Non-Separable Joint Cost Functions and Returns to Scope: Identification and Testing

Nicole Scholz

March 2026

Abstract

Understanding economies of scope is central to merger analysis and the regulation of natural monopolies, yet existing tools cannot account for imperfect input markets and within-firm productivity heterogeneity. This paper relaxes both. I derive a corrected testable condition for non-jointness in residual cost functions and extend it to production frontiers. I then demonstrate that ignoring pre-determined inputs causes researchers to underestimate returns to scale while leaving returns to scope consistently identified. Finally, I establish non-parametric point identification of the joint cost function when firm-product-time-specific Hicks-neutral productivities follow an AR(1) process, showing that standard GMM moments recover the cost function without imposing separability or functional form restrictions on the technology.

1 Introduction

Merger simulations are one of the most prominent and important uses of demand estimation in economics. To correctly simulate demand it is necessary to recover the firms' cost functions from its first order conditions, identification of which has been proven among others by Berry and Haile (2014).

However, while demand estimation has advanced significantly since its early days, advancements in its treatment of the production side have lagged behind. Most strikingly, current identification results typically assume that firms are equally productive at producing each of their outputs, which is in conflict with the realities of production (see e.g. Orr, 2022). Additionally, they abstract away from structural wedges in input markets, like input market power, adjustment costs, and pre-determined inputs. Therefore, current estimates are reduced form approximations rather than their true structural counterparts that are necessary to draw robust conclusions about counterfactual scenarios like mergers.

Furthermore, it is key to understand economies of scope both in merger analysis and in the regulation of natural monopolies. However, the conventional wisdom that production of a good is independent of production of another, also called non-joint, if and only if the joint cost function can be written as the sum of individual cost functions (Hall, 1973) no longer holds when some inputs are pre-determined as optimal re-allocation leads to a different type of interdependence (Shumway et al., 1984). While we currently have a good understanding of the functional form restrictions of structurally independent production processes when the products are produced by technologies with common input intensities, large gaps remain for more flexible functional forms. Since most economic analysis models production using Cobb Douglas or translog specifications it is imperative to understand how these can be accommodated in the multi-product setting.

In this paper I develop new tools for economists and regulators interested in the cost function and returns to scope, I exemplify the importance of combining demand and production data when input markets may not be perfectly competitive. From this motivation, I show how to adjust the non-jointness condition when firms have pre-determined inputs and extend the existing non-parametric identification results for the joint cost function to allow for within-firm productivity differences and fixed inputs.

I start by describing the current state of cost estimation from the demand side and the recent advances in the estimation of the production side in section 2. Section 3 lays out the formal production model and basic definitions. Section 4 exemplifies the issues with ignoring pre-determined inputs and adjustment costs using Khmelnitskaya et al. (2025)'s approach to the estimation of returns to scale and scope. Therefore, in section 5, I generalise Hall (1973)'s test of non-jointness to production frontiers and cost functions with pre-determined inputs. Having developed these tools, I generalise the existing non-parametric identification results for cost functions to allow for within-firm productivity differences and pre-determined inputs in section 6. Section 7 concludes.

2 Literature Review

2.1 Demand estimation

There is already a large and well-known literature identifying the marginal cost function from demand estimation. These results are crucial for the validity of counterfactual exercises like merger simulations.

While the identification of marginal costs follows from identification of the demand system, the identification of the entire marginal cost function requires additional assumptions. The literature abstracts from input market imperfections, instead assuming that firms can buy inputs freely on perfectly competitive markets. Berry and Haile (2014) also make the assumption that the marginal cost of a good does not depend on the productivities (unobserved cost shocks) of all the other products. Under the assumption that a subset of instruments enters the marginal cost function linears, they non-parametrically identify the marginal cost function.

Khmelnitskaya et al. (2025) instead model production using a constant elasticity of substitution (CES) production frontier while allowing for firm-product specific productivities. Their model allows for direct recovery of returns to scope and scale parameters using an instrumental variable regression without needing to identify the cost function.

2.2 Production estimation

Khmelnitskaya et al. (2025)'s model is rooted in recent advances in the estimation of production technologies. This literature has received renewed interest from academics and competition authorities studying firms' input market power (e.g. Autor et al., 2020; CMA, 2024). The renewed interest has also derived from increasing availability of product level production datasets.

However, product level datasets have also raised new challenges as most datasets only report inputs aggregated to the firm level, rather than at the product level. This has led to the development of a wide variety of approaches, including methods to impute input allocations and estimate production functions (e.g. De Loecker et al., 2016; Orr, 2022; Valmari, 2023).

A separate strand of the literature has instead turned to estimating the production possibility frontier to side-step the issue of unobserved input allocations. However, the estimation of production frontiers is not as innocuous as it may first appear. As shown by Cairncross et al. (2023), the commonly used, separable, specifications of the production frontier impose important limitations on the underlying technologies. In particular, the technology is generally forced to be joint – the only exception being the CES production frontier. However, as noted by Hall (1973), non-joint technologies can only be represented by a separable frontier if relative input intensities are the same across products. That is, non-jointness and separability together imply that the product specific production functions differ only by a Hicks-neutral term.

These insights mean that new methods need to be developed to tackle industries where different products are known to require different input mixes – Cairncross et al. (2023) give capital intensive electric vehicles vs labour intensive combustion vehicles as an example. However, to generalise methods, a number of hurdles need to be overcome. Firstly, models in modern IO typically assume a mix of input and output market power and the existence of inputs that are fixed in the short run, like capital, and inputs that may have short-run adjustment costs, like labour. This is in contrast to the existing theory and methods to distinguish joint and non-joint production like those developed in Hall (1973), which rely on all inputs and outputs being freely traded in perfectly competitive markets.

In particular, the existence of fixed inputs induces jointness in production in the short run even when the technology is non-joint, as cost minimisation leads to a re-allocation of fixed inputs when a firm produces more of one input (Shumway et al., 1984). This means that testing for non-jointness is not as easy as testing whether the cross-output terms are zero as has been done for example by Warzynski (2022WP). Importantly, even with a simple model of non-joint production with Cobb Douglas production functions, the production frontier does not generally have a closed form solution (Jensen, 2025) when the production functions have different parameters. Therefore, even in the most basic (commonly used) data generating processes, it may be necessary to non-parametrically estimate the frontier.

This paper contributes to the multi-product production estimation literatures by providing a non-parametric identification result for the joint cost function and deriving non-jointness conditions when cost functions appear joint due to fixed inputs. In the next section I will introduce the standard production model and define key terms.

3 Model

I will assume that there are J firms each producing N goods over T periods. Let $\vec{Y}_{i,j,t} \geq 0$ denote the quantity of good i produced by firm j at time t. Where it is clear, I will suppress the subscripts to reduce notational clutter.

I will assume that each firm produces a vector of outputs, $\vec{Y}$, given some technology, which will be modelled by its transformation function, $T(\vec{Y}, \vec{X})$. The transformation function describes by how much the output vector can be scaled while still being able to produce it with the given vector of inputs.[1] Therefore, the firm's production is pareto efficient if $T(\vec{Y}, \vec{X}) = 1$. The set of all $(\vec{Y}, \vec{X})$ such that $T(\vec{Y}, \vec{X}) = 1$ is called the *production possibility frontier*.

I will further assume that each firm has a vector of firm-product-time-specific productivities, $\vec{A}$, allowing both for between- and within-firm productivity differences. As

[1]For this reason, the transformation function is sometimes also called the distance function.

is standard, I will assume that productivity is Hicks-neutral so that

$$T(\vec{Y}, \vec{X}, \vec{A}) \coloneqq T(\vec{Y} \oslash \vec{A}, \vec{X}), \tag{1}$$

where $\oslash$ denotes element-wise division.

Whilst the transformation function models how firms turn inputs into outputs, the cost function describes how costly it is to produce a set of outputs. It is simply the solution to the firm's cost minimisation problem.

$$C(\vec{Y}, \vec{W}, \vec{A}) \coloneqq \min_{\vec{X}} \vec{W}^\top \vec{X} \text{ s.t. } T(\vec{Y}, \vec{X}, \vec{A}) \leq 1, \tag{2}$$

where $\vec{W}$ is the vector of input prices. As is customary, I will denote the vector of log productivities by $\vec{\omega}$.

Lastly, it is sometimes possible to represent the production technology using product-specific production functions,

$$\frac{Y_i}{A_i} = F_i(\vec{X}_i),$$

that describe the relationship between outputs and inputs separately.

A technology is *non-joint* if the firm could be split into one firm for each product without affecting production. More formally, a production technology is non-joint if in addition to the product-specific production functions, there exist input allocations, $\vec{X}_i \geq \vec{0}$, such that inputs are allocated exclusively and exhaustively

$$\sum_i \vec{X}_i = \vec{X}.$$

If a technology features returns to scope in marginal costs it is also non-joint.[2]

[2] It is worth noting here that the production function literature has largely assumed away fixed costs of production

4 Returns to Scale and Scope

4.1 The CES cost function

Khmelnitskaya et al. (2025) study the estimation of returns to scale and scope from a structural perspective. As they allow for within-firm productivity heterogeneity, but maintain the assumption that all inputs can be freely bought on perfectly competitive markets, their paper is a good example for the necessary caution economists should maintain when interpreting counterfactual simulations based purely on demand estimates.

They assume that firms produce goods using a production function featuring non-rival inputs like managerial time that can simultaneously be used in all product lines at the same time, based on Chapter 15 of Baumol et al. (1982). Let X^r and X^p denote the amount of input X assigned to the rival and public task respectively. Then the*production function* takes the form of

$$Y_{i,j,t} = \frac{A_{i,j,t}}{B} \left(\prod_X \left(X^r_{i,j,t}\right)^{\beta^r_X} \left(X^p_{j,t}\right)^{\beta^p_X} \right), \tag{3}$$

where $B := \frac{\prod_X (\beta^r_X)^{\beta^r_X} (\beta^p_X)^{\beta^p_X}}{\prod_X (\beta^r_X + \beta^p_X)^{\beta^r_X + \beta^p_X}}$ is a constant, $X^r_{i,j,t}, X^p_{j,t} \geq 0$ and $X^p_{j,t} + \sum_i X^r_{i,j,t} = X_{j,t}$.

Letting $\alpha := \sum_X \beta^r_X$ denote the returns to scale in rival tasks and $\phi := \sum_X \beta^r_X + \beta^p_X$ denote the returns to scale in all tasks, Khmelnitskaya et al. (2025) show that these production functions give rise to the following constant elasticity of substitution (CES) transformation function

$$T(\vec{Y}, \vec{X}, \vec{A}) = \frac{\left(\sum_i \left(\frac{Y_{i,j,t}}{A_{i,j,t}} \right)^{\frac{1}{\alpha}} \right)^{\alpha}}{\prod_X (X_{j,t})^{\beta^r_X + \beta^p_X}}. \tag{4}$$

Additionally assuming that all inputs are bought on perfectly competitive markets it is then possible to derive the following cost function

$$C(\vec{Y}, \vec{W}, \vec{\omega}) = g(\vec{W}) \left(\sum_i \left(\frac{Y_i}{\omega_i} \right)^{\frac{1}{\alpha}} \right)^{\frac{\alpha}{\phi}}, \tag{5}$$

where

$$g(\vec{W}) = \phi \prod_X \left(\frac{W_x}{\beta_x} \right)^{\frac{\beta_X}{\phi}} .$$

The model nests non-joint production with a common homogeneous technology when $\phi = \alpha$.

To identify ϕ and α Khmelnitskaya et al. (2025) assume that firms sell their goods in different markets and introduce transportation costs, modelled by product-market-specific iceberg trade costs – that is $\tau_{i,j,c,t} \geq 1$ units of good i must be sent to market c for 1 unit to arrive, so that $Y_{i,j,t} = \sum_c Y^c_{i,j,t} \tau_{i,j,c,t}$ and

$$C(\vec{Y}, \vec{W}, \vec{\omega}) = g(\vec{W}) \left(\sum_i \left(\sum_c \frac{Y_{i,c}\tau_{i,c}}{\omega_i} \right)^{\frac{1}{\alpha}} \right)^{\frac{\alpha}{\phi}} \tag{6}$$

They show that log marginal costs can be written as

$$\ln MC^c_{i,t} = \frac{\phi - \alpha}{\phi} \ln S_{i,j,t} + \frac{1-\phi}{\phi} \ln \left(\frac{Y^c_{i,j,t}}{S^c_{i,j,t}} \right) + \underbrace{\ln \left(\frac{1}{\phi} g(\vec{W}) \right) - \frac{1}{\phi} \ln \left(\frac{\omega_{i,j,t}}{\tau_{i,j,c,t}} \right)}_{:=\tilde{\omega}_{i,j,c,t}}, \tag{7}$$

where

$$S_{i,j,t} = \frac{\sum_c MC^c_{i,j,t} Y^c_{i,j,t}}{\sum_i \sum_c MC^c_{i,j,t}} = \frac{\left(\sum_c \frac{Y^c_{i,j,t}\tau_{i,j,c,t}}{\omega_i} \right)^{\frac{1}{\alpha}}}{\sum_{i'} \left(\sum_c \frac{Y^c_{i',j,t}\tau_{i',j,c,t}}{\omega_{i'}} \right)^{\frac{1}{\alpha}}}$$

is the share of input costs of product i produced by firm j relative to the input costs of all products produced by the firm and

$$S^c_{i,j,t} = \frac{MC^c_{i,j,t} Y^c_{i,j,t}}{\sum_c MC^c_{i,j,t}} = \frac{Y^c_{i,j,t}\tau_{i,j,c,t}}{\sum_c Y^c_{i,j,t}\tau_{i,j,c,t}}$$

is the share product i being produced to be sold in market c. They show that, assuming market entry is mostly driven by demand-side factors, variation in the number of cities a product is shipped to identifies scale economies, while scope economies are identified by exogenous variation that shifts the output of other product lines, e.g. demand shifters.

Now consider the case where capital is actually fixed. Then in appendix A I show that the cost function becomes

$$C(\vec{Y},\vec{W},K,\vec{\omega}) = g^f(\vec{W})K^{-\beta_k/\phi^f}\left(\sum_i\left(\sum_c \frac{Y_{i,c}\tau_{i,c}}{\omega_i}\right)^{\frac{1}{\alpha}}\right)^{\frac{\alpha}{\phi^f}}, \tag{8}$$

where $\phi^f = \phi - \beta_k$ is the short run returns to scale and $g^f(\vec{W}) = \phi^f \prod_{X\neq K}\left(\frac{W_X}{\beta_X}\right)^{\beta_X/\phi^f}$ is the pre-determined input equivalent of $g(\vec{W})$. The (log) marginal costs adjust similarly.

$$\ln MC^c_{i,j,t} = \frac{\phi^f-\alpha}{\phi^f}\ln S_{i,j,t} + \frac{1-\phi^f}{\phi^f}\ln\left(\frac{Y^c_{i,j,t}}{S^c_{i,j,t}}\right) + \underbrace{\ln\left(\frac{1}{\phi^f}g^f(\vec{W})\right) - \frac{\beta_k}{\phi^f}\ln K - \frac{1}{\phi^f}\ln\left(\frac{\omega_{i,j,t}}{\tau_{i,j,c,t}}\right)}_{:=\tilde{\omega}_{i,j,c,t}} \tag{9}$$

Therefore, a researcher estimating the returns to scale and scope without adjusting for fixed inputs will be able to correctly estimate returns to scope, but will underestimate returns to scale.

4.2 Adjustment costs

We have so far derived the cost function when firms purchase capital in a perfectly competitive market and when firms face fixed inputs. A more general class of models is the model with firms buying inputs on perfectly competitive input markets but facing adjustment costs.

For a simple model that naturally bridges the two extreme cases I will use a model with linear adjustment costs. Assume that capital is rented at a competitive rate of W_K.[3] At period $t-1$ the firm chooses how much capital to rent for the next period, K^0_t. After observing its productivity in period t, $\vec{\omega}_t$, the firm can rent additional units or sublet already units at a rate of $W_K + a$ and $W_K - a$ respectively, where the adjustment cost may reflect installation costs or a market with imperfect liquidity. The

[3]It is possible to instead have capital follow an accumulation process with the firm buying the difference between the desired and depreciated capital stock on a competitive market; however, this leaves the depreciated stock as a hidden state variable, making it necessary to have capital data to recover the returns to scale and scope even without adjustment costs.

firm's objective is then

$$\sum_X W_X^\top X + a|K - K^0| \text{ s.t. } T(\vec{Y}, \vec{X}, \vec{\omega}), \tag{10}$$

where $a \in [0, \infty)$ measures the adjustment costs. This model nests perfect competition with $a = 0$ and fixed capital at the limit as $a \to \infty$.

The subjective price faced by the firm then depends on the amount of capital it chooses relative to K^0,

$$\tilde{W}_K = \begin{cases} W_K - a & \text{if } K < K^0 \\ W_K & \text{if } K = K^0 \\ W_K + a & \text{if } K > K^0. \end{cases} \tag{11}$$

Let $\mathcal{Y} = \left(\sum_i \left(\frac{Y_i}{A_i}\right)^{\frac{1}{\alpha}}\right)^\alpha$ be the output index. Cost minimisation allows us to solve for capital

$$K = \frac{\beta_K^{(1-\beta_K)/\phi} \mathcal{Y}^{1/\phi} \prod_{X \neq K} \left(\frac{W_X}{\beta_X}\right)^{\beta_X/\phi}}{\tilde{W}_K^{(1-\beta_K)/\phi}}. \tag{12}$$

Substituting in K^0 for K provides cutoff values for $\mathcal{Y}$ such that $K > K^0$ and $K < K^0$ respectively:

$$\mathcal{Y}^+ = \frac{(W_K + a)^{(1-\beta_K)}}{\beta_K^{(1-\beta_K)} (K^0)^\phi \prod_{X \neq K} \left(\frac{W_X}{\beta_X}\right)^{\beta_X}}, \; \mathcal{Y}^- = \frac{(W_K - a)^{(1-\beta_K)}}{\beta_K^{(1-\beta_K)} (K^0)^\phi \prod_{X \neq K} \left(\frac{W_X}{\beta_X}\right)^{\beta_X}}. \tag{13}$$

The cost function is then characterised by three regimes, depending on value of the output index relative to its cutoffs.

$$C(\vec{Y}, \vec{W}, K^0, a, \vec{\omega}) = \begin{cases} C(\vec{Y}, \vec{W} - a \cdot 1_K, \vec{\omega}) + aK^0 & \text{if } \mathcal{Y}(\vec{Y}) < \mathcal{Y}^- \\ C(\vec{Y}, \vec{W}, K^0, \vec{\omega}) + W_K K^0 & \text{if } \mathcal{Y}(\vec{Y}) \in [\mathcal{Y}^-, \mathcal{Y}^+] \\ C(\vec{Y}, \vec{W} + a \cdot 1_K, \vec{\omega}) - aK^0) & \text{if } \mathcal{Y}(\vec{Y}) > \mathcal{Y}^+, \end{cases} \tag{14}$$

noting that K^0 equals the observed K if $\mathcal{Y} \in [\mathcal{Y}^-, \mathcal{Y}^+]$.

There are two important take-aways to note. Firstly, the range of output values for which the cost function behaves equivalently to the case with fixed capital increases with the severity of the adjustment costs, a. Secondly, the subjective input price now becomes an endogenous variable of $\vec{Y}$.

The two exceptions to this endogeneity are exactly the cases of perfect competition and the limit case of fixed capital inputs. This has important consequences for the recovery of returns to scale and scope. For intermediate cases the log marginal cost function now becomes

$$\ln MC^c_{i,j,t} = \begin{cases} \frac{\phi-\alpha}{\phi}\ln S_{i,j,t} + \frac{1-\phi}{\phi}\ln\left(\frac{Y^c_{i,j,t}}{S^c_{i,j,t}}\right) + \ln\left(\frac{1}{\phi}g(\vec{W}-a\cdot 1_K)\right) - \frac{1}{\phi}\ln\left(\frac{\omega_{i,j,t}}{\tau_{i,j,c,t}}\right) \\ \frac{\phi^f-\alpha}{\phi^f}\ln S_{i,j,t} + \frac{1-\phi^f}{\phi^f}\ln\left(\frac{Y^c_{i,j,t}}{S^c_{i,j,t}}\right) + \ln\left(\frac{1}{\phi^f}g^f(\vec{W})\right) - \frac{\beta_k}{\phi^f}\ln K - \frac{1}{\phi^f}\ln\left(\frac{\omega_{i,j,t}}{\tau_{i,j,c,t}}\right) \\ \frac{\phi-\alpha}{\phi}\ln S_{i,j,t} + \frac{1-\phi}{\phi}\ln\left(\frac{Y^c_{i,j,t}}{S^c_{i,j,t}}\right) + \ln\left(\frac{g(\vec{W}+a\cdot 1_K)}{\phi}\right) - \frac{1}{\phi}\ln\left(\frac{\omega_{i,j,t}}{\tau_{i,j,c,t}}\right), \end{cases} \tag{15}$$

making it impossible to absorb $g(\vec{W})$ into the composite error term unless we have instruments that shift $S_{i,j,t}$ and $Y^c_{i,j,t}/S^c_{i,j,t}$ without shifting $\mathcal{Y}$.

If the endogeneity problem did not exist a researcher regressing $\ln MC^c_{i,j,t}$ on $\ln S_{i,j,t}$ and $\ln\frac{Y^c_{i,j,t}}{S^c_{i,j,t}}$ would recover a weighted average of the coefficents,

$$\beta_S = \pi_0\frac{\phi_f-\alpha}{\phi_f} + (1-\pi_0)\frac{\phi-\alpha}{\phi} \text{ and } \beta_{\frac{Y^c}{S^c}} = \pi_0\frac{1-\phi_f}{\phi_f} + (1-\pi_0)\frac{1-\phi}{\phi}$$

where $\pi_0 = \Pr(\mathcal{Y} \in [\mathcal{Y}^-, \mathcal{Y}^+])$ is the fraction of observations for which firms do not adjust their capital.

Attempting to recover ϕ naively using $\hat{\phi} = \frac{1}{1+\beta_{\frac{Y^c}{S^c}}}$ as in the case of no adjustment costs would give an estimate for returns to scales of

$$\hat{\phi} = \frac{\phi\phi^f}{\pi_0\phi + (1-\pi_0)\phi^f}$$

while α could still be recovered by $\hat{\alpha} = \hat{\phi}(1-\beta_S)$ since

$$\hat{\phi}\beta_S = \frac{\pi_0\phi(\phi_f-\alpha) + (1-\pi_0)\phi^f(\phi-\alpha)}{\pi_0\phi + (1-\pi_0)\phi^f} = \frac{\phi^f\phi}{\pi_0\phi + (1-\pi_0)\phi^f} - \alpha = \hat{\phi} - \alpha.$$

However, even α is unlikely to be recovered correctly in practice once endogeneity of $\tilde{W}_K$ is taken into account.

While the model looks very similar to the structural threshold model of Kourtellos et al. (2016), the fact that the threshold variable, $\mathcal{Y}$, is not just endogenous but also unobservable as it depends on the transportation-cost-adjusted productivities introduces complexities that mean standard methods cannot be used.

This highlights the need of complementing demand with supply side data, which can be used to recover the full CES frontier using existing methods like those of Caselli et al. (2025). However, the Khmelnitskaya et al. (2025)'s CES frontier can only accommodate technologies without returns to scope if each good is produced with the same homogeneous production function. This rules out industries like car manufacturing where motor cars are labour intensive while electric cars are capital intensive Cairncross et al. (2023).

In the next section, I discuss how to test whether production is non-joint or exhibits returns to scope, and develop a non-parametric identification result for the joint cost function to accommodate functional forms like translog production functions.

5 Non-jointness

Hall (1973) studied the conditions that features like non-jointness and separability of the transformation function impose on the joint cost function when firms can buy all inputs on perfectly competitive markets, including his famous general theorem on non-jointness.

Theorem 5.1 (General Theorem on Non-Jointness (Hall, 1973))**.** *A technology is non-joint if and only if the joint cost function can be written as the sum of independent cost functions for each kind of output:*

$$C(\vec{Y}, \vec{W}) = \sum_{i=1}^{N} C^{(i)}(Y_i, \vec{W})$$

This result has allowed for a great number of papers that would not otherwise have been possible given the available data at the time. However, with increasing availability of detailed production data it is useful and necessary to update this result to account for fixed inputs and input market power.

According to Hall's theorem, if all inputs are bought on a perfectly competitive market, the joint cost function is non-joint if and only if all cross-output terms are identically zero. Let $C_{Y_iY_j}$ denote the second order partial derivative of the joint cost function with respect to Y_i and Y_j.

Corollary 1. *The joint cost function, $C(\vec{Y}, \vec{W})$, is non-joint if and only if*

$$C_{Y_iY_j} = 0 \; \forall i \neq j,$$

where $C_{Y_iY_j} = \frac{\partial^2 C}{\partial Y_i \partial Y_j}$

However, when capital is fixed, the implication of non-jointness changes since we need to condition each product-specific cost function on the optimally allocated capital input.

Corollary 2. *If all flexible inputs are bought on a competitive market, a technology is non-joint if and only if*

$$C(\vec{Y}, \vec{W}, K) = \min_{\{K_i\}_{i=1}^N} \sum_{i=1}^{N} C^{(i)}(Y_i, \vec{W} | K_i),$$

subject $K_i \geq 0 \; \forall i$ *and* $\sum_i K_i = K$.

This now also means that the cross effect is non-zero exactly because it equals the re-allocation term.

$$C_{Y_iY_j} = C^{(i)}_{K_i} \frac{\partial K_i}{\partial Y_j}.$$

Nonetheless, it possible to understand whether a technology is non-joint from the residual joint cost function.

Propostion 5.1. *The residual joint cost function, $C(\vec{Y}, \vec{W}, K)$, represents a non-joint technology if and only if*

$$C_{Y_iY_j} C_{KK} - C_{Y_iK} C_{Y_jK} = 0 \qquad (16)$$

Proof. For the if direction, assume that the technology represented by $C(\cdot)$ is non-joint, then, by Corollary 2 ,$C(\vec{Y},\vec{W},K) = \min_{\vec{K}} \sum_{i=1}^{N} C^{(i)}(Y_i,\vec{W}|K_i)$ and

$$C_{Y_i} = C_{Y_i}^{(i)} + \sum_{j=1}^{N} C_{K_j}^{(j)} \frac{\partial K_j}{\partial Y_i},$$

but, since cost minimisation implies that $C_{K_i}^{(i)} = C_{K_j}^{(j)}\ \forall i,j$ and since $\frac{\partial K_N}{\partial Y_i} = -\sum_{j=1}^{N-1} \frac{\partial K_j}{\partial Y_i}$ as capital is allocated exclusively and exhaustively, this simplifies to

$$C_{Y_i} = C_{Y_i}^{(i)}.$$

Similarly,

$$C_K = C_{K_i}^{(i)} \sum_{i=1}^{N} \frac{\partial K_i}{\partial K} = \sum_{i=1}^{N} C_{K_i}^{(i)} \frac{\partial K_i}{\partial K} = C_{K_i}^{(i)},$$

for any i since $\sum_{i=1}^{N} \frac{\partial K_i}{\partial K} = \frac{\partial K}{\partial K} = 1$.

Differentiating these expressions again gives us

$$C_{Y_iY_j} = C_{Y_iK_i}^{(i)} \frac{\partial K_i}{\partial Y_j},$$

$$C_{Y_iK} = C_{Y_iK_i}^{(i)} \frac{\partial K_i}{\partial K},$$

$$C_{KK} = C_{K_iK_i}^{(i)} \frac{\partial K_i}{\partial K}.$$

However, note that, since $C_{Y_iK} = C_{KY_i}$ as $C(\cdot)$ is continuously twice differentiable, and we can also write

$$C_{Y_jK} = C_{K_iK_i}^{(i)} \frac{\partial K_i}{\partial Y_j}.$$

Then putting it all together

$$C_{Y_iY_j} C_{KK} = C_{Y_iK_i}^{(i)} C_{K_iK_i}^{(i)} \frac{\partial K_i}{\partial Y_j} \frac{\partial K_i}{\partial K}$$

and

$$C_{Y_iK} C_{Y_jK} = C_{Y_iK_i}^{(i)} C_{K_iK_i}^{(i)} \frac{\partial K_i}{\partial Y_j} \frac{\partial K_i}{\partial K}.$$

For the only if direction, assume that there exist N product specific residual cost functions, $\{C^{(i)}(Y_i, \vec{W}|K_i)\}_{i=1}^N$, such that $C(\vec{Y}, \vec{W}, K) = \min_{\vec{K}} \sum_{i=1}^N C^{(i)}(Y_i, \vec{W}|K_i)$ with $K_i \geq 0$ and $\sum_i K_i = K$. Then by Corollary 2 the technology is non-joint. □

The advantage of using cost functions is that they can be estimated with no or, in the case of residual cost functions, minimal production data. A disadvantage of the cost function approach is that this simplicity relies on the assumption of perfectly competitive input markets and no adjustment costs.

However, recently there has been an increased interest in estimating the underlying production technology to study firms' input market power,including from competition regulators CMA (e.g. 2024). Therefore, it is worth considering what the restrictions non-jointness imposes on production frontiers, which directly model the input-output relationship and thereby avoid restrictions on input market power.[4]

Definition 1. The marginal rate of substitution (MRTS) is the rate at which one input can be substituted for another while keeping the level of outputs fixed. Formally, define the MRTS between inputs X_a and X_b as

$$MRTS^{a,b}(\vec{Y}, \vec{X}) := -\frac{\partial X_a}{\partial X_b}\bigg|_{(\vec{Y},\vec{X})} = -\frac{T_{X_b}}{T_{X_a}}\bigg|_{(\vec{Y},\vec{X})}.$$

Since the cross derivative, $T_{Y_iY_j}$, keeps inputs constant, testing for non-jointness is again complicated by the fact that re-optimisation leads to an interdependence even when the underlying technology is non-joint. However, we can use the fundamental result that an observed input allocation is cost minimising relative to the shadow price vector for inputs, which is proportional to its MRTS. That is, $\vec{X}$ is the cost minimising input choice when $MRTS^{m,n} = \frac{W_{X_m}}{W_{X_n}}$ $\forall m, n$. Therefore, the following corollary is a direct implication of Hall's general theorem on non-jointness.

Corollary 3. *A technology is non-joint if and only if the joint cost function can be written as the sum of independent cost functions for each kind of output, where the*

[4] The class of production technologies modelled by production frontiers is a superset of those modelled by production functions. Since most production datasets do not contain input allocations, studying production frontiers also avoids the problem of having to impute the allocations from firms' optimisation conditions.

vector of input prices is the normalised vector of MRTSs. That is, where

$$\vec{W}^{MRTS} = \left(1, \frac{T_{x_2}}{T_{x_1}}, \frac{T_{x_3}}{T_{x_1}}, ..., \frac{T_{x_M}}{T_{x_1}}\right).$$

This result is intuitive since it states that the marginal cost of producing more of product i is independent of the amount of any other product produced when measuring costs using the firm's subjective input price vector. We can use this result to derive a test of whether a technology is non-joint.

Propostion 5.2. *The technology represented by the production frontier* $T(\vec{Y}, \vec{X}) = 1$ *is non-joint if and only if, for all* $(\vec{Y}, \vec{X})$, $(\vec{Y}', \vec{X}')$ *with* $Y_i = Y_i'$,

$$\exists a \in \mathbb{R} \ s.t. \ \frac{T_{X_m}(\vec{Y}, \vec{X})}{T_{X_m}(\vec{Y}', \vec{X}')} = a \ \forall m$$

$$\implies \frac{T_{Y_i}(\vec{Y}, \vec{X})}{T_{Y_i}(\vec{Y}', \vec{X}')} = a \tag{17}$$

Proof. For the only if direction assume that, whenever $Y_i = Y_i'$ and

$$\frac{T_{X_m}(\vec{Y}, \vec{X})}{T_{X_m}(\vec{Y}', \vec{X}')} = a \ \forall m,$$

then $\frac{T_{Y_i}(\vec{Y}, \vec{X})}{T_{Y_i}(\vec{Y}', \vec{X}')} = a$.

$$\frac{T_{X_m}(\vec{Y}, \vec{X})}{T_{X_m}(\vec{Y}', \vec{X}')} = \frac{T_{X_1}(\vec{Y}, \vec{X})}{T_{X_1}(\vec{Y}', \vec{X}')} = a$$

$$\implies MRTS^{m,n} = \frac{T_{X_b}(\vec{Y}, \vec{X})}{T_{X_m}(\vec{Y}, \vec{X})} = \frac{T_{X_n}(\vec{Y}', \vec{X}')}{T_{X_m}(\vec{Y}', \vec{X}')} = MRTS^{m,n\prime}$$

Therefore, the vector of shadow input prices is the same across both observations, $\vec{W}^{MRTS} = \vec{W}^{MRTS\prime}$. Furthermore, since $T_{Y_i}(\vec{Y}, \vec{X})/T_{Y_i}(\vec{Y}', \vec{X}')$ whenever $Y_i = Y_i'$, and since, from the lagrangian, $MC_{y_i}(\vec{Y}, \vec{X}) = \frac{T_{X_m}(\vec{Y}, \vec{X})}{T_{Y_i}(\vec{Y}, \vec{X})}$ the marginal costs are the same. Then by integrating and using that $C(\vec{0}, \vec{W}, \vec{A}) = 0$ to eliminate the constant by Corollary 3 we get that the technology is non-joint.

The if direction similarly uses Corollary 3. The conclusion then follows from the definition of the MRTS and MC. □

While we now have the tools to test for non-jointness, we still need to understand what class of models to study if we want to be able to nest non-joint production with common production function specifications like Cobb Douglas or translog.

6 Non-parametric cost function

6.1 Separable technologies and standard production models

Definition 2. A production frontier is *separable* in inputs and outputs if there exist functions $G(\vec{Y})$ and $H(\vec{X})$ such that

$$T(\vec{Y}, \vec{X}) = G(\vec{Y})/H(\vec{X}).$$

The same holds for the joint cost function.

Separable technologies have a number of properties that make them popular in the literature. For example, they make it possible to explicitly solve the firm's profit maximising first order conditions to recover the unobserved productivities. However, as shown by Cairncross et al. (2023), a separable production frontier that is homogeneous in inputs is non-joint if and only if the production function is common across products and equals the input component, $H(\cdot)$.[5] These specifications, therefore, non-trivially restrict the technologies that they can model.

More importantly, even in the simple case where two products are produced with separate constant returns to scale Cobb Douglas production functions with different input intensities, the production frontier does not have an explicit solution except for a handful of parameters Jensen (2025, p. 718). Naturally, this failure extends to the joint cost function. This makes it necessary to at least use non-separable and at worst non-parametric specifications when aiming to nest standard non-joint and joint technologies. Fortunately, as I will show below the joint cost function is non-parametrically identified when we impose the standard productivity process.

[5] In other words, non-jointness requires that $\frac{Y_i}{A_i} = H(\vec{X}_i)\ \forall i$.

6.2 Non-parametric identification

Here, I will address the simpler identification problem of cost functions. The joint cost function has advantages over production frontiers as marginal costs can be directly inferred making it unnecessary to identify the shadow cost parameter on the feasibility constraint, which requires an additional first order condition. Additionally, since the cost function aggregates information we only need to observe the price of an input if it is flexible or its quantity if it is pre-determined, but not both.

Naturally, these benefits come at a cost. In particular, we need to assume that all inputs are either pre-determined or bought on a competitive market.[6] It is also worth noting that I assume that, other than productivity, there are no hidden state variables. For technologies that can be modelled using production functions this means pre-determined inputs can be freely allocated between different production processes without any adjustment costs (see e.g. Orr, 2022; Valmari, 2023).

Assumption 1. The inputs can be partitioned into flexible inputs bought on a competitive market, $\vec{X}^{(1)}$, with input prices, $\vec{W}$, and pre-determined inputs, $\vec{K}$. The set of pre-determined inputs may be empty.

Let $\omega_{i,j,t} = \ln(A_{i,j,t})$ denote firm j's log productivity for producing good i at time t, so that $\vec{\omega}_{j,t}$ is the firm's vector of log productivities. Furthermore, let $\zeta_{i,j,t}$ be the (unexpected) innovation in productivity. I will make the standard assumption that productivity follows an AR(1) process. As it turns out, this structure allows us to offset the additionally generality of having firm-product specific productivities without needing to restrict how these, or any of the instruments, enter the marginal cost function.

Assumption 2. Productivity evolves according to an AR(1) process:

$$\omega_{i,j,t} = \rho\omega_{i,j,t-1} + \zeta_{i,j,t},$$

where $\rho \in (0,1)$, $E_i[\zeta_{i,j,t}] = 0$ and $\vec{\zeta}_{j,t}$ is iid.

This formulation allows for rich within-firm dependence in productivity subject to assumption 7, but does necessitate that the persistence parameter is common across

[6] If an input was not bought on a competitive market the economist would need to observe the adjustment cost or price schedule the firm faces, which is almost never the case in practice.

products.

It is worth also re-stating the Hicks-neutrality assumption.

Assumption 3. $\vec{\omega}$ is a vector of product-specific Hicks-neutral productivities. That is,

$$C(\vec{Y},\vec{W},\vec{K},\vec{\omega}) = C(\vec{Q},\vec{W},\vec{K},\vec{0})$$

where $Q_i = Y_i e^{-\omega_i}$

The multi-product literature typically recovers the firm's productivities by inverting its profit maximising first order conditions,

$$\vec{mr} = \vec{mc}(\vec{Y},\vec{W},\vec{\omega}). \tag{18}$$

Therefore, it is typically assumed that the marginal revenues can be identified from the demand side separately. However, as marginal revenues are equilibrium objects, they are not independently distributed from the productivity shocks. Therefore, to reduce the complexity of the system of equations studied in this section, I will assume that firms sell their outputs in perfectly competitive markets.

Assumption 4. Firms sell their products on perfectly competitive markets with observed prices, $\vec{P}$, which are distributed independently of $\vec{\zeta}$.

We also need to restrict the class of joint cost functions to those that allow for the inversion of the firm's profit maximisation condition to recover the unobserved productivities. Therefore, we need the system of the firm's marginal costs, $\vec{mc} := \partial C/\partial\vec{Y}$, to satisfy Hadamard's global function inversion theorem:

Theorem 6.1 (Hadamard's global function inversion theorem)**.** *A C^1-map $f : \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism if and only if the Jacobian determinant is non-vanishing everywhere and $||f(x)|| \to \infty$ whenever $||x|| \to \infty$.*[7]

Assumption 5 ensures that the system of marginal costs satisfies Hadamard's invertibility condition everywhere. That the marginal cost system is proper naturally derives from the basic assumption that $\lim_{\omega_i \to -\infty} mc_i = \infty$ and $\lim_{\omega_i \to \infty} mc_i = 0$ while $mc_i(\vec{Y},\vec{W},\vec{K},\vec{\omega}) > 0$ for any finite $\vec{\omega}$.

[7]In $\mathbb{R}^n$ properness and the limit condition used here are equivalent.

Assumption 5. The cost function, $C(\cdot)$, is thrice continuously differentiable[8] and

$$det\left(\frac{\partial \vec{mc}(\vec{Y},\vec{W},\vec{K},\vec{\omega})}{\partial \vec{\omega}}\right) \neq 0 \;\forall(\vec{Y},\vec{W},\vec{K},\vec{\omega}).$$

However, as this is an abstract condition it is useful to consider what meaningful economic restrictions can be sufficient for invertibility of the productivities. A natural candidate is that increasing the productivity of a product i impacts its own marginal cost more than the total impact it has on the marginal costs of all other goods. The proof that assumption 6 implies assumption 5 is in appendix B.1

Assumption 6.

$$\left|\frac{\partial mc_i}{\partial \omega_i}\right| > \sum_{j\neq i}\left|\frac{\partial mc_j}{\partial \omega_i}\right|$$

As discussed in appendix B.2 assumption 6 holds for example for the cost function implied by Cairncross et al. (2023)'s production frontier, which was discussed in section 4.1.

We now know that we can recover the productivities by inverting the firm's first order conditions with respect to its output price,

$$\vec{\omega} = \vec{mc}^{-1}(\vec{Y},\vec{W},\vec{P}). \tag{19}$$

However, to recover non-parametric identification I also need to make an assumption on the amount of variation in the data. This takes the form of an assumption on the underlying distribution of observables.

Assumption 7. For every $(\vec{Y}_{t-1},\vec{W}_{t-1},\vec{K}_{t-1},\vec{P}_{t-1})$, for every $i \in (1,2,...,N)$, there exists $\left(\vec{Y}_t^{(i)},\vec{W}_t^{(i)},\vec{K}_t^{(i)},\vec{P}_t^{(i)}\right)$ such that

$$\frac{\partial f_\zeta\left(\vec{mc}^{-1}\left(\vec{Y}_t^{(i)},\vec{W}_t^{(i)},\vec{K}_t^{(i)},\vec{P}_t^{(i)}\right) - \rho\,\vec{mc}^{-1}\left(\vec{Y}_{t-1},\vec{W}_{t-1},\vec{K}_{t-1},\vec{P}_{t-1}\right)\right)}{\partial \zeta} \propto \vec{e}_i,$$

where $\vec{e}_i$ is the i-th standard basis vector of $\mathbb{R}^N$.

[8]We need the marginal costs to be twice continuously differentiable to apply Matzkin (2008)'s result below.

To avoid notational clutter, I will also drop the vector notation in the theorems and proof below. To further ease notation let $s = \vec{Y}_t$ denote the endogenous variables, $z^{(1)} = (\vec{W}_t, \vec{K}_t, \vec{P}_t)$ denote the contemporaneous exogenous variables. Furthermore, let $z^{(2)} = (\vec{Y}_{t-1})$ and $z^{(3)} = (\vec{W}_{t-1}, \vec{K}_{t-1}, \vec{P}_{t-1})$ denote their lagged counterparts, and let $z = (z^{(1)}, z^{(2)}, z^{(3)})$ be the full vector of exogenous variables. Lastly, let $r(\cdot) = \vec{mc}(s, z^{(1)}) - \rho\, \vec{mc}^{-1}(z^{(2)}, z^{(3)})$ be the system of equations that maps to our unobserved productivity shocks, $\vec{\zeta}$, as a function of observables. Then it is instructive to recall Matzkin (2008)'s nonparametric identification result.

Theorem 6.2 (Matzkin (2008))**.** *There exists $f_{\tilde{\zeta}}$ such that $(\tilde{r}, f_{\tilde{\zeta}})$ is observationally equivalent to (r, f_ζ) if and only if for all (s, z) the matrix*

$$\begin{pmatrix} \left(\frac{\partial \tilde{r}(s,z)}{\partial s}\right)' & \Delta_s(s,z,r,\tilde{r}) + \left(\frac{\partial r(s,z)}{\partial s}\right)' \frac{log\left(f_\zeta(r(s,z))\right)}{\partial \zeta} \\ \left(\frac{\partial \tilde{r}(s,z)}{\partial z}\right)' & \Delta_z(s,z,r,\tilde{r}) + \left(\frac{\partial r(s,z)}{\partial z}\right)' \frac{log\left(f_\zeta(r(s,z))\right)}{\partial \zeta} \end{pmatrix} \tag{20}$$

where

$$\Delta_s(s,z,r,\tilde{r}) = \frac{\partial}{\partial s} \log det\left(\frac{\partial r(s,z)}{\partial s}\right) - \frac{\partial}{\partial s} \log det\left(\frac{\partial \tilde{r}(s,z)}{\partial s}\right)$$

and

$$\Delta_z(s,z,r,\tilde{r}) = \frac{\partial}{\partial z} \log det\left(\frac{\partial r(s,z)}{\partial s}\right) - \frac{\partial}{\partial z} \log det\left(\frac{\partial \tilde{r}(s,z)}{\partial s}\right)$$

and $det(\cdot)$ is the determinant, has rank N.

Matzkin also points out that, if r is identified, so is the distribution of errors, f_ζ. Therefore, I will throughout refer to observationally equivalent $\tilde{r}$ to mean the observationally equivalent pair $(\tilde{r}, f_{\tilde{\zeta}})$.

I will make one further assumption on the distribution of observables, namely that it is possible for a firm to have the same observables over two periods.

Assumption 8. $f_{s,z}(s, z^{(1)}, s, z^{(1)}) \neq 0$ for any $(s, z^{(1)})$, where $f_{s,z}$ is the observed marginal distribution of the endogenous and all exogenous variables.

With this we have the necessary tools to prove the following nonparametric identification result on the cost function.

Theorem 6.3. *Given assumptions 1-8, the cost function is non-parametrically point identified.*

Proof. We can use the separable structure of $\vec{\zeta}$ to write the matrix in equation (20) as a matrix with four blocks of rows

$$\begin{pmatrix} \left(\tilde{mc}_y^{-1}(s,z^{(1)})\right)' & \Delta_s(s,z,r,\tilde{r}) + \left(mc_y^{-1}(s,z^{(1)})\right)' \gamma_\zeta(s,z) \\ \left(\tilde{mc}_x^{-1}(s,z^{(1)})\right)' & \Delta_{z^{(1)}}(s,z,r,\tilde{r}) + \left(mc_x^{-1}(s,z^{(1)})\right)' \gamma_\zeta(s,z) \\ -\tilde{\rho}\left(\tilde{mc}_y^{-1}(z^{(2)},z^{(3)})\right)' & \Delta_{z^{(2)}}(s,z,r,\tilde{r}) - \rho\left(mc_y^{-1}(z^{(2)},z^{(3)})\right)' \gamma_\zeta(s,z) \\ -\tilde{\rho}\left(\tilde{mc}_x^{-1}(z^{(2)},z^{(3)})\right)' & \Delta_{z^{(3)}}(s,z,r,\tilde{r}) - \rho\left(mc_x^{-1}(z^{(2)},z^{(3)})\right)' \gamma_\zeta(s,z) \end{pmatrix}, \tag{21}$$

where $\gamma_\zeta(s,z) = \frac{log\left(f_\zeta(mc^{-1}(s,z^{(1)}) - \rho mc^{-1}(z^{(2)},z^{(3)}))\right)}{\partial\zeta}$.

Step 1 - $\Delta_s = \Delta_{z^{(1)}} = \Delta_{z^{(2)}} = \Delta_{z^{(3)}} = 0$ *and* $\rho = \tilde{\rho}$
The separability between $(s, z^{(1)})$ and $(z^{(2)}, z^{(3)})$ means that $\Delta_{z^{(2)}} = \Delta_{z^{(3)}} = 0$ since, for any $v \in \{s, z^{(1)}, z^{(2)}, z^{(3)}\}$, $\Delta_v = \frac{\partial}{\partial v}\log\left(det\left(mc_y^{-1}(s,z^{(1)})\right)\right) - \frac{\partial}{\partial v}\log\left(det\left(\tilde{mc}_y^{-1}(s,z^{(1)})\right)\right)$. We will also use the insight that $\Delta_v(s,z,r,\tilde{r}) = \Delta_v(s,z^{(1)},mc^{-1},\tilde{mc}^{-1})$ and therefore is independent of ρ and $\tilde{\rho}$ in the next step.

Since the matrix has $N+1$ columns, by Matzkin's theorem, we know that, for all observationally equivalent $\tilde{r}$, the rightmost column can be written as a linear combination of the first N columns. Additionally, from Matzkin's proof we know that the vector by which we multiply the first N columns to equate them to the last column, $\gamma_{\tilde{\zeta}}$, also equals $\partial f_{\tilde{\zeta}}(\tilde{r}(s,z))/\partial\tilde{\zeta}$. Evaluating the first and third row block of the matrix at $(s,z) = (s, z^{(1)}, s, z^{(1)})$ and dividing both sides of the third row block by $-\tilde{\rho}$ allows us to conclude that

$$\Delta_s(s,z,mc^{-1},\tilde{mc}^{-1}) = \left(\frac{\rho}{\tilde{\rho}} - 1\right) mc^{-1}(s,z^{(1)})'\gamma_u(s,z^{(1)},s,z^{(1)}).$$

However, since Δ_s does not depend on ρ, the only way for this equality to hold is if $\tilde{\rho} = \rho$, thereby identifying the AR(1) persistence parameter. Furthermore, when $\tilde{\rho} = \rho$, the right-hand side is equal to 0. Following the same steps for the 2nd and 4th row block then allows us to conclude that $\Delta_s = \Delta_{z^{(1)}} = 0$.

Step 2 - $\left(\tilde{mc}^{-1}(y,x)'\right)^{-1} mc^{-1}(y,x) = \Phi$
The next step is to show that $\tilde{mc}^{-1}$ and mc^{-1} are related by a constant matrix. Due to the underlying structure $\gamma_\zeta(s,z) = \gamma_\zeta(mc^{-1}(s,z^{(1)}), mc^{-1}(z^{(2)},z^{(3)}))$.

Fix an arbitrary $(s, z^{(1)})$. Then, (if they exist) pick any $(z^{(2)}, z^{(3)})$, $(z^{(2)\prime}, z^{(3)\prime})$ such that $mc^{-1}(z^{(2)}, z^{(3)}) = mc^{-1}(z^{(2)\prime}, z^{(3)\prime})$. Then $\gamma_\zeta(s, z^{(1)}, z^{(2)}, z^{(3)}) = \gamma_\zeta(s, z^{(1)}, z^{(2)\prime}, z^{(3)\prime})$, so that $\gamma_{\tilde{\zeta}}(s, z^{(1)}, z^{(2)}, z^{(3)}) = \gamma_{\tilde{\zeta}}(s, z^{(1)}, z^{(2)\prime}, z^{(3)\prime})$. Using this fact and letting

$$\Phi(y, x) := \left(\tilde{mc}_y^{-1}(y, x)'\right)^{-1} mc_y^{-1}(y, x)'$$

and

$$b := mc^{-1}(z^{(2)}, z^{(3)}) = mc^{-1}(z^{(2)\prime}, z^{(3)\prime}),$$

we can use the first equation block to show that

$$\left(\Phi(z^{(2)}, z^{(3)}) - \Phi(z^{(2)\prime}, z^{(3)\prime})\right) \gamma_\zeta(mc^{-1}(s, z^{(1)}), b) = 0.$$

Using assumption 7 to let $\gamma_\zeta(mc^{-1}(s, z^{(1)}), b)$ vary over the basis vectors, this equality implies that $\Phi(y, x)$ is at most a function of $mc^{-1}(y, x)$.

Therefore, for any arbitrary $(s, z^{(1)})$, $(z^{(2)}, z^{(3)})$, we can now solve for $\gamma_{\tilde{\zeta}}$ from the first and third equation block to get

$$\left(\Phi(mc^{-1}(s, z^{(1)})) - \Phi(mc^{-1}(z^{(2)}, z^{(3)}))\right) \gamma_\zeta = 0.$$

Furthermore, fixing $(z^{(2)}, z^{(3)})$ and again using assumption 7 to pick $(s^i, z^{i,(1)})$ so that $\gamma_\zeta \propto e_i$ and letting $a(z^{(2)}, z^{(3)}) := mc^{-1}(s^i, z^{i,(1)})$ gives that the i-th row of the respective matrices are the same,

$$\Phi_i(b) = \Phi_i(a(z^{(2)}, z^{(3)})).$$

Now, by assumption 8, we can also set $(z^2, z^3) = (s^i, z^{i,(1)})$ and repeat this process to get

$$\Phi_i(a(z^{(2)}, z^{(3)})) = \Phi_i(a(s^i, z^{i,(1)})).$$

Using the AR(1) form, assumptions 7 and 8, and abusing notation to instead write $a(b) = \vec{\zeta}^{e_i} - \rho b$ we can then define the above procedure of repeatedly using the mapping n times as

$$a^n(b) = \left(\sum_{k=1}^{n} (-\rho)^{k-1}\right) \vec{\zeta}^{e_i} - \rho^n b.$$

Notably,

$$\lim_{n\to\infty} a^n(b) = \frac{\rho^2}{1-\rho^2}\vec{\zeta}^{e_i}$$

is independent of the starting point. Then using that

$$\Phi_i(b) = \Phi_i(a(b)) = \Phi_i(a^2(b)) = ... = \Phi_i(a^n(b))$$

and taking the limit, which is independent of b, we get that the i-th row of Φ is constant

$$\Phi_i(b) = \lim_{n\to\infty} \Phi_i(a^n(b)) = \Phi_i\left(\frac{\rho^2}{1-\rho^2}\vec{\zeta}^{e_i}\right) \ \forall b.$$

Repeating this process for each i allows us to conclude that Φ is a constant and invertible matrix.

Step 3 - Hicks-neutrality and mean zero shocks imply $\Phi = I$

Let λ_i denote the ith column of Φ^{-1} and let mc_i be the $i-th$ element of mc. Since $\tilde{mc}_Y^{-1} = mc_y^{-1}\Phi^{-1}$,

$$\frac{\partial \tilde{mc}_i^{-1}}{\partial Y_i} = \sum_{k=1}^{N} \frac{\partial \tilde{mc}_k^{-1}}{\partial Y_i}\lambda_{i,k}.$$

Then integrating with respect to Y_i gives

$$mc_i^{-1} = \sum_{k=1}^{N} mc_k^{-1}\lambda_{i,k} + B_i(\vec{Y}_{-i}, \vec{X}),$$

where Y_{-i} denotes the vector of all outputs other than Y_i. However, since the same procedure works for each element of $\vec{Y}_{-i}$ and $\vec{X}$, we can conclude that B_i is a constant. Since

$$0 = E[\tilde{\zeta}_i] = E[\tilde{mc}_i^{-1}(\cdot) - \rho\tilde{mc}_i^{-1}(\cdot)] = \sum_{k=1}^{N} E[\zeta_k]\lambda_{i,k} + B_i = B_i,$$

we can conclude that $B_i = 0$ for all i.

Therefore, the productivity vector is identified up to a set of linearly independent linear combinations

$$\vec{\tilde{\omega}}(\vec{Y},\vec{W},\vec{K},\vec{P}) = \Phi^{-1}\vec{\omega}(\vec{Y},\vec{W},\vec{K},\vec{P}).$$

However, note that this means that

$$\tilde{mc}(\vec{Y},\vec{W},\vec{K},\vec{\omega}) = mc(\vec{Y},\vec{W},\vec{K},\Phi^{-1}\vec{\omega}).$$

Therefore,

$$\tilde{C}(\vec{Y},\vec{W},\vec{K},\vec{\omega}) = C(\vec{Y},\vec{W},\vec{K},\Phi^{-1}\vec{\omega}) + C_0(\vec{W},\vec{K},\vec{\omega}).$$

But we can conclude that $C_0 = 0$ since we need that $\lim_{\vec{Y}\to\vec{0}} \tilde{C}(\vec{Y},\vec{W},\vec{K},\vec{\omega}) = 0$.

Assumption 3 requires that $\frac{\partial\tilde{C}}{\partial Y_i} = -\frac{1}{Y_i}\frac{\partial\tilde{C}}{\partial\omega_i}$. However,

$$\frac{\partial\tilde{C}(\vec{Y},\vec{W},\vec{K},\vec{\omega})}{\partial\omega_i} = \sum_{k=1}^{N}\frac{\partial C(\vec{Y},\vec{W},\vec{K},\Phi^{-1}\vec{\omega})}{\partial\omega_k}\lambda_{i,k} = -\sum_{k=1}^{N} y_k\frac{\partial C(\vec{Y},\vec{W},\vec{K},\Phi^{-1}\vec{\omega})}{\partial Y_k}\lambda_{i,k}$$
$$\neq -Y_i\frac{\partial C(\vec{Y},\vec{W},\vec{K},\Phi^{-1}\vec{\omega})}{\partial Y_i} \;\forall(\vec{Y},\vec{W},\vec{K},\vec{\omega}) \quad (22)$$

unless there exists a function $H(\cdot)$ such that $C(\vec{Y},\vec{W},\vec{K},\vec{\omega}) = H\left(\mathcal{Y},\vec{W},\vec{K}\right)$ with $\mathcal{Y} = \prod_{k=1}^{N} Y_k e^{-\omega_k}$ in which case it would violate the invertibility conditions of assumption 5. We can conclude that $\Phi^{-1} = \Phi = I$. □

The proof shows that the usual GMM moments

$$E[\vec{\zeta}_t|\vec{W}_t,\vec{K}_t,\vec{Y}_{t-1},\vec{W}_{t-1},\vec{K}_{t-1},\vec{P}_{t-1};\theta] = 0$$

can be used to estimate the system of marginal cost functions. The cost function can then be recovered through integration.

Additionally, since the set of fixed input is allowed to be empty this result extends the existing non-parametric cost function identification results to the case of arbitrary technologies with product-firm-time specific productivities. If the researcher is confident that all inputs are flexible and bought on perfectly competitive input markets this means that we can recover the structural cost function using only demand data and data on input prices, which are typically available.

6.3 Finite sample testing of non-jointness

The non-parametric identification result is very useful as it shows that remaining agnostic on the jointness of the technology is possible. However, it is important to acknowledge that non-parametric convergence rates are significantly slower than their parametric counterparts. Additionally, the number of instruments required for estimation explodes in the precision of the approximation and higher order powers of observable variables tend quickly become approximately collinear. For this reason, work on production functions for single product firms has tended to favour a second order translog approximation, allowing for flexible approximations of arbitrary production functions while requiring only a reasonable number of instruments. In the multi-product setting the number of parameters will also explode as the number of goods grows. Nonetheless, the translog specification remains manageable for small numbers of goods.

Take a firm that buys labour and materials on a perfectly competitive market. Since the cost function is homogeneous of degree 1 in input prices it is standard to normalise the input price vector. Let $W = W_l/W_m$ denote the relative price of labour to materials. Then if we use a translog approximation for the level of the joint cost function[9] we get

$$\begin{aligned} C(\vec{Y}, W, K) = \beta_0 + \beta_w \ln W + \beta_k \ln K + \sum_i \beta_{y_i}(\ln Y_i - \omega_i) + \beta_{w^2}(\ln W)^2 \\ + \beta_{wk} \ln W \ln K + \beta_{k^2}(\ln K)^2 + \sum_i \beta_{y_i w}(\ln Y_i - \omega_i) \ln W + \\ \sum_i \beta_{y_i k}(\ln Y_i - \omega_i) \ln K + \sum_i \sum_j \beta_{y_i y_j}(\ln Y_i - \omega_i)(\ln Y_j - \omega_j) \end{aligned} \quad (23)$$

Taking the second order derivatives to compute for what parameter values the technology is non-joint we see that

$$C_{Y_i Y_j} = \frac{2\beta_{y_i y_j}}{Y_i Y_j},$$

[9] Production models typically assume that nothing gets produced with no inputs so that it is typical to approximate the log cost function, $\ln C$, instead so that $\lim_{\vec{Y} \to \vec{0}} C = 0$. When approximating the level equation with a translog we instead generally get that $\lim_{\vec{Y} \to \vec{0}} C = \pm\infty$ since $\lim_{Y_i \to 0} \ln Y_i = -\infty$.

$$C_{Y_iK} = \frac{\beta_{y_ik}}{Y_iK},$$

and

$$C_{KK} = \frac{1}{K^2}\left(2\beta_{k^2} - \beta_k - \beta_{wk}\ln W - 2\beta_k^2 \ln K - \sum_i \beta_{y_ik}(\ln Y_i - \omega_i)\right).$$

Plugging these into equation (16) and multiplying through by $Y_iY_jK^2$ we can see that the technology is non-joint if and only if

$$2\beta_{y_iy_j}\left(2\beta_{k^2} - \beta_k - \beta_{wk}\ln W - 2\beta_{k^2}\ln K - \sum_l \beta_{y_ll}(\ln Y_l - \omega_l)\right) - \beta_{y_ik}\beta_{y_jk} = 0 \quad (24)$$

Therefore, whenever capital interacts with outputs ($\beta_{y_ik}, \beta_{y_jk} \neq 0$), as is generally the case since capital allocations depend on the output vector, the technology is non-joint whenever at least one of β_{wk}, β_{k^2}, or any β_{y_ik} is non-zero as their respective terms are non-constant. If we instead approximated the log cost function these expression would become even more complicated, depending more generally on the first-order log-cost terms.[10] This exemplifies an important pitfall. Whilst the non-parametric framework can accommodate both joint and non-joint technologies, any finite approximation may be structurally joint.[11]

7 Conclusion

In this paper I derived a new non-parametric identification result for joint cost functions when the unobserved Hicks-neutral productivities are firm-product-time specific and enter the firm's cost function non-separably. I show that this generality is necessary when the researcher wants to stay agnostic about whether production is joint. A crucial implication is that this generality comes at the cost of non-parametric convergence rates and an exploding numbers of instruments, effectively leaving researchers

[10]The condition would then become

$$\left(2\beta_{y_iy_j} + \frac{\partial \ln C}{\partial \ln Y_i}\frac{\partial \ln C}{\partial \ln Y_J}\right)\left(2\beta_{k^2} - \frac{\partial \ln C}{\partial \ln K} + \left(\frac{\partial \ln C}{\partial \ln K}\right)^2\right) = \left(\beta_{y_ik} + \frac{\partial \ln C}{\partial \ln Y_i}\frac{\partial \ln C}{\partial \ln K}\right)\left(\beta_{y_jk} + \frac{\partial \ln C}{\partial \ln Y_j}\frac{\partial \ln C}{\partial \ln K}\right).$$

[11]Another way to look at this is that the approximation bias induces jointness mechanically.

to choose between estimating separable technologies, severely restricting non-jointness, or imposing non-jointness and estimating production functions.

I also derived testable conditions for non-jointness for production frontiers and joint cost functions. Furthermore, I showed the importance of incorporating production information for identifying returns to scope and scale when there are frictions in input markets.

Future work should extend the identification results to production frontiers and more general demand models. It is also important to understand how well the GMM moments recover parameters in finite sample. Lastly, some work should compare how well different finite sample approximations can capture non-joint production, testing at what percentage of observations the non-jointness conditions canot be rejected locally. More importantly, additional work needs to be dedicated to finding a sieve family that can globally capture jointness and to discussing the trade-offs researchers need to make when choosing a sieve family.

References

David Autor, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 2020.

J. William Baumol, John C. Panzar, and Robert D. Willig. *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt College Pub, 1982.

Steven T. Berry and Philip A. Haile. Identification in differentiated products markets using market level data. *Econometrica*, 2014.

John Cairncross, Peter Morrow, Scott Orr, and Swapnika Rachapalli. Multi-product markups. 2023.

Mauro Caselli, Arpita Chatterjee, and Shengyu Li. Productivity and quality of multi-product firms. *International Finance Discussion Papers 1430*, 2025.

CMA. Competition and market power in uk labour markets. 2024.

Jan De Loecker, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. Prices, markups, and trade reform. *Econometrica*, 84(2):445–510, 2016. ISSN 0012-9682. doi: 10.3982/ecta11042.

Robert E. Hall. The specification of technology with several kinds of output. *Journal of Political Economy*, 1973.

Bjarne S. Jensen. Transformation functions: Cost, revenue functions, autarky general equilibrium, comparative advantage. In *The Elements and Dynamic Systems of Economic Growth and Trade Models*, pages 697–765. Springer, 2025. doi: 10.1007/978-3-031-52493-$6_2$3.

Ekaterina Khmelnitskaya, Guillermo Marshall, and Scott Orr. Identifying scale and scope economies using product market data. 2025.

Andros Kourtellos, Thanasis Stengos, and Chih Ming Tan. Structural threshold regression. *Econometric Theory*, 2016.

Rosa Matzkin. Identification in nonparametric simultaneous equations models. *Econometrica*, 2008.

Scott Orr. Within-firm productivity dispersion: Estimates and implications. *Journal of Political Economy*, 2022. doi: .org/10.1086/720465.

Richard Shumway, Rulon D. Pope, and Elizabeth K. Nash. Allocatable fixed inputs and jointness in agricultural production: Implications for economic modeling. *American Journal of Agricultural Economics*, 1984.

Nelli Valmari. Estimating production functions of multiproduct firms. *Review of Economic Studies*, 90(6):3315–3342, January 2023. ISSN 1467-937X. doi: 10.1093/restud/rdad005.

Emmanuel Dhyne; Amil Petrin; Valerie Smeets; Frederic Warzynski. Theory for extending single-product production function estimation to multi-product settings. 2022WP.

A Shared tasks cost function with fixed inputs

B Own dominance of productivity

B.1 Own dominance implies non-vanishing Jacobian

In this subsection I will show that assumption 6 implies that the Jacobian determinant is non-vanishing everywhere. The proof follows directly from the Gershgorin circle theorem.

Let A be a complex $n \times n$ matrix, with entries $a_{i,j}$. Let $R_i = \sum_{j \neq i} |a_{i,j}|$ be the sum of absolute values of the non-diagonal entries in the i-th row. Let $D(a_{ii}, R_i) \subseteq C$ be a closed disc centered at a_{ii} with radius R_i.

Theorem B.1 (Gershgorin circle theorem). *Every eigenvalue of A lies within $\cup_i D(a_{ii}, R_i)$.*

Lemma 1. $\sum_{k \neq i} \left| \frac{\partial mc_k}{\partial \omega_i} \right| < \left| \frac{\partial mc_i}{\partial \omega_i} \right| \; \forall(\vec{y}, \vec{w}, \vec{K}, \vec{\omega}) \implies det \left(\frac{\partial mc(\vec{y}, \vec{w}, \vec{K}, \vec{\omega}}{\partial \omega} \right) \neq 0 \; \forall(\vec{y}, \vec{w}, \vec{K}, \vec{\omega})$

Proof. By assumption, for every $(\vec{y}, \vec{w}, \vec{K}, \vec{\omega})$,

$$\sum_{k \neq i} \left| \frac{\partial mc_k}{\partial \omega_i} \right| < \left| \frac{\partial mc_i}{\partial \omega_i} \right|$$

Therefore,

$$0 \notin \cup_i D \left(\frac{\partial mc_i}{\partial \omega_i}, \sum_{k \neq i} \left| \frac{\partial mc_k}{\partial \omega_i} \right| \right)$$

proving that $det \left(\frac{\partial mc(\vec{y}, \vec{w}, \vec{K}, \vec{\omega}}{\partial \omega} \right) \neq 0$ since the determinant is the product of all eigenvalues. $\square$

B.2 Invertibility of Cairncross et al. (2023)'s system

Cairncross et al. (2023) propose a CES hproduction frontier and show how it can be used to model non-joint technologies where, except for a product specific Hicks-neutral term, A_j, each product is produced with the same homogeneous production function, and joint production where the firm can first use some of its inputs to produce a non-rival intermediate good as in Chapter 15 of Baumol et al. (1982).

Definition 3. The CES frontier is given by

$$T(\vec{Y},\vec{X}) = \left(\sum_j (Y_j/A_j)^{\frac{1}{\phi\beta}}\right)^{\phi\beta} \Big/ F(\vec{X}), \tag{25}$$

where $F(\vec{X})$ is homogeneous of degree ϕ.

The input-output separability of the frontier has the advantage that evaluating the relative magnitude of marginal costs at a given point is equivalent to evaluating the derivative of the frontier with respect to the outputs and is independent of the input price vector.

Verifying the condition of assumption 6 is as simple as calculating the second order derivatives. Let $F := \sum_j (Y_j/A_j)^{\frac{1}{\phi\beta}}$, then

$$T_{Y_k\omega_i} = -\frac{\phi\beta - 1}{\phi\beta}\frac{F^{\phi\beta}}{e^{\omega_i}Y_k}, \text{ if } i \neq k$$

$$T_{Y_i\omega_i} = -\frac{F^{\phi\beta}}{e^{\omega_i}Y_i}.$$

Therefore,

$$\begin{aligned} &\sum_{k\neq i}\left|\frac{\partial mc_k}{\partial\omega_i}\right| < \left|\frac{\partial mc_i}{\partial\omega_i}\right| \\ \iff &\sum_{k\neq i}|T_{Y_k\omega_i}| < |T_{Y_i\omega_i}| \\ \iff &\frac{\phi\beta-1}{\phi\beta}\sum_{k\neq i}\frac{1}{Y_k} < \frac{1}{Y_i} \\ \iff &\phi\beta \leq 1 \end{aligned}$$

the cost function resulting from the frontier satisfies the condition if and only if the output function has at most constant returns to scale.

Cairncross et al. (2023) show that the technology is non-joint if and only if $\beta = 1$. This means that the condition is satisfied if the production technology is at most constant returns to scale.