagen for instance produces the same models for the Asian market in plants that are geographically closer to Asia than European plants. Additionally, there might be unobserved production locations within Europe that produce the same models as the production locations included in the dataset. For this reason, I allocate the seven countries for which I observe price data to the respective production locations, where I assume that the closest production location produces the model for the country. This procedure requires not only searching for the assembled cars of the production locations available in the dataset, but also determining whether there exist other unobserved plants in Europe that produces the same model. To provide an example, Volkswagen Slovakia and Martorell near Barcelona manufacture the model Seat Ibiza at the same time, but Volkswagen only reports individual balance sheet data for the Slovakian plant. Because there are two production locations within Europe that manufacture the same car, I split the countries such that prices from Spain and France do not feed into plant-level prices and characteristics of the Slovakian plant, because the plant near Barcelona is closer to these two countries. Similarly, some of the observed plants in the dataset produce the same models at the same time. The Peugeot plants in Sochaux and Mulhouse in France both produce the Peugeot 307, but are located closely to each other. From the Sochaux website, I extract the information that the 307 model was the plant's topseller in the first decade of the 21st century. The topsellers at Mulhouse are the models 205 and 106. Using this information, I allocate the 307 model to the Sochaux plant for all countries.

Moreover, the available data does not contain price and characteristics data for all models produced at the production locations. Particularly prices and characteristics of high-end sportscars cannot be accounted for when constructing the plant-level variables. This constitutes only a small share of produced cars, because the dataset does not contain any production locations with focus on high-end sportscars. I solve the second challenge of unobserved quantities by weighting plant-level prices and characteristics with sales of the produced models in the respective countries. I assume that relative sales of the models contains information on the relative number of models produced at the respective production location.

Given these data challenges, it is important to mention that the constructed plant-level prices and characteristics provide an approximation for only the true values. Nevertheless, several plants focus on production of only one or two car segments.

Each plant i's output price $\hat{P}_{it}$ depends on the individual prices P_{jt} of the car models j that the plant produces in a given year t. The car prices P_{jt} are weighted by each car model's share s_{ijt} in the plant's total sales S_{it}, which is the sum of sales of the individual cars S_{ijt}, such that $s_{ijt} = \frac{S_{ijt}}{S_{it}}$ with $S_{it} = \sum_J S_{ijt}$.

I employ $\hat{P}_{it}$ to construct output units $\hat{Q}_{it}$ as:

$$\hat{Q}_{it} = \frac{R_{it}}{\hat{P}_{it}}, \tag{20}$$

with R_{it} representing the sum of the plants' revenues from all cars that the plants produce in a given year t, such that $R_{it} = \sum_j R_{ijt}$. As pointed out in De Loecker et al. (2016), in a purely vertically differentiated industry, there exists a one-to-one mapping between prices and product quality. In this case, price variation might fully capture quality variation of products, such that $\hat{Q}_{it} = Q_{it}$.

Constructing Sales-Weighted Plant-Level Characteristics As discussed in Berry (1994), the quality of car models can be represented by a function of observed and unobserved product characteristics. I introduce observed product characteristics X_{jt} to the production function and assume that the unobserved characteristics do not determine the firms' input demand. A possible way to add unobserved characteristics to the procedure is to combine the approach with demand estimation following Berry, Levinsohn, and Pakes (1995).[16]

I construct sales-weighted plant-level product characteristics as:

$$\hat{X}_{it} = \sum_j s_{ijt} X_j, \tag{21}$$

where X_j summarizes the observed product characteristics X of the car models j that are assembled in production plant i at time t. These are for instance horsepower, length, width, or cylinder. Similar to the construction of plant-level prices, I weight the characteristics by each car model's share of total sales s_{ijt}.

[16] In the authors' notation, unobserved characteristics are ξ_j.

6.2.1 Approximating The Material Input Share

The fixed share of material inputs to produce one unit of output (κ_{it}) is unobserved. In practice, κ_{it} represents the quantity of car components that the manufacturers require to produce one car. This number strongly varies within and between car manufacturers, depending on (i) the car models that they produce, and on (ii) the level of aggregation that manufacturers buy their inputs in; whether they rather buy whole modules, components, or individual parts.[17]

De Loecker and Scott (2022) approximate κ_{it} non-parametrically with capital, labor, materials, firm-level wages, year dummies, regional dummies, and their interactions. Similarly, I approximate κ_{it} with

$$\kappa_{it} = h\left(\chi_{it}, l_{it}, w_{it}, m_{it}, k_{it}, D_t\right). \tag{22}$$

The specification if κ_{it} is driven by the Leontief property of the production function.

7 Data

The empirical application relies on a subset of the database of the European car industry from Hahn (2024). I refer to this paper for a detailed description and summary statistics of the dataset.

The employed subset of the data comprises two primary sources: (1) the Orbis balance sheet database provided by Bureau Van Dijk, and (2) the JATO database that contains product characteristics and prices at the country-level.

Financial Statements. I collect financial statements for 31 European car manufacturing plants from 2002 to 2019. The plants belong to ten car manufacturing groups, where three plants are independently owned and assemble cars for various car manufacturers; such as the Magna Steyr production plant in Austria. Within the observed time period, six of the plants change ownership.[18] The data on financial statements only includes production plants that report unconsolidated balance sheet information, meaning that the individual plants provide information on their annual sales, number of employees, etc. Many European plants, specifically in Germany, solely provide consolidated accounts that include subsidiaries. In this case, the reported number of employees for instance does not only include the specific production location, but also numerous other plants that might not operate within the same industry. To provide an example, the Volkswagen production plant in Wolfsburg, Germany, solely reports consolidated financial statements for this plant and hundreds of subsidiaries. These are scattered all over

[17] For other industries where all plants require the same share of material inputs with $\kappa_{it} = \kappa_t$, fixed effects already fully control for κ_t.

[18] Three GM plants are bought by PSA in 2017, one Ford plant is acquired by Geely in 2010, and a jointly owned plant by Fiat and PSA is fully acquired by PSA in 2012.

the world and partly operate in other industries, for instance Volkswagen insurance or banking. Estimating the output elasticities and productivity relying on these consolidated accounts yields a composite weighted measure of all these plants, which challenges interpretation of the results.

Product Characteristics and Prices. The JATO database comprises information on car characteristics and list prices within seven European countries, including Belgium, France, Germany, Great Britain, Italy, Netherlands and Spain for the period 1998 to 2018.

Table 1 presents summary statistics for the financial statements, product characteristics, and prices of car manufacturers from 2002 to 2018. All monetary values are shown in thousands of Euros. Materials represents material input expenditures. The data reveals a substantial spread in the financial statistics of European car manufacturers, particularly within the 25th to 75th percentile range. For example, the number of employees varies from approximately 2,300 to 7,700. The disparity in wages is even more pronounced; the first quartile of employees earns about 60,000 Euros, while the last quartile earns 329,000 Euros, a more than fifty fold increase.

The average selling price of the cars is 22,000 Euros. These cars typically feature 82 horsepower, a cylinder capacity of 1,592 cc, and average dimensions of 175 cm in width, 418 cm in length, and 151 cm in height. On average, they require five liters of gasoline per 100 kilometers.

The "quantity" column represents the constructed output quantity that I use as the dependent variable in the production function estimation. I construct the output quantity by dividing plant-level sales ($P_{it}Q_{it}$) by the constructed plant-level prices (P_{it}). On average, the production locations assemble 142,000 cars annually.

	Capital	Revenue	Materials	Wages	Price	Employees	Horsepower	Cylinder	Width	Length	Height	Liter	Quantity
median	351	1926	1396	117	21	3500	84	1575	175	425	149	5	109167
mean	829	4026	3511	298	22	6568	82	1592	175	418	151	5	141824
p25	197	669	197	60	14	2250	60	1306	169	396	146	5	58027
p75	628	3357	2897	235	29	5546	97	1804	181	449	153	6	186122
p90	1240	5178	4629	329	33	7747	114	2097	183	473	159	7	316730

Notes: Monetary values in thousands.

Table 1: Summary Statistics - Financial Statements, Prices, and Characteristics

Table 2 shows the correlations between the variables presented in Table 1. The financial metrics capital, revenue, material input expenditure, wages, and the number of employees exhibit strong positive correlations, ranging around 0.73. All financial information is also positively correlated with the output quantity, ranging around 0.25. The output quantity, however, is significantly negatively correlated with all product characteristics, for instance, quantity and horsepower (-0.298) or height (-0.175). This suggests that cars with exceptionally high horsepower, such as sports cars, or particularly large cars, tend to sell less.

Table 2 also sheds light on the correlations among product characteristics and prices. Due to the small sample size and potentially high multicollinearity among characteristics, selecting which product characteristics to include into the production function is critical. All characteristics are highly correlated with prices. For instance, the correlation between cylinder capacity and horsepower is 0.853, and between length and width is 0.887. Therefore, only a selected set of characteristics is included in the estimation procedure.

	Capital	Revenue	Materials	Wages	Price	Employees	Horsepower	Cylinder	Width	Length	Height	Liter	Quantity
Capital	1												
Revenue	0.738***	1											
Materials	0.736***	0.998***	1										
Wages	0.734***	0.986***	0.986***	1									
Price	0.0146	-0.0618	-0.0628	-0.0363	1								
Employees	0.726***	0.980***	0.980***	0.990***	-0.0748	1							
Horsepower	0.00867	-0.0874	-0.0871	-0.0811	0.961***	-0.116*	1						
Cylinder	-0.0303	-0.0550	-0.0453	-0.0322	0.838***	-0.0591	0.853***	1					
Width	-0.0231	-0.0513	-0.0592	-0.0452	0.807***	-0.0831	0.840***	0.766***	1				
Length	-0.0537	-0.0694	-0.0660	-0.0396	0.708***	-0.0799	0.737***	0.755***	0.887***	1			
Height	-0.0562	0.0109	0.00960	-0.0118	0.310***	-0.0109	0.331***	0.404***	0.266***	0.0739	1		
Liter	-0.0886	-0.0881	-0.0934	-0.0759	0.359***	-0.0756	0.395***	0.666***	0.312***	0.357***	0.420***	1	
Quantity	0.239***	0.276***	0.285***	0.253***	-0.277***	0.251***	-0.298***	-0.208***	-0.193***	-0.126**	-0.175***	-0.145**	1

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Correlations - Financial Statements, Prices, and Characteristics

8 Results

Table 3 presents the production function estimates, excluding and including various product characteristics. Due to the small sample size, it is not feasible to control for all car characteristics in the $\alpha(.)$-function of the estimation procedure. Therefore, I include only the cylinder variable to control for quality variation in the first stage, as it accounts for the largest share of output quantity variation once labor and capital are controlled for (see Table 4 from Hahn (2024) presented in Appendix B). Given that many car characteristics are highly correlated, including a single characteristic can account for a substantial variation in car quality across several dimensions. As mentioned in the data section, the cylinder variable is highly correlated with other characteristics such as horsepower, length, and car prices.

The production function specification (1) represents the standard specification that is frequently taken to the data, where the dependent variable represents the firms' revenues and characteristics are not included into the procedure. The production function coefficient on labor is relatively high (1.240) and the capital coefficient is insignificant, most likely due to small sample size and insufficient variation in the variable. The specification (2) has the constructed output quantity as dependent variable, which is sales divided by the constructed sales-weighted plant-level prices. With this specification, the labor coefficient decreases to 0.981. Specifications (3)-(9) include characteristics in the first and second stages of the estimation procedure. Including cylinder to procedure as in specification (3) reduces the coefficient of labor further (0.976). Recall that the included characteristics are sales-weighted plant-level product characteristics of the cars, thus including characteristics to the first stage controls for variation in quality and remaining price variation that might not be captured with the constructed output variable. Introducing characteristics to the first stage of the procedure allows for separating unobserved productivity from the measurement error, accounting for product quality variation within and between production locations. For specifications (4)-(9), I subsequently add more characteristics to first stage of the procedure. It shows that the coefficient of labor further decreases and capital stays insignificant. Once the characteristic "width" is added to the procedure, adding additional characteristics (length) does not affect the estimated production functions coefficients anymore, accounting for all endogenous and unobserved quality variation.

	(1) revenues	(2) q	(3) q	(4) q	(5) q	(6) q	(7) q	(8) q	(9) q
labor	1.240***	0.981***	0.976***	0.956***	0.976***	0.964***	0.954**	0.882**	0.882**
	(0.295)	(0.239)	(0.230)	(0.268)	(0.286)	(0.270)	(0.296)	(0.271)	(0.284)
capital	0.0181	0.770	-0.148	-0.108	-0.129	-0.124	-0.122	0.0406	0.0406
	(0.688)	(0.554)	(0.176)	(0.170)	(0.170)	(0.157)	(0.172)	(0.177)	(0.194)
cylinder			-1.886***	-1.925	-1.931	-1.910	-1.909	-2.074	-2.074
			(0.295)	(1.069)	(1.419)	(1.432)	(1.337)	(1.504)	(1.536)
a(.)									
cylinder			-1.886	-1.925	-1.931	-1.910	-1.909	-2.074	-2.074
			(1.280)	(30.01)	(15.17)	(13.66)	(16.11)	(13.31)	(20.00)
Characteristics in the First Stage:									
cylinder				✓	✓	✓	✓	✓	✓
horsepower					✓	✓	✓	✓	✓
height						✓	✓	✓	✓
liter							✓	✓	✓
width								✓	✓
length									✓

Block-bootstrapped standard errors (at the plant-level) in parentheses. 1.000 replications

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Production Function Estimates

9 Discussion

Unbiased production function estimation relies on several critical factors, ranging from data availability and product differentiation within the assessed industry, to competitive environments in input and product markets.

Depending on the extent of product differentiation and the competitive environments, including price indices or plant-level prices might fully control for price and quality variation in production function estimation. The more complex competitive environments and the more differentiated products become, however, the more flexible the quality control function needs to be in terms of demand and conduct assumptions.

Various approaches have been developed to control for quality and price variation in industries with vertically differentiated products. These range from estimating sales-generating production functions with input expenditures, over constructing plant-level input and output prices, to assumptions on pass-through of input and output price variation. In cases where products are highly differentiated and output is measured in sales, controlling for quality could be conducted with the establishment of demand systems or expressing price variation based on observables.

For industries with highly differentiated products, where product attributes change over time, including product characteristics to the estimation procedure controls for quality variation driven by technological changes in environments of imperfect competition. The car industry provides a good example of this because inputs and output quality varies between car models, which are highly differentiated and advance technologically over time. Estimating a production function spanning several years including characteristics allows for varying product traits over time. For industries that experience frequent technological advancements, it becomes relevant to account for the quality bias in the production function.

10 Conclusion

The methodology of production function estimation became a frequently discussed topic in the industrial organization literature and has been applied to many industries by recent work. Unbiased estimation of production functions requires adequate controls for product quality and potential price variation within and between observation units.

In this paper, I provide an overview of state-of-the-art approaches that have been developed to control for these biases in production function estimation. Further, I discuss an alternative approach including product characteristics that controls for quality variation in highly differentiated product industries. This is relevant for products such as cars or high-end electronic devices that undergo regular changes because of technological advancements or adjusting consumer preferences. Based on a dataset of European car manufacturers, I demonstrate that including characteristics helps to mitigate latent quality variation that is correlated with input demand; leading to a decrease in the estimated production function coefficients.

As linking product characteristics to production locations becomes more feasible for different industries, a trajectory for future research might be to include characteristics to production functions of other sectors. Constructing comparable input and output units by including hedonic prices and hedonic quantity controls might offer an alternative approach to account for quality variation in highly differentiated product industries.

A Pass-through and Revenue Generating Production Functions

$$s_{it} = e_{it}\alpha + \omega_{it} + p_{it} - z_{it}\alpha + \varepsilon_{it}. \tag{23}$$

The production function coefficients are identified if the output price bias fully offsets the input price bias with $p_{it} - z_{it}\alpha = 0$. As discussed by the authors, this situation arises when additional assumptions hold, that is (1) monopolistic competition on product markets, (2) firms produce horizontally differentiated products, facing the same constant elasticity of substitution (CES) demand system, (3) input neutral input price variation, with $z_{it}^{h} = \lambda_{it} \forall h = 1, 2...., H$. H represents the total number of inputs, and (4) constant returns to scale (CRS).

Assumptions (1) and (2) are required to assure that variation in p_{it} is fully reflected in variation in z_{it}, such that $\Delta p_{it} = \Delta z_{it}$. Monopsony power in the product market, for instance, does not allow firms to flexibly adjust p_{it} according to changes in z_{it}. Uni-dimensional product differentiation is required because vertically and horizontally differentiated products do not necessarily allow for a one-to-one mapping of input to output quality and prices. Assumptions (3) and (4) are required to ensure that both biases offset each other, such that $p_{it}^{*} = z_{it}^{*}\alpha$, with $\alpha = 1$. In cases where $\alpha \neq 1$, the production function still contains an error term that might be correlated with unobserved productivity ω_{it}. To illustrate this argument, I employ the example by De Loecker and Goldberg (2014) and assume a simple Cobb-Douglas production function that consists of two substitutable inputs: labor l_{it} and capital k_{it}. Labor expenditure is represented by wages W_{it} multiplied by the number of employees L_{it} and capital expenditure is represented by the returns to capital r_{it} multiplied by the capital stock K_{it}. In this case, the structural error term $\hat{\varepsilon}_{it}$ contains unobserved price variation in w_{it}, k_{it} and outputs p_{it}. Let us assume that similarly to output prices, wages and returns to capital are deflated with industry-level deflators $\bar{w}_t$ and $\bar{r}_t$, such that $w_{it}^{*} = w_{it} - \bar{w}_t$ and $r_{it}^{*} = r_{it} - \bar{r}_t$. In this case, the structural error term is represented by the following equation:

$$\hat{\varepsilon}_{it} = \varepsilon_{it} + p_{it}^{*} - \alpha_w w_{it}^{*} - \alpha_k r_{it}^{*} \tag{24}$$

$$= \varepsilon_{it} + p_{it}^{*} - \alpha_w \lambda_{it} - \alpha_k \lambda_{it} \tag{25}$$

$$= \varepsilon_{it} + p_{it}^{*} - (\alpha_w + \alpha_k)\lambda_{it} \tag{26}$$

Combining equation (24) with assumption (4) on constant returns to scale, which implies $\alpha_w + \alpha_k = 1$ and assumption (1) and (2) leads to both input and output price biases cancel out each other, such that the realized measurement error represents the i.i.d measurement error $\hat{\varepsilon}_{it} = \varepsilon_{it}$.

B Characteristics Selection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	q	q	q	q	q	q	q	q
f(.)								
l	0.844**	0.888***	0.872**	0.870***	0.848**	0.847**	0.847**	0.837**
	(0.247)	(0.227)	(0.234)	(0.230)	(0.230)	(0.229)	(0.231)	(0.234)
k	0.0711	0.0675	0.0268	0.0253	0.0112	0.0136	0.0145	0.0149
	(0.257)	(0.211)	(0.220)	(0.219)	(0.215)	(0.210)	(0.215)	(0.216)
a(.)								
price		-0.963*	0.651	0.760	1.579	1.598	1.594	1.563
		(0.405)	(0.883)	(0.834)	(1.214)	(1.205)	(1.201)	(1.176)
cylinder			-3.181*	-3.467*	-2.429*	-2.489*	-2.500*	-2.316*
			(1.257)	(1.261)	(0.974)	(1.035)	(1.041)	(1.031)
height				0.871	0.505	0.573	0.639	0.979
				(1.961)	(1.975)	(2.066)	(1.864)	(1.634)
horsepower					-1.959	-2.018	-1.998	-1.885
					(1.526)	(1.528)	(1.567)	(1.590)
length						0.325	0.335	1.525
						(2.226)	(2.193)	(3.169)
liter							-0.0368	0.0143
							(0.258)	(0.287)
width								-4.872
								(10.68)
Observations	296	296	296	296	296	296	296	296
R^2	0.630	0.693	0.721	0.721	0.729	0.730	0.730	0.731
Adjusted R^2	0.606	0.672	0.700	0.700	0.708	0.707	0.706	0.706

Notes: OLS-regressions. Robust standard errors in parentheses, clustered at the plant-level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: OLS-Regressions

References

Ackerberg, Daniel, Caves, Kevin, and Frazer, Garth (2015). "Identification Properties of Recent Production Function Estimators". *Econometrica* 83.6, pp. 2411–2451.

Avignon, Rémi and Guigue, Etienne (2022). "Markups and Markdowns in the French Dairy Market". *Mimeo*.

Berry, Steven (1994). "Estimating Discrete-Choice Models of Product Differentiation". *RAND Journal of Economics*, pp. 242–262.

Berry, Steven, Levinsohn, James, and Pakes, Ariel (1995). "Automobile Prices in Market Equilibrium". *Econometrica* 63.4, pp. 841–890.

Blundell, Richard and Bond, Stephen (1998). "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models". *Journal of Econometrics* 87.1, pp. 115–143.

Braguinsky, Serguey, Ohyama, Atsushi, Okazaki, Tetsuji, and Syverson, Chad (2015). "Acquisitions, Productivity, and Profitability: Evidence from the Japanese Cotton Spinning Industry". *American Economic Review* 105.7, pp. 2086–2119.

Collard-Wexler, Allan and De Loecker, Jan (2015). "Reallocation and Technology: Evidence from the US Steel Industry". *American Economic Review* 105.1, pp. 131–71.

De Loecker, Jan (2011). "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity". *Econometrica* 79.5, pp. 1407–1451.

De Loecker, Jan (2013). "Detecting Learning by Exporting". *American Economic Journal: Microeconomics* 5.3, pp. 1–21.

De Loecker, Jan and Ackerberg, Daniel (2024). "Production Function Identification Under Imperfect Competition". *Mimeo*.

De Loecker, Jan, Goldberg, Pinelopi, Khandelwal, Amit, and Pavcnik, Nina (2016). "Prices, Markups, and Trade Reform". *Econometrica* 84.2, pp. 445–510.

De Loecker, Jan and Goldberg, Pinelopi Koujianou (2014). "Firm Performance in a Global Market". *Annual Review of Economics* 6.1, pp. 201–227.

De Loecker, Jan and Scott, Paul (2022). "Markup Estimation using Production and Demand Data. An Application to the US Brewing Industry". *Mimeo*.

De Loecker, Jan and Syverson, Chad (2021). "An Industrial Organization Perspective on Productivity". *Handbook of Industrial Organization*. Vol. 4. 1. Elsevier, pp. 141–223.

De Roux, Nicolás, Eslava, Marcela, Franco, Santiago, and Verhoogen, Eric (2021). "Estimating Production Functions in Differentiated-Product Industries with Quantity Information and External Instruments". *IZA Discussion Paper No. 14006*.

Dhyne, Emmanuel, Petrin, Amil, Smeets, Valerie, and Warzynski, Frederic (2020). "Theory for Extending Single-Product Production Function Estimation to Multi-Product Settings". *Aarhus University Working Paper*.

Doraszelski, Ulrich and Jaumandreu, Jordi (2013). "R&D and Productivity: Estimating Endogenous Productivity". *Review of Economic Studies* 80.4, pp. 1338–1383.

Feenstra, Robert and Levinsohn, James (1995). "Estimating Markups and Market Conduct with Multidimensional Product Attributes". *Review of Economic Studies* 62.1, pp. 19–52.

Forlani, Emanuele, Martin, Ralf, Mion, Giordiano, and Muûls, Mirabelle (2016). "Unraveling Firms: Demand, Productivity and Markups Heterogeneity". *CEPR Discussion Paper* DP11058.

Foster, Lucia, Haltiwanger, John, and Syverson, Chad (2008). "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* 98.1, pp. 394–425.

Gandhi, Amit, Navarro, Salvador, and Rivers, David (2020). "On the Identification of Gross Output Production Functions". *Journal of Political Economy* 128.8, pp. 2973–3016.

Gibbon, Alexandra and Schain, Jan Philip (2023). "Rising markups, common ownership, and technological capacities". *International Journal of Industrial Organization* 89, p. 102900.

Gonzales, Xulia, Lach, Saul, and Miles, Daniel (2021). "Production function estimation when (only) price changes are observed". *Mimeo*.

Griliches, Zvi (1964). "Notes on the Measurement of Price and Quality Changes". *Models of Income Determination*. Princeton University Press, pp. 381–418.

Griliches, Zvi (1971). "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change". *Price Indexes and Quality Change: Studies in New Methods of Measurement*. Harvard University Press, pp. 55–87.

Hahn, Nadine (2024). "Who Is in the Driver's Seat? Markups, Markdowns, and Profit Sharing in the Car Industry". *Mimeo*.

Hallak, Juan Carlos and Sivadasan, Jagadeesh (2013). "Product and Process Productivity: Implications for Quality Choice and Conditional Exporter Premia". *Journal of International Economics* 91.1, pp. 53–67.

Hendel, Igal and Spiegel, Yossi (2014). "Small Steps for Workers, a Giant Leap for Productivity". *American Economic Journal: Applied Economics* 6.1, pp. 73–90.

Itoga, Takaaki (2019). "Within-Firm Reallocation and the Impacts of Trade under Factor Market Imperfection". *Mimeo, Penn State University*.

Katayama, Hajime, Lu, Shihua, and Tybout, James (2009). "Firm-Level Productivity Studies: Illusions and a Solution". *International Journal of Industrial Organization* 27.3, pp. 403–413.

Klette, Jakob and Griliches, Zvi (1996). "The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous". *Journal of Applied Econometrics* 11.4, pp. 343–361.

Kremer, Michael (1993). "The O-Ring Theory of Economic Development". *Quarterly Journal of Economics* 108.3, pp. 551–575.

Kugler, Maurice and Verhoogen, Eric (2012). "Prices, Plant Size, and Product Quality". *Review of Economic Studies* 79.1, pp. 307–339.

Levinsohn, James and Petrin, Amil (2003). "Estimating Production Functions using Inputs to Control for Unobservables". *Review of Economic Studies* 70.2, pp. 317–341.

Maican, Florin and Orth, Matilda (2021). "Determinants of Economies of Scope in Retail". *International Journal of Industrial Organization* 75, p. 102710.

Melitz, Marc and Ottaviano, Gianmarco (2000). *Estimating Firm-level Productivity in Differentiated Product Industries*. Tech. rep. Mimeo, Harvard University.

Mertens, Matthias (2023). “Labor Market Power and Between-Firm Wage (In) Equality”. *International Journal of Industrial Organization* 91, p. 103005.

Olley, Steven and Pakes, Ariel (1996). “The Dynamics of Productivity in the Telecommunications Equipment Industry”. *Econometrica* 64.6, pp. 1263–1297.

Orr, Scott (2022). “Within-Firm Productivity Dispersion: Estimates and Implications”. *Journal of Political Economy* 130.11, pp. 2771–2828.

Pozzi, Andrea and Schivardi, Fabiano (2016). “Demand or Productivity: What Determines Firm Growth?” *RAND Journal of Economics* 47.3, pp. 608–630.

Raval, Devesh (2024). “Economies of Scope and Common Inputs in Multi-Output Production”. *Mimeo*.

Rosen, Sherwin (1974). “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition”. *Journal of Political Economy* 82.1, pp. 34–55.

Rubens, Michael (2022). “Management, Productivity, and Technology Choices: Evidence from US Mining Schools”. *RAND Journal of Economics*.

Smeets, Valerie and Warzynski, Frederic (2013). “Estimating Productivity with Multi-Product Firms, Pricing Heterogeneity and the Role of International Trade”. *Journal of International Economics* 90.2, pp. 237–244.

Treuren, Leonard (2022). “Wage Markups and Buyer Power in Intermediate Input Markets”. *KU Leuven Discussion Paper Series No. 22.06*.

Triplett, Jack E (1969). “Automobiles and Hedonic Quality Measurement”. *Journal of Political Economy* 77.3, pp. 408–417.

Valmari, Nelli (2023). “Estimating Production Functions of Multiproduct Firms”. *Review of Economic Studies* 90.6, pp. 3315–3342.

Verhoogen, Eric A (2008). “Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector”. *The Quarterly Journal of Economics* 123.2, pp. 489–530.

Download ZEW Discussion Papers:

https://www.zew.de/en/publications/zew-discussion-papers

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

//
IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de