# Who Can Produce What? Firm Capabilities and Product Entry

Luca Macedoni[a], John Morrow[b], Vladimir Tyazhelnikov[c]

September 2026

**Abstract**

*Which* firms supply *which* products? Modeling multi-product firms under heterogeneous markups, we recover firms' costs for produced products and develop an algorithm inspired by recommender systems to predict costs for unproduced products. Lower predicted costs imply increased product adoption and identify which firms enter when export demand rises, including firms crossing broad product classifications. We use these counterfactual costs to quantify how reductions in fixed and marginal costs induce product entry and lower price indices. In a price-stabilization exercise following a cost shock, entry subsidies can cost less than subsidies to incumbents, particularly for smaller shocks when capable potential entrants remain available.

**JEL Codes:** F1, F6, D2, L1.
**Keywords:** Multi-product firms, firm capabilities, classification, industrial policy.

---

*Acknowledgments.* We thank Nick de Roos for early suggestions and Axel Roepstorff and Kristian Maretty Højen for excellent research assistance. We are grateful for comments from Kirill Borusyak, Magnus Tolum Buus, Wanyu Chung, Andreas Hafler, Luca Lodi, Guzman Ourens, Catherine Thomas, and participants at the 1st SICKCL, Economics of Global Interactions, KCL Seminars, ETSG, LSE Trade Workshop, Kent Macro Workshop, Australasian Trade Workshop, WEAI, Nottingham, Zurich ETH, Paris Trade & Macro Workshop, Asia-Pacific Trade Seminar, PSE/Sciences Po, InsTED, Lille, Organisational Economics Workshop, UC Louvain, Tilburg GIFT Conference, BCAM Macro, CEP Retreat, ESCOE, ITSG, Workshop on the Impact of Global Trade Shocks, Firms/Workers/Inst Linz, RES, SMU, INSE at Peking University, Universitat Pompeu Fabra, HKS Growth Lab, NYU Trade Day, CUHK Trade Workshop, Geneva Grad Institute, EUI Trade and Development, QMUL, ULB, Maastricht, University of Verona, the Competition and Markets Authority, SPRU, University of Pavia, University of Milan, UNSW, Aston Business School, and University of Zurich. An earlier version of this project appeared as "Firms in Product Space."

[a]University of Milan. Financial support from the Carlsberg Foundation is gratefully acknowledged.
[b]Queen Mary University of London, CEP, and CEPR. Corresponding Author: john.morrow@qmul.ac.uk.
[c]The University of Sydney.

# 1 Introduction

Which products are potentially produced together? When demand for a product increases, can we predict which firms will supply it? This is important for policy as knowing *which* products a firm is likely to adopt is central in the current policy push toward national self-reliance in key sectors, ranging from "green" technologies to microprocessors to defense. In the United States, the CHIPS incentives subsidize investments in semiconductor production and supply chains, while clean-energy policy combines production-based support, such as the Advanced Manufacturing Production Credit, with investment support through the Advanced Energy Project Credit. In Europe, the Chips Act and Important Projects of Common European Interest similarly use public support to develop strategic production capabilities. A central problem in designing such policies is not only which products to support, but *which firms* could become suppliers. In particular, firms with the capabilities to enter a targeted product may currently operate outside its conventional industry classification. Designing and evaluating these policies therefore requires knowing which products domestic firms can realistically produce and what kinds of support would be required to induce production.

We build a framework that provides firm-level measures of counterfactual productive capabilities and allows policymakers to quantify the feasibility and welfare implications of industrial policy interventions. By predicting which firms have the capabilities to produce critical goods or pointing out limited capability, the model can inform targeted policies and can help distinguish sectors where domestic supply is feasible from those where it would require substantial capability accumulation.

A key obstacle is that until a firm produces a product, costs and benefits of this decision are unknown, at least outside of the firm. Our central idea is that the sales, markup and cost distribution of a firm's existing product portfolio, together with the distribution of co-occurring absolute advantages across all firms, contains valuable information about the firm's absolute advantage in unproduced products. Much like streaming or e-commerce platforms that predict a user's next choice by combining individual histories with aggregate patterns across users, we draw on the computer science recommender-system literature to infer a firm's absolute advantage in *unproduced* products based on observed co-production patterns within and across firms. This idea aligns with the multi-product firm literature, which emphasizes the role of core competencies and supply-side linkages across products in shaping firms' production portfolios (Eckel and Neary, 2010; Nocke and Yeaple, 2014; Mayer et al., 2014; Eckel et al., 2023). One challenge is that predicting product adoption conditional on all possible combinations of co-production would require unrealistically rich data. While such data demands might be met by online

platforms with millions of users, they typically exceed what standard production surveys in economics can provide.

We address this challenge by constructing a multi-dimensional *product space* from observed firm-product absolute advantages. In this space, both products and firms are represented as points. The distance between two products summarizes how efficiently they tend to be co-produced across firms when co-produced. This implies that when a firm produces a given product efficiently, it is more likely, based on the behavior of other firms, to also produce products that are nearby in the product space. In analogy to streaming platforms, two movies are "close" if they tend to be watched and rated highly by the same users, so that watching one predicts watching the other.

A natural approach to think about product co-production is to use product characteristics (e.g. inputs, technologies, or end uses) when they are observable, which is often the case for *demand characteristics* in narrow industries for final goods. In this paper, we wish to understand *latent supply capabilities* in co-production across all (or most) products at a granular level. We therefore adopt a more parsimonious strategy and infer them directly from revealed co-production efficiency patterns, without committing to any single underlying mechanism.[1] The practical benefit is that our approach is less data demanding: comprehensive product characteristics are often unavailable at high levels of disaggregation and difficult to define consistently across very different products, whereas co-production patterns and absolute advantages remain observable.

The product space we construct therefore holds information about co-production efficiency in the location of products. We locate each firm as a sum of the products they produce as in the recommender system literature. Once products and then firms are located, the distance between a firm and any unproduced product provides a counterfactual measure of the firm's absolute advantage in that product. These inferred advantages form the basis for our empirical analysis.

We combine this approach with a heterogeneous firm model under monopolistic competition with CoPaTh demand (Matsuyama and Ushchev, 2017), which allows costs and markups to vary across firms and products. In the model, product adoption depends on observable factors such as market size and firm productivity, but crucially also on a firm's product-specific absolute advantage, which governs marginal costs and therefore equilibrium prices. Interpreting our inferred counterfactual advantages as model primitives enables counterfactual exercises, such as analyzing how shocks affect prices and welfare through the extensive margin of product scope.

[1] This is fully compatible with a characteristics-based approach as observable characteristics can be viewed as micro-foundations for the latent capabilities we recover, and different underlying mechanisms can be observationally equivalent. For our purposes, efficient co-production is a sufficient statistic to construct counterfactual absolute advantages.

We apply our algorithm to data from the Danish production survey, which covers all manufacturing firms with at least ten employees and reports production at the product level, which we define as an 8-digit Combined Nomenclature (CN) code-unit. We link these data to additional firm-level administrative registers, including customs records. To recover observed absolute advantages for produced products, we consider two complementary approaches: one based on inverting demand and constant markups, and one based on production function estimation following De Loecker et al. (2016) with heterogeneous markups. Notably, the model rationalizes heterogeneous markups at the firm-product level, allowing us to construct counterfactual variable profits at a highly granular level. Results across the measures are qualitatively similar but quantitatively distinct.

Applying our algorithm, we first recover pairwise distances between products and then distances between firms and products that capture production complementarities spanning broad sectors in contrast to the mixed rationales behind standard classifications.[2] These distances reveal economically sensible patterns. For example, our classification reveals bottled beer is close to soft drinks, even though they belong to different 4-digit CN codes (alcohol is the distinction), and sweet syrups, which are in different 2-digit CN codes to sauces (sugar is the distinction) but are close in our classification. Such patterns reflect explicit similarities in production capabilities and help explain why only about half of product adoptions are in a firm's best selling 2-digit CN code.[3]

We empirically validate our approach by testing two predictions about *which* firms produce *which* products. While shocks, such as changes in productivity, shifts in demand or trade policy, are often central to models of firm behavior, much less is known about *which* firms are most likely to respond along the extensive margin of products. We show that firms are significantly more likely to begin producing products in which they have stronger counterfactual absolute advantage. Importantly, our continuous measure retains predictive power within discrete Combined Nomenclature categories. For further validation, we combine our measures of counterfactual advantage with plausibly exogenous variation in product-level demand (Hummels et al., 2014), finding that stronger counterfactual absolute advantage predicts product adoption.

We then use our counterfactual measures to quantify product entry and price-index responses to reductions in potential entrants' fixed and marginal costs. These exercises reveal how firms' latent capabilities shape the scope for expanding supply. Marginal cost reductions generate stronger entry responses under constant markups, yet slightly larger

[2]These classifications can be both haphazard and endogenous to policy objectives. See the examples of Jacobs and ONeill (2003) and Grant (2023).

[3]This continuous classification provides a flexible measure of supply relatedness between products, industries, or activities that can be used beyond the applications of this paper. We provide the classification at the 4-digit level at this webpage.

price-index reductions under heterogeneous markups. Fixed cost reductions induce entry under both specifications, with substantially larger price-index reductions under heterogeneous markups. Entry responses also vary with the size of the cost reduction, rather than exhibiting the constant elasticities assumed in standard theoretical benchmarks.

Finally, we build on these entry responses to compare alternative policies for stabilizing prices following an increase in incumbent firms' marginal costs. We ask whether subsidizing the fixed costs of product entry can achieve the same price-index target at a lower fiscal cost than subsidizing incumbent production. The framework predicts counterfactual profits, markups, entry and price-index effects, allowing us to compare the cost of policies that subsidize incumbent production with entry subsidies that reduce prices through competition. We find that entry subsidies are often a cheaper way to stabilize prices, particularly following smaller cost shocks. As shocks become larger, the share of products for which entry is cheaper declines, consistent with exhausting the pool of profitable potential entrants. Under heterogeneous markups, entry subsidies are preferred for fewer products than under constant markups, but those products account for a larger share of sales. More broadly, the exercise shows that the product space can inform policy along a margin that standard industry classifications miss: the set of existing firms with latent capabilities to enter a targeted product market.

This section continues with a literature review. Section 2 presents the theoretical model. Section 3 introduces a method to recover latent absolute advantages. Section 4 describes the data and estimation strategy. Section 5 empirically tests the model's predictive power for product adoption. Section 6 uses the counterfactuals of the model to evaluate the extensive-margin responses to cost shocks and policies to stabilize prices. Section 7 concludes.

### Literature Review

There has been a rapid expansion of research on multi-product firms, especially in the context of international trade.[4] For the typical model of this literature, a firm is a collection of product varieties, which may be linked by supply or demand linkages. Supply linkages arise from flexible manufacturing, economies and dis-economies of scope, and the presence of core competency and non-core competency products (Eckel and Neary, 2010; Nocke and Yeaple, 2014; Mayer et al., 2014; Eckel et al., 2015; Arkolakis et al., 2021; Macedoni and Xu, 2022). While the literature has established there are "significant departures from the theoretical benchmark of core competencies" (Fontagne et al., 2018), progress in formalizing and quantifying these departures has remained limited. On the demand side, linkages mainly include cannibalization effects and demand com-

[4]For details, see the review by Irlacher (2024).

plementarities (Feenstra and Ma, 2007; Eckel and Neary, 2010; Dhingra, 2013; Bernard et al., 2018; Flach and Irlacher, 2018; Macedoni, 2022). However, the role of particular product bundles in shaping supply-side dynamics and firm growth remains understudied. A contribution of our approach is to show that considering *which* product bundles a firm produces relative to *all other firms* matters for product adoption and growth.

The literature on co-production (Bernard et al., 2010; Goldberg et al., 2010) shows that some pairs of products are often produced together, while others are almost never produced by the same firm. In this approach, when a firm expands its product range, it will likely choose products that are often co-produced within other firms, conditional on their current product mix, for a wide variety of possible explanations for linkages across inputs and outputs of firms (Boehm et al., 2022; Dhyne et al., 2022; Ding, 2023; Jakel et al., 2023). We build on such co-production concepts and in our approach, a firm is more likely to expand into products that are closer to its existing capabilities. Here, closeness is determined not only by the frequency of co-production across products, but also by the efficiency with which firms co-produce them.

The concept of a product space and importance of product complexity was popularized by Hidalgo et al. (2007), Hausmann and Klinger (2006), Hausmann et al. (2007), and Hausmann and Hidalgo (2011), who define product distance based on the co-occurrence of Revealed Comparative Advantage (RCA) across countries.[5] A key insight from this literature is that proximity to products helps explain cross-country differences in income growth.[6] Instead of relying on RCA, we construct product distances from firm-level absolute advantages, yielding a product space micro-founded with monopolistic competition models and econometric estimation of costs and markups. This allows us to predict product adoption even when co-production is sparse. This also addresses the firm-product sparsity limitation of RCA-based approaches which results in the majority of combinations showing zero co-occurrence of RCA across firms.

Our approach to the product space deliberately departs from methods based on product characteristics, which underlie models of hedonic demand (Lancaster, 1966; Rosen, 1974; Berry et al., 1995; Feenstra and Levinsohn, 1995). For example, in Pellegrino (2025), firms are positioned in a characteristics space, where proximity reflects the degree of overlap in product attributes. In contrast, our framework can be interpreted as the *dual* of this characteristics-based approach: here, proximity between products reflects similarities on the supply side rather than the demand side. Specifically, two products are

[5]The approach has since evolved (Hidalgo, 2021), with applications ranging from research fields (Guevara et al., 2016) to pioneer firm survival (Jara-Figueroa et al., 2018) and industry transitions (Neffke et al., 2011), and more recently to green products (Mealy and Teytelboym, 2022).

[6]Hidalgo et al. (2007) show that countries tend to develop RCA in goods close to those they already specialize in, facilitating faster export upgrading. This mirrors our finding that firms tend to adopt products with higher counterfactual absolute advantage.

close if firms frequently exhibit joint efficiency in producing them. This method enables us to analyze a broad set of otherwise incomparable products and requires minimal data: in the simplest case, firm-level sales data are sufficient.

Our focus on firms in a product space links our paper to an expanding body of research that uses various measures of similarity *between firms* for different applications. For example, in the R&D literature, spillovers across firms depend on the proximity between technologies and products across firms. A common method in this field involves calculating the overlap between firms' technology classes, as indicated by patents, and the overlap between their product sales (Jaffe, 1986; Bloom et al., 2013).[7]

A growing literature uses text analysis to develop new classifications of firm activities and outputs. Hoberg and Phillips (2016) create firm locations from word vectors of SEC filings, in which firms have new relative locations each year based on cosine similarity measures which generally better explain profitability and growth than SIC or NAICS classifications. They show that, following a negative demand shock, firms either relocate toward areas of high common demand or differentiate by reducing product similarity, underscoring that the relative similarity of product portfolios is an important determinant of firm dynamics. In contrast to their methodology, we generate a new production-based classification with standard product classification codes common to administrative data. Existing crosswalks then allow us to link to many potential data sets and policy changes as used in the trade literature.

We next lay out the model framework which we will use to motivate the empirical measures, estimating equations and counterfactual exercises in the rest of the paper.

# 2 Multi-Product Firms and Product Adoption

This section presents a model of multi-product firms engaged in monopolistic competition using the Constant Pass Through (CoPaTh) setting of Matsuyama and Ushchev (2020) which nests the well known CES demand system. The novel feature we leverage here is the presence of markups that vary within products across firms and within firms across varieties. However, markups are constant at the firm-variety level as they do not depend on other economic conditions such as the toughness of competition. Firms absolute advantages vary by product and drive product adoption, which additionally depends on variables that can be estimated with fixed effects. We derive product adoption estimating equations that motivate our construction of the product space to predict counterfactual absolute advantages on *unproduced* products.

[7] Escolar et al. (2023) provide a summary of the approaches used in the literature for locating firms using patents data.

## 2.1 Consumers and Firms

There is a discrete set of differentiated products indexed by $\nu$. Each product $\nu$ is produced by a continuum of firms indexed by $\omega$, each offering a horizontally differentiated variety. The pair $(\omega, \nu)$ denotes firm $\omega$'s variety of product $\nu$.

Consumers have nested preferences and aggregate income $I$. At the top level, consumption is aggregated across products $\nu$ with Cobb-Douglas preferences. Within each nest, preferences are of the CoPaTh class introduced by Matsuyama and Ushchev (2020).[8] Let $q_{\omega,\nu,t}$ and $p_{\omega,\nu,t}$ denote the quantity and price of variety $(\omega, \nu)$ at time $t$, and $\boldsymbol{q}_{\nu,t}$ and $\boldsymbol{p}_{\nu,t}$ denote their respective vectors. The CoPaTh aggregator over consumption bundles is denoted $Q(\boldsymbol{q}_{\nu,t})$, wherein variety $(\omega, \nu)$ has an elasticity of substitution $\sigma_{\omega,\nu}$.

The first layer of the utility function is

$$U\left(\left\{\boldsymbol{q}_{\nu,t}\right\}\right) = \sum_{\nu} \psi_{\nu} \ln Q\left(\boldsymbol{q}_{\nu,t}\right) \quad \text{with} \quad \sum \psi_{\nu} = 1,$$

where $\psi_{\nu}$ are the constant, Cobb-Douglas expenditure shares. The total revenues generated by product $\nu$ satisfy $R_{\nu} = \psi_{\nu} I$. Consumers choose $\{q_{\omega,\nu,t}\}$ to maximize utility subject to the budget constraint

$$\max_{q_{\omega,\nu,t}} U\left(\left\{\boldsymbol{q}_{\nu,t}\right\}\right) \quad \text{subject to} \quad I = \sum_{\nu} \int_{\omega \in \Omega_{\nu,t}} p_{\omega,\nu,t} q_{\omega,\nu,t} d\omega.$$

A key distinction between CoPaTh and CES preferences is the presence of two different price indices. For the price vector $\boldsymbol{p}_{\nu,t}$, the ideal price index $P(\boldsymbol{p}_{\nu,t})$ captures the cost of one unit of utility. The aggregator $A(\boldsymbol{p}_{\nu,t})$, in contrast, enters the expression for individual firm demand and determines expenditure shares. These two indices generally differ, and coincide only under CES preferences when $\sigma_{\omega,\nu} = \sigma_{\nu}$ for all $\omega$.

Let $s_{\omega,\nu,t}$ denote the expenditure share on variety $(\omega, \nu)$ at time $t$. The CoPaTh structure implies

$$s_{\omega,\nu,t} = \left(p_{\omega,\nu,t} / A\left(\boldsymbol{p}_{\boldsymbol{\nu},\boldsymbol{t}}\right)\right)^{1-\sigma_{\omega,\nu}}. \tag{1}$$

The aggregator $A(\boldsymbol{p}_{\nu,t})$ is implicitly defined by the following equation

$$\int_{\Omega_{\nu}} \left(\frac{p_{\omega,\nu,t}}{A(\boldsymbol{p}_{\nu,t})}\right)^{1-\sigma_{\omega,\nu}} d\omega = 1, \tag{2}$$

[8] Matsuyama and Ushchev (2020) provide a general framework for homothetic demand systems. A notable subclass is the Homothetic Single Aggregator family, which exhibits constant price elasticities and therefore constant markups at the variety level which we use here. For details, see the original paper.

which is obtained by integrating over the shares in Equation (1). Since the left hand side of Equation (2) is increasing in $A$, this equation admits a unique solution for $A\left(\boldsymbol{p_{\nu,t}}\right)$.

The ideal price index $P\left(\boldsymbol{p_{\nu,t}}\right)$ is given by

$$P\left(\boldsymbol{p_{\nu,t}}\right) = A\left(\boldsymbol{p_{\nu,t}}\right) \exp\left(\int_{\Omega\nu} \frac{1}{1-\sigma_{\omega,\nu}} \left(\frac{p_{\omega,\nu,t}}{A\left(\boldsymbol{p_{\nu,t}}\right)}\right)^{1-\sigma_{\omega,\nu}} d\omega\right). \tag{3}$$

Under CES preferences with $\sigma_{\omega,\nu} = \sigma$, this expression reduces to

$$P\left(\boldsymbol{p}_{\nu,t}\right) = \left(\int_{\Omega_\nu} p_{\omega,\nu,t}^{1-\sigma} d\omega\right)^{1/(1-\sigma)} \exp^{1/(1-\sigma)} = A\left(\boldsymbol{p_{\nu,t}}\right) \exp^{1/(1-\sigma)}, \tag{4}$$

which is the familiar CES price index times a constant. Firm $\omega$ can produce product $\nu$ at a fixed cost $f_{\omega,\nu}$ and constant marginal cost $m_{\omega,\nu,t}$, which may vary over time.[9] Conditional on producing $\nu$, firm $\omega$ sets the price

$$p_{\omega,\nu,t} = \frac{\sigma_{\omega,\nu}}{\sigma_{\omega,\nu}-1} m_{\omega,\nu,t} = \mu_{\omega,\nu} m_{\omega,\nu,t}, \tag{5}$$

where $\mu_{\omega,\nu} \equiv \sigma_{\omega,\nu}/(\sigma_{\omega,\nu}-1)$ are defined as markups and $\boldsymbol{\mu_\nu}$ the vector of markups for each product $\nu$. Markups can vary across varieties within a firm and across firms, depending on $\sigma_{\omega,\nu}$.

Let $\overline{m}_{\nu,t}$ denote the average cost of actually produced varieties $\nu$ across firms at time $t$, and define the firm's Absolute Advantage Index as

$$\text{Absolute Advantage Index:} \qquad a_{\omega,\nu,t} \equiv \frac{m_{\omega,\nu,t}}{\overline{m}_{\nu,t}}. \tag{6}$$

This measure is invariant to the choice of units for each product $\nu$. The index represents the cost for a firm to produce a variety as the percentage of the product average. The lower $a_{\omega,\nu,t}$, the better a firm $\omega$ is in making product $\nu$, so it is an *inverse measure of absolute advantage*. We refer to $a_{\omega,\nu,t}$ as advantage for brevity, with $\boldsymbol{a_{\nu,t}}$ as the vector of advantages for product $\nu$ at time $t$.

Note from Equation (5) that $\boldsymbol{p}_{\nu,t}/\overline{m}_{\nu,t} = \boldsymbol{\mu}_\nu \odot \boldsymbol{a}_{\nu,t}$ where $\odot$ is component-wise multiplication. Then using the fact that $A$ is homogeneous of degree one, the revenue and profit of firm $\omega$ from producing $\nu$ at time $t$ are

$$r_{\omega,\nu,t} = \left(\frac{\mu_{\omega,\nu} a_{\omega,\nu,t}}{A(\boldsymbol{\mu}_\nu \odot \boldsymbol{a}_{\nu,t})}\right)^{1-\sigma_{\omega,\nu}} R_\nu, \qquad \pi_{\omega,\nu,t} = \frac{r_{\omega,\nu,t}}{\sigma_{\omega,\nu}}. \tag{7}$$

[9] Matsuyama and Ushchev (2020) identify each variety with a firm, but we generalize to the multi-product case with minimal changes. We assume a fixed number of firms without modeling entry. In the empirical section, we allow the set of firms and costs to evolve as in the data.

## 2.2 Product Adoption Equations

We next derive an estimating equation for product adoption. For products a firm already produces, the advantage $a_{\omega,\nu,t}$ defined in Equation (6) is observed in levels. For an unproduced product, however, this object is unobserved. Our approach below recovers a counterfactual advantage $\tilde{a}_{\omega,\nu,t}$ from the firm's location in the product space. As the product space identifies a firm's capabilities across products relative to its overall efficiency, we write the corresponding counterfactual advantage in levels as

$$a^{cf}_{\omega,\nu,t} \equiv \frac{\tilde{a}_{\omega,\nu,t}}{\varphi_{\omega,t}}, \tag{8}$$

where $\varphi_{\omega,t} > 0$ is a firm-wide efficiency shifter common across products. Higher $\varphi_{\omega,t}$ corresponds to greater firm-level efficiency and therefore a lower counterfactual marginal cost for any given $\tilde{a}_{\omega,\nu,t}$. This decomposition separates the relative firm–product capabilities that we recover below from differences in firms' overall efficiency.

Letting $A_{\nu,t} \equiv A(\boldsymbol{\mu}_\nu \odot \boldsymbol{a}_{\nu,t})$ and using Equation (7), the counterfactual variable profits from producing an unproduced product $\nu$ are therefore

$$\pi^{cf}_{\omega,\nu,t} = \frac{R_\nu}{\sigma_{\omega,\nu}} \left( \frac{\mu_{\omega,\nu}\tilde{a}_{\omega,\nu,t}}{\varphi_{\omega,t}A_{\nu,t}} \right)^{1-\sigma_{\omega,\nu}}. \tag{9}$$

The event that a firm adopts product $\nu$, $\text{Intro}_{\omega,\nu,t} > 0$, occurs when $\pi^{cf}_{\omega,\nu,t} > f_{\omega,\nu}$. Taking logs and rearranging, product adoption occurs when

$$-\ln \tilde{a}_{\omega,\nu,t} + \ln \varphi_{\omega,t} + \ln A_{\nu,t} > \frac{1}{\sigma_{\omega,\nu}-1} \ln \frac{\sigma_{\omega,\nu} f_{\omega,\nu}}{R_\nu} + \ln \frac{\sigma_{\omega,\nu}}{\sigma_{\omega,\nu}-1}. \tag{10}$$

Equation (10) forms the basis for estimating product adoption probabilities using firm-level production data. Introducing an error term $\epsilon_{\omega,\nu,t}$ and grouping terms, the adoption condition becomes

$$-\ln \tilde{a}_{\omega,\nu,t} + \underbrace{\ln \varphi_{\omega,t}}_{\text{Firm-Time}} + \underbrace{\ln A_{\nu,t}}_{\text{Product-Time}} - \underbrace{\left( \frac{\ln\left(\sigma_{\omega,\nu} f_{\omega,\nu}/R_\nu\right)}{\sigma_{\omega,\nu}-1} + \ln \frac{\sigma_{\omega,\nu}}{\sigma_{\omega,\nu}-1} \right)}_{\text{Firm-Product}} + \epsilon_{\omega,\nu,t} > 0. \tag{11}$$

Equation (11) can be estimated with a linear probability model and firm-time, product-time and firm-product fixed effects. Firm-time fixed effects absorb the firm-wide efficiency term $\ln \varphi_{\omega,t}$ together with other time-varying firm-level shocks. Product-time fixed effects absorb changes in product-specific market conditions, here summarized by the aggregator $A_{\nu,t}$. Firm-product fixed effects control for components that are fixed within a firm–product pair over time, including the fixed cost $f_{\omega,\nu}$ and demand elasticity $\sigma_{\omega,\nu}$.

The inclusion of firm-time fixed effects implies that identification relies on within-firm variation across products, and therefore requires a sample of multi-product firms.

The remaining obstacle is that $\tilde{a}_{\omega,\nu,t}$ is not observed for products the firm does not produce. The next section develops a method to recover these counterfactual relative advantages from observed co-production patterns. For the product adoption exercises, their relative values are sufficient because $\varphi_{\omega,t}$ is absorbed by firm-time fixed effects. For the quantitative counterfactuals in Section 6, we instead require counterfactual advantages in levels and therefore recover $\varphi_{\omega,t}$ to implement the mapping in Equation (8).

# 3 Product Spaces and Counterfactual Advantages

This section establishes a link between observed firms' behaviors and their advantages for produced and unproduced products. Conditional on observing revenues, prices, and demand elasticities, we can use the model above to back out the set of advantages for produced products. Our goal is to map these advantages to counterfactual advantages for unproduced products. This mapping could contain arbitrarily rich economies of scope from co-production.[10] We reduce the dimensionality of this problem by representing each of $N$ products as locational vectors wherein distance between products represents the joint efficiency of co-producing them. This process necessarily loses information in the representation but nonetheless provides a granular set of counterfactual absolute advantages for individual firm-product pairs that we later validate.

Our algorithm adapts a method from the computer science literature, namely recommender systems, popularized by online platforms that provide purchase or viewing recommendations. In our context, we apply these ideas to a product space, which we define as a vector space with many dimensions as the number of products. In the space, each product $\nu$ is represented by a vector $\ell_\nu$ and each firm is represented by a vector $\ell_\omega$. A recommender system might recover user (firm) and item (product) locations simultaneously to explain consumer ratings (advantages), $a_{\omega,\nu}$, by fitting product and firm vectors to solve

$$\min_{\ell_\omega,\ell_\nu} \sum_{\omega,\nu} \left(a_{\omega,\nu} - \ell_\omega \cdot \ell_\nu\right)^2 .$$

The solution locates each firm as a weighted linear combination of its product vectors. While this approach is popular as it can be computed efficiently in parallel[11], as we have

[10] For instance, suppose that a firm's ability to produce a new product depends on the set of products it currently produces. For instance, whether a firm produces product 1 and product 2, only 1, or only 2, or neither, affects its likelihood of adopting product 3. Since for $N$ products there are $2^N$ possible combinations of products, estimating the effect of extensive margin combinations would impose unrealistic data requirements.

[11] See for instance Koren et al. (2009); Bengio et al. (2014).

access to a single confidential data server, we take a related but different approach in three steps.

First, similar to collaborative filtering (Sarwar et al., 2001), we construct a set of product vectors. We use observed co-production to construct a pairwise measure of joint efficiency between co-produced products, $d_{\nu,\tau}$, for products $\nu$ and $\tau$, where a smaller value implies stronger joint production advantage. We then construct product vectors, say $\ell_\nu$ and $\ell_\tau$, such that the Euclidean distance between these vectors approximately matches the pairwise distance, $|\ell_\nu - \ell_\tau| \approx d_{\nu,\tau}$.

In the second step, we locate each firm $\omega$ by a vector $\ell_\omega$, defined as a convex combination of the vectors corresponding to the products it produces. The weights are higher for products in which a firm has stronger observed advantage. Subsequently we separately estimate or recover a firm-level productivity term, so it is this relative location within the product space that we require.

In the third step, having located both products and firms in the product space, we use the distance between a firm and a product it does not produce to recover a counterfactual measure of its advantage in that product. Products further from the firm imply weaker counterfactual advantage. These inferred advantages provide counterfactual values for every firm-product pair and form the basis for the subsequent empirical analysis.

## 3.1 Measuring Efficiency in Co-production

To address dimensionality challenges in our analysis and to increase precision recovering a highly sparse product space, we divide our sample into sectors based on product codes.[12] The construction of product spaces is applicable only to a set of products that are either co-produced or linked by chains of co-production and we call such a set of products a *cluster*. We will further refine our selection of products in the next section to belong to a cluster so we will assume this is true for the construction we now detail.

Our goal is to obtain distances between all products within a cluster, including products that are not directly co-produced by any firm. We first construct distance estimates for directly co-produced product pairs. Let $T(a_{\omega,\nu})$ denote an increasing transformation of firm $\omega$'s advantage in product $\nu$, with $T(0) = 0$. Suppose products $\nu$ and $\tau$ are both produced by firm $\omega$, and interpret $T(a_{\omega,\nu})$ and $T(a_{\omega,\tau})$ as distances from the firm. The triangle and reverse triangle inequalities imply the firm-specific bounds

[12] We begin with the following 15 sectors: Animal products (CN 2-digit 01-05), Vegetable products (06-15), Foodstuffs (16-24), Mineral products (25-27), Chemical products (28-38), Plastics and rubber (39-40), Leather and Fur (41-43), Wood products (44-49), Textiles (50-63), Footwear and headgear (64-67), Stone and glass (68-71), Metals (72-83), Machinery and electrical (84-85), Transportation (86-89), and Miscellaneous (90-97). We then drop any sector with less than 10 products, to be consistent with our model of monopolistic competition with differentiated goods, which leaves us with 12 sectors.

$$|T(a_{\omega,\nu}) - T(a_{\omega,\tau})| \leq d_{\nu,\tau} \leq T(a_{\omega,\nu}) + T(a_{\omega,\tau}). \tag{12}$$

For directly co-produced products, we average these bounds across the $N_{\nu,\tau}$ firms producing both products:

$$\underline{d}^{dir}_{\nu,\tau} \equiv \frac{1}{N_{\nu,\tau}} \sum_{\omega} |T(a_{\omega,\nu}) - T(a_{\omega,\tau})|, \tag{13}$$

and

$$\overline{d}^{dir}_{\nu,\tau} \equiv \frac{1}{N_{\nu,\tau}} \sum_{\omega} [T(a_{\omega,\nu}) + T(a_{\omega,\tau})]. \tag{14}$$

We take the midpoint of these average bounds as the initial distance estimate for directly co-produced products:

$$D_{\nu,\tau} \equiv \frac{\underline{d}^{dir}_{\nu,\tau} + \overline{d}^{dir}_{\nu,\tau}}{2}. \tag{15}$$

We then use the available entries of the distance matrix $D$ to infer distances for product pairs that are not directly co-produced. Let $K_{\nu,\tau}$ denote the set of intermediary products $k$ for which both $D_{\nu,k}$ and $D_{k,\tau}$ are available. If these pairwise distances were generated by a metric, then for every $k \in K_{\nu,\tau}$ the triangle inequalities would imply the candidate interval

$$|D_{\nu,k} - D_{k,\tau}| \leq d_{\nu,\tau} \leq D_{\nu,k} + D_{k,\tau}. \tag{16}$$

When several intermediary products connect $\nu$ and $\tau$, we combine the resulting restrictions by taking the greatest candidate lower bound and the smallest candidate upper bound:

$$\underline{d}_{\nu,\tau} \equiv \max_{k \in K_{\nu,\tau}} |D_{\nu,k} - D_{k,\tau}|, \tag{17}$$

and

$$\overline{d}_{\nu,\tau} \equiv \min_{k \in K_{\nu,\tau}} (D_{\nu,k} + D_{k,\tau}). \tag{18}$$

If the observed pairwise distances were generated by a metric, these expressions would provide, respectively, lower and upper bounds on the missing distance. In our application, the entries of $D$ are empirical distance estimates, so we interpret them as metric-implied candidate bounds rather than exact bounds on an underlying true distance. These candidate bounds can in principle cross. Finally, we define the distance between two products

as the midpoint of the two candidate values:

$$d_{\nu,\tau} \equiv \frac{\underline{d}_{\nu,\tau} + \overline{d}_{\nu,\tau}}{2}. \tag{19}$$

After filling distances that can be connected through one intermediary, we augment the distance matrix with the newly imputed values and repeat the procedure. Successive iterations therefore recover distances between products connected through increasingly long chains of co-production. Since both the directly constructed and subsequently imputed entries are empirical estimates rather than exact metric distances, the completed distance matrix is not guaranteed to satisfy every metric restriction or to admit an exact representation in Euclidean space.

We allow for a different transformation $T(a_{\omega,\nu})$ to reduce sensitivity to extreme values. We consider three, with the logarithmic transformation as our baseline specification:

1. **Log (Baseline):** $T(a_{\omega,\nu}) \equiv \log(1 + a_{\omega,\nu})$. Adding one ensures that the transformation remains non-negative, including when $a_{\omega,\nu} < 1$.

2. **Level:** $T(a_{\omega,\nu}) \equiv a_{\omega,\nu}$. Here, we use advantage directly.

3. **Inverse Hyperbolic Sine:** $T(a_{\omega,\nu}) \equiv \log\left(a_{\omega,\nu} + \sqrt{1 + a_{\omega,\nu}^2}\right)$. This provides an alternative monotonic transformation that behaves similarly to the logarithm for large values while remaining well defined near zero (see Chen and Roth (2024)).

In each case, we first transform observed firm-product advantages using $T(\cdot)$ and then apply the same procedure above to construct pairwise product distances throughout the cluster. We now proceed to our second step.

## 3.2 Assigning Product Locations

Given a complete set of pairwise distances between products in each cluster $c$, we represent these distances as a collection of vectors the same dimension as the number of products representing product locations, which we define as a product space $\text{PS}_c$. We compute these for each period so that products and firms can move over time.

We now generate product vectors to reflect the pairwise distances above.[13] We start by arbitrarily numbering products, and for each product $k$ we set all but the first $k-1$ coordinates of its location vector $\ell_k$ to zero. Therefore we locate product 1 as $\ell_1 = 0$ at the origin. For the second product, we choose the first coordinate equal to the distance between products 1 and 2: $d_{1,2} = \|\ell_1 - \ell_2\|$. For the third product, we now face the possibility that

[13] This step is similar to a Cholesky decomposition in linear algebra, however we require approximation to fit the final vector space.

the empirical joint efficiency measures do not exactly reflect the geometry of Euclidean space. Therefore we find a product vector $\ell_3$ to approximately satisfy $d_{2,3} = \|\ell_2 - \ell_3\|$ and $d_{1,3} = \|\ell_1 - \ell_3\|$. To do so, we fit the first two coordinates of the product vector $\ell_3$ to be close to each pairwise distance by solving $\min_{\ell_3} \sum_{i=1}^{2} \left(\|\ell_i - \ell_3\| / d_{i,3} - 1\right)^2$ at each step. We iterate this procedure for each product $k$, finding its first $k-1$ coordinates to approximate distances to each product $i < k$: $d_{i,k} = \|\ell_i - \ell_k\|$. Having located products, we proceed to our final third step.

## 3.3 Locating Firms

We now locate each multi-product firm $\omega$ in the product space to a vector $\ell_\omega$. We do so by finding $\ell_\omega$ to solve

$$\min_{\ell_\omega} \sum_{\nu} \left( \frac{\|\ell_\nu - \ell_\omega\|}{T\left(a_{\omega,\nu}\right)} \right)^2, \tag{20}$$

where the sum is taken over every product $\nu$ produced by firm $\omega$. This projects the firm's location onto the product vectors it produces with the explicit form

$$\ell_\omega = \sum_{\nu} \frac{T\left(a_{\omega,\nu}\right)^{-2}}{\sum_k T\left(a_{\omega,k}\right)^{-2}} \ell_\nu. \tag{21}$$

The location of a firm in Equation (21) is a weighted average of the location of the products that it makes, with zero weight on products it does not produce. Note that the firm's location captures its *relative* advantage in each product, so a firm cannot be close to every product or far from every product. Furthermore, since the product vectors do not perfectly reflect Euclidean space, firm-product distances need not be proportional to observed advantages.

This solution is intuitive considering the case of equidistant products (all $\|\ell_\nu - \ell_\tau\|$ equal) and equal advantages (all $a_{\omega,\nu}$ equal). For two products, the firm is on the midpoint of the line between products. For three products, the firm is centered in an equilateral triangle, and so forth.

Given the firm's location, we can then recover the counterfactual (relative) advantages for unproduced products by inverting $T\left(\tilde{a}_{\omega,\tau}\right)$ in the following equation

$$\|\ell_\tau - \ell_\omega\| = T\left(\tilde{a}_{\omega,\tau}\right). \tag{22}$$

The proximity of a firm to unproduced products given in Equation (22) comes from information embedded in the product space and the location of the firm. Together they

determine each firm's counterfactual (relative) advantages to produce products.

We now turn to the data and construction of clusters and product spaces which will form the environment for our estimates of firm behavior.

# 4 Data and Estimation Procedure

In this section, we present the data, construct clusters and product spaces, and report summary statistics. For each sector and year, we construct two different product spaces based on measuring advantages when the underlying markups are constant or variable and compare these to product spaces based on the Combined Nomenclature and on the product relatedness measures from the product complexity literature.

## 4.1 Data

We rely on firm-product-level data from Danish firms, spanning from 2000 to 2018, provided by Denmark Statistics. We use two data sources: the Production Statistics (VARS) and the Trade Statistics (UHDI). The Production Statistics is a survey in which manufacturing firms with at least 10 employees are required to report their sales in quantities and values for each product they produce. Sales are recorded independently of the market in which the product is sold, thereby including both domestic and export sales. This is similar to PRODCOM data available in many European countries, and we combine it with trade statistics to isolate production for the domestic market. In the Trade Statistics, firms report their exports and imports by product and destination. Products are reported according to the 8-digit level of the Combined Nomenclature (CN) code, equivalent to the Harmonized System classification at the 6-digit level, with the firm (CVRNR) being the reporting unit.

Our data preparation closely follows Buus et al. (2022), who have provided the code for the estimation of marginal costs using the method proposed by De Loecker et al. (2016). To account for changes in product categories over time, we employ the algorithm proposed by Van Beveren et al. (2012), aggregating categories to the so-called CN8+ level. A product is defined at both the CN8+ and the unit of measurement level. For most CN8+ categories, firms report the same unit of measurement (kg, number, etc.). In some rare instances, the same product code is recorded with different units, and we consider these as separate products.

For the production function estimation, we also rely on the Firm Statistics Register and the Firm Accounts Statistics, which provide balance sheet, employment, and finance information on the universe of private sector Danish firms.

## 4.2 Advantage Estimation

Our model in Section 2.1 defines the advantage of a firm $\omega$ in product $\nu$ in year $t$ as the ratio between the firm's marginal cost $m_{\omega,\nu,t}$ and the average marginal cost across firms for the same product, $\overline{m}_{\nu,t}$. To compute these advantages, we must first estimate marginal costs, which are typically unobserved by the researcher. We will treat all firm-products within each CN8+ category as differentiated but otherwise identical, abstracting from properties like quality, indicators of which we lack across the entire span of the CN classification. We consider two approaches consistent with our CoPaTh framework. First, we analyze the specific case of CES preferences, where the elasticity of substitution $\sigma_{\omega,\nu} = \sigma$ is constant across all firms and products. In this setting, we invert the demand system to obtain firm-level marginal costs. Second, we adopt the approach of by De Loecker et al. (2016) (DGKP) to estimate production functions and markups. These markups allow us to back out the elasticity $\sigma_{\omega,\nu}$ and recover marginal costs as the ratio between unit values and estimated markups.

While the CES method relies on strong assumptions such as constant markups and a particular nesting of demand across varieties, it does not rely on production function estimation and is therefore widely applicable. By leveraging sophisticated methods and rich data, the DGKP method controls for economies of scope, scale and selection into multi-product status as well as productivity improvements that accompany production adoption. The DGKP results are therefore our preferred estimates despite more stringent data requirements.[14] Using the CES and the DGKP approach, we derive two distinct measures for the marginal costs for product $\nu$ of firm $\omega$, $m^i_{\omega,\nu}$, where $i$ =CES, DGKP. This implies observed advantages $a^i_{\omega,\nu} = m^i_{\omega,\nu}/\overline{m}^i_\nu$. We now detail each construction.

### 4.2.1 Demand Inversion (CES)

By combining the production statistics with the trade statistics, we can compute the total domestic sales for each firm-product-year as the difference between the total production value and total export value. We apply the same procedure to derive the domestic quantity.[15]

We calculate the marginal costs for each firm-product by assigning a value of $\sigma = 5$ for all sectors, using the domestic unit values to compute the price index of Equation

[14]The industrial organization literature has proposed a number of alternatives for estimating multi-product firm behavior, see De Roux et al. (2021); Dhyne et al. (2022); Orr (2022). Some of these approaches would not be in line with our theoretical model and would lead to different theoretical formulas for our counterfactuals. These methods are of course limited to estimating advantages for *produced* products. Our paper complements this line of work by developing an algorithm to recover advantages for *unproduced* products, conditional on estimates.

[15]We exclude observations with negative domestic sales or quantity, which are likely due to firms engaging in carry-along trade (Bernard et al., 2018).

(4) and combining it with domestic sales to compute the marginal costs from Equation (7) and advantages from Equation (6). As we only use information on Danish firms and their domestic sales, in computing the marginal cost, we implicitly ignore foreign firms exporting to Denmark in the calculation of aggregate revenues and price indexes. This does not pose an issue in measuring domestic advantage, since its derivation only requires data on domestic product expenditure shares.

### 4.2.2 Production Function Estimation (DGKP)

We apply the procedure outlined by De Loecker et al. (2016), which integrates the estimation of a production function and markups, to derive marginal costs for each firm-product and thereby absolute advantage. In this scenario, there is no need to assume a demand function to determine marginal costs, as these are inferred from the production decisions of firms. We closely follow Buus et al. (2022) in estimating markups and refer readers to their paper for further details.

The key challenge in estimating a production function for multi-product firms arises from a lack of information on how various inputs are allocated to each product. To address this challenge, De Loecker et al. (2016) propose estimating production functions for single-product firms and then dividing production activity by product, and we follow their approach. The production function is estimated using a control function approach that accounts for unobserved heterogeneity in both productivity and input prices. We follow Buus et al. (2022) and also incorporate product-specific export status, number of export destinations, and the square of the number of export destinations into the control functions. However, unlike Buus et al. (2022), we do not include information on export support, as it is only available for a subset of the years.

To calculate markups, this approach uses the results of cost minimization of flexible inputs.[16] The markup is defined as the wedge between the revenue share of a variable production input and its output elasticity. By estimating the production function, we determine the output elasticity of materials and we interpret the ratio of this elasticity to the revenue share of materials as the price-cost markup. Table A.1 in the Appendix reports, by sector, the means and standard deviations of the estimated output elasticities, returns to scale, and markups. The estimates are similar to those in Buus et al. (2022), with most clusters exhibiting approximately constant returns to scale.

Once the markups are obtained, we calculate the marginal cost as the ratio between unit value and markup. To avoid outliers, we calculate the absolute growth rate of firm-product marginal cost changes and exclude the top 2% of those. Moreover, we drop the top and bottom 3% of estimated markups. We amalgamate the CES and DGKP samples

[16] For details, see De Loecker and Warzynski (2012) and De Loecker (2021).

to ensure that the estimation of the product space, according to the two different marginal costs, is derived using an identical set of firm-products.

In summary, for each firm-product-year in our sample, we estimate the advantages using either the CES demand function (CES) or a production function estimation (DGKP).

## 4.3 Cluster Construction and Analysis

Within each sector, we identify the clusters of all products connected by co-production to each other in *any* year within each sector. We further refine the sample of products to avoid the possibility that one cluster in one year breaks down by two or more separate clusters, or sub-clusters. If there are more than one sub-cluster per cluster each year, we keep the sub-cluster with the largest number of products in it, and drop the products outside the largest cluster in each year. Similarly, we exclude firms that are always single-product in our product space construction, validation, and counterfactual analysis. Finally we retain the common set of products across years in the panel. This procedure drops approximately 15% of observations at the firm-product-year level and 19% at the product-year level. We thereby find a single cluster for each sector spanning all years.

By definition, each cluster stands distinct from the others. Given this distinction, we will refer to a firm as a firm-cluster for the rest of our analysis. Notably, firms that extend across multiple clusters represent 8% of the total number of firms and account for 23% of overall sales. These are firms whose products span multiple sectors, e.g., a firm making both chemical products and plastic products. In a robustness test, we employ a more aggregated sector definition, yielding similar outcomes (refer to Appendix D).[17] When dealing with larger clusters, the proportion of firms spanning multiple clusters diminishes, making up 5% of firms and contributing 4% to sales. Furthermore, with larger clusters, the procedure that drops the smallest sub-clusters per cluster each year leads us to drop 8% of observations at the firm-product-year level and 12% at the product-year level.

Table 1 presents summary statistics that describe the average size of each cluster, both in terms of products and firms. Among these, the Foodstuffs cluster has the highest count of products and firms, while the Transportation cluster registers the lowest. The Metals and Machinery and Electrical clusters both tally above-average numbers in terms of products and firms. Meanwhile, the Chemical and Textile clusters stand out with a high ratio of products to firms.

[17] In this context, the sectors are categorized as: Animals/Vegetables/Food (CN 2-digit 01-24), Minerals/Chemicals/Plastics (25-40), Textiles/Footwear (41-43, 50-67), Stone/Metals (68-83), Machinery/Transportation (84-89), and Miscellaneous (44-49, 90-97).

Table 1: Cluster Descriptive Statistics

| | Number of Products | | Number of Firms | |
|---|---|---|---|---|
| Cluster | Avg. | Std. | Avg. | Std. |
| Animal Products | 106 | 35 | 37 | 19 |
| Foodstuffs | 227 | 19 | 107 | 13 |
| Mineral products | 21 | 7 | 14 | 8 |
| Chemical Products | 148 | 26 | 46 | 11 |
| Plastics and rubber | 54 | 13 | 41 | 10 |
| Wood products | 19 | 7 | 38 | 25 |
| Textiles | 178 | 33 | 39 | 22 |
| Stone and glass | 15 | 5 | 25 | 15 |
| Metals | 103 | 23 | 95 | 21 |
| Machinery and Electrical | 127 | 16 | 92 | 12 |
| Transportation | 10 | 3 | 10 | 2 |
| Miscellaneous | 48 | 11 | 89 | 34 |

In each cluster, we compute average (Avg.) and standard deviation (Std.) of the number of products (first two columns) and number of firms (last two columns) across years.

Table 2 reports the descriptive statistics for the identified clusters. On average, a cluster comprises 88 products and is associated with 53 firms. The distributions of both products and firms exhibit a right skew and display significant interquartile variation. Notably, the counts of products and firms both witnessed a decline around the time of the 2008 financial crisis.

Table 2: Cluster Descriptive Statistics

| | Number of Products | | | | | Number of Firms | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Year | Avg. | Std. | Med. | 25P. | 75P. | Avg. | Std. | Med. | 25P. | 75P. |
| 2000 | 88 | 87 | 53 | 27 | 131 | 62 | 44 | 56 | 26 | 84 |
| 2001 | 90 | 83 | 70 | 18 | 138 | 62 | 42 | 54 | 35 | 87 |
| 2002 | 91 | 80 | 67 | 23 | 141 | 68 | 40 | 67 | 39 | 95 |
| 2003 | 90 | 77 | 62 | 23 | 137 | 61 | 36 | 54 | 39 | 79 |
| 2004 | 97 | 77 | 80 | 30 | 150 | 73 | 38 | 68 | 44 | 113 |
| 2005 | 94 | 76 | 82 | 18 | 149 | 65 | 41 | 53 | 38 | 107 |
| 2006 | 98 | 81 | 89 | 18 | 153 | 60 | 43 | 56 | 20 | 100 |
| 2007 | 68 | 58 | 47 | 23 | 102 | 41 | 32 | 32 | 16 | 66 |
| 2008 | 73 | 61 | 63 | 22 | 108 | 41 | 27 | 33 | 20 | 63 |
| 2009 | 77 | 66 | 71 | 14 | 123 | 44 | 30 | 42 | 18 | 67 |
| 2010 | 75 | 63 | 63 | 13 | 113 | 47 | 33 | 36 | 19 | 79 |
| 2011 | 83 | 67 | 78 | 23 | 127 | 49 | 36 | 36 | 19 | 78 |
| 2012 | 88 | 69 | 84 | 24 | 136 | 46 | 32 | 38 | 21 | 80 |
| 2013 | 88 | 71 | 66 | 21 | 141 | 46 | 35 | 32 | 19 | 78 |
| 2014 | 91 | 74 | 68 | 16 | 155 | 47 | 36 | 40 | 20 | 78 |
| 2015 | 93 | 76 | 66 | 21 | 156 | 46 | 38 | 33 | 17 | 82 |
| 2016 | 94 | 73 | 92 | 20 | 151 | 47 | 38 | 37 | 17 | 81 |
| 2017 | 97 | 76 | 93 | 18 | 151 | 46 | 39 | 39 | 9 | 79 |
| 2018 | 97 | 76 | 98 | 25 | 146 | 48 | 37 | 39 | 17 | 81 |
| **Average** | **88** | **73** | **73** | **21** | **137** | **53** | **37** | **44** | **24** | **83** |

In each year, we compute average (Avg.), standard deviation (Std.), median (Med.), and 25th and 75th percentiles (25P. and 75P.) of the number of products (first five columns) and number of firms (last five columns) across clusters. In each year, there are 12 clusters (for 12 sectors defined as groups of CN 2-digit codes). The last row (Average) reports the average of the statistics across years.

## 4.4 Product Space Based on Combined Nomenclature (CN)

To evaluate our method against a discrete classification system, we consider an alternate product space derived from the CN classification system. In the CN classification, the distances between firms and products, as well as between firms themselves, are discrete. We assume these distances can have one of four distinct values.

Recall that a product $\nu$ is defined as a combination of a unit and CN8+ code. For every firm-cluster in a given year, the core product $\bar{\nu}_\omega$ is defined as the product $\nu$ with the largest sales for the firm $\omega$. Then, the distance of a firm $\omega$ to a product $\nu$ is given by

$$d_{\omega,\nu} = 1 + \mathbf{1}_{\text{CN6}(\bar{\nu}_\omega)\neq\text{CN6}(\nu)} + \mathbf{1}_{\text{CN4}(\bar{\nu}_\omega)\neq\text{CN4}(\nu)} + \mathbf{1}_{\text{CN2}(\bar{\nu}_\omega)\neq\text{CN2}(\nu)}.$$

To elucidate, a firm has a distance of one to all products that fall under the same CN

6-digit code of its core product. It has a distance of two to products under the same CN 4-digit code of its core, but a different CN 6-digit code, and so on.

## 4.5 Product Space Based on Product Relatedness (RCA)

To further validate the efficacy of our approach, we compare the performance of our model in predicting product adoption using a product space based solely on co-production, as it is typically done in the literature on the product space for countries (Hidalgo et al., 2007). Extending this concept to firm co-production, we compute a measure of proximity between products based on how frequently firms have Revealed Comparative Advantage (RCA) in their production at the same time, and then construct a measure of relatedness of firms to products. To compare with our baseline exercise, we compute the measure of product proximity and firm-product relatedness within our clusters.

First, we define the $RCA_{\omega,\nu}$ of a firm $\omega$ for product $\nu$ as the share of firm $\omega$ revenues on product $\nu$, relative to the aggregate sales share of product $\nu$ in the total cluster sales

$$RCA_{\omega,\nu} = \frac{r_{\omega,\nu}}{\sum_{\nu} r_{\omega,\nu}} \Big/ \frac{\sum_{\omega} r_{\omega,\nu}}{\sum_{\omega}\sum_{\nu} r_{\omega,\nu}}. \tag{23}$$

Then, we compute a measure of proximity between product $\nu$ and $\nu'$ as

$$\phi_{\nu,\nu'} = \min\{\text{Prob}(RCA_{\omega,\nu} > 1 | RCA_{\omega,\nu'} > 1),\ \text{Prob}(RCA_{\omega,\nu'} > 1 | RCA_{\omega,\nu} > 1)\}. \tag{24}$$

$\text{Prob}(RCA_{\omega,\nu} > 1 | RCA_{\omega,\nu'} > 1)$ is the probability that a firm has an RCA on $\nu$ conditional on having an RCA on $\nu'$. It is computed as the share of firms that have an RCA on $\nu$ out of the total number of firms that have an RCA on $\nu'$.

Finally, we compute product relatedness of RCA between firm $\omega$ and a product $\nu$ as

$$\text{Relatedness}_{\omega,\nu} = \sum_{\nu'} \phi_{\nu,\nu'} \cdot 1\left(RCA_{\omega,\nu'} > 1\right) \Big/ \sum_{\nu'} \phi_{\nu,\nu'}, \tag{25}$$

where $1\left(RCA_{\omega,\nu'} > 1\right)$ is one when $RCA_{\omega,\nu'} > 1$ and zero otherwise. The distance of a firm to a product is then given by

$$d_{\omega,\nu} = 1 - \text{Relatedness}_{\omega,\nu}. \tag{26}$$

## 4.6 Product Space Analysis

We now have all the elements to determine the location of products and firms and the respective distances. For each cluster-year, we construct the product space using the

procedure outlined in Section 3, and the three transformations for advantage detailed above, with the log specification in the main text and the others in the appendix for robustness.

Here, we present descriptive information about product spaces. The summary statistics for the estimated distances in the year 2000 are shown in Table 3. Table 4 presents the correlation between various measures for the year 2000. All correlations are statistically significant and positive. Notably, the correlation between DGKP and CES distances is 26%,[18] while there is a 29% correlation between the CN and RCA measures, showing the affinity of these pairs of measures.

Product spaces maintain stability over time. We examine the distribution of the growth rate in the distance between a firm and a product, defined as the difference $d_{\omega,\nu,t} - d_{\omega,\nu,t-1}$, since the distances are expressed in logarithms. Both the average and the median firm-to-product growth rate of distance is close to zero on a year to year basis. In the CES framework, distances vary by 5 percentage points within the 25-75 percentile range, indicating general stability of our approach. Distance growth for DGKP is more volatile, varying by 14 percentage points for the same range.

Table 3: Estimated Distances: Summary Statistics (2000)

| | Product-to-Product | | Firm-to-Product | |
|---|---|---|---|---|
| | CES | DGKP | CES | DGKP |
| Average | 0.39 | 0.75 | 0.23 | 0.58 |
| Std. Dev. | 0.23 | 0.32 | 0.09 | 0.21 |
| 5th Perc. | 0.11 | 0.27 | 0.10 | 0.20 |
| 10th Perc. | 0.15 | 0.38 | 0.13 | 0.31 |
| 25th Perc. | 0.23 | 0.56 | 0.17 | 0.46 |
| 50th Perc. | 0.33 | 0.69 | 0.22 | 0.58 |
| 75th Perc. | 0.50 | 0.91 | 0.28 | 0.70 |
| 90th Perc. | 0.70 | 1.11 | 0.34 | 0.81 |
| 95th Perc. | 0.82 | 1.29 | 0.39 | 0.88 |

The table reports the distribution of product-to-product and firm-to-product distances in the year 2000. The sample comprises all product-to-product and all firm-to-product distances across all clusters. For confidentiality reasons, we divide distances data in 100 bins, compute the average distance within bin, and report these values in the rows with percentiles.

[18] There is a higher correlation of 41% between the estimated marginal costs in these two approaches.

Table 4: Correlations Between Measures (2000)

| Distance | CES | DGKP | CN | RCA |
|---|---|---|---|---|
| CES | 1 | | | |
| DGKP | 0.258*** | 1 | | |
| CN | 0.097*** | 0.149*** | 1 | |
| RCA | 0.121*** | 0.134*** | 0.288*** | 1 |

***: significant at 99%, ** at 95%, * at 90%. The table reports the correlation values across distances in our four specifications.

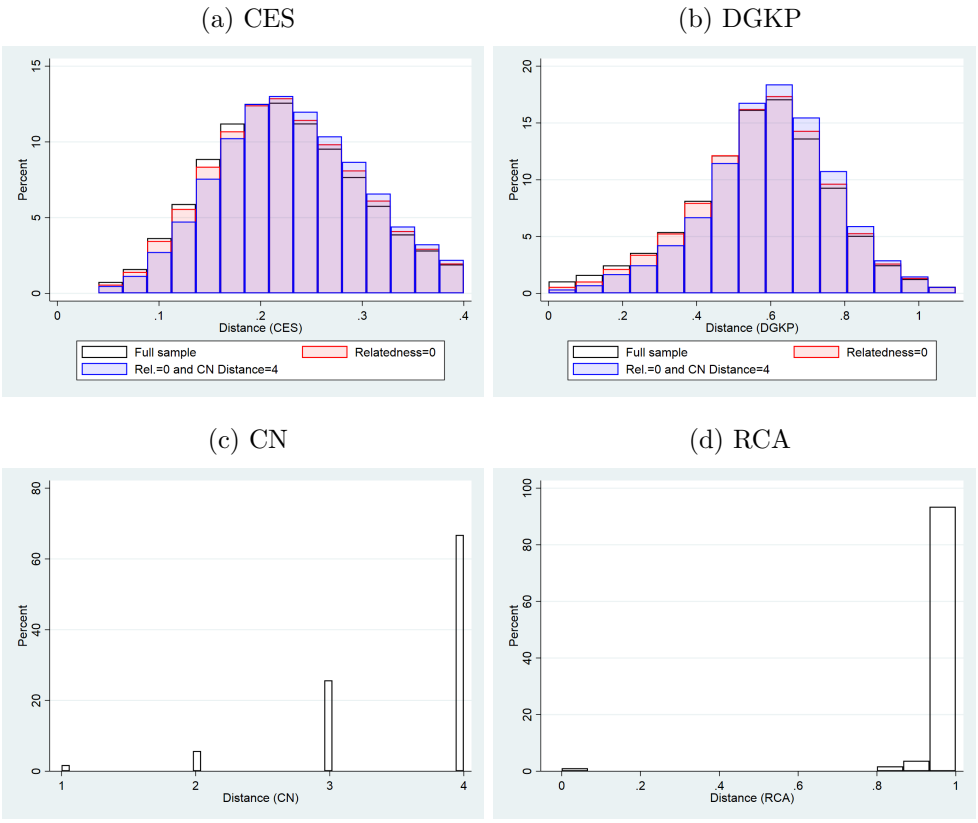
Figure 1 illustrates the distribution of firm-product distances in 2000 across four specifications. The CES and DGKP approaches yield continuous distributions, while the CN approach categorizes distances into four groups, with most observations at distance four, corresponding to different CN 2-digit sectors. Using the RCA measure, over 80% of observations are at the maximum distance of one, indicating that many products are not co-produced within the same year.

Figures 1a and 1b display the distribution of firm-product distances for the CES and DGKP methods under two conditions: when RCA-based relatedness is zero (red bars) and when both RCA-relatedness is zero and CN distance is at its maximum (blue bars). These distributions shift to the right and exhibit considerable variance. This rightward shift results from the correlation between distance measures based on advantages and those derived from CN classifications and RCA co-occurrence. The substantial variance in CES and DGKP distances, even within these restrictive samples, indicates that relying solely on discrete classification systems or RCA co-occurrence metrics fails to capture the full granularity of counterfactual advantages.

**Graphical Representation.** Due to confidentiality, we are unable to provide a visual representation of product spaces at the CN 8-digit level but can do so with aggregation to the CN 4-digit level. Our clustering algorithm produces a single cluster at this level, meaning that each product is directly or indirectly linked to all others. We first illustrate the distribution of pairwise distances across products in Figure 2, which bins distances into ten bins for confidentiality. Figure 2 shows substantial variation in distances both within and across CN product categories. While groups of nearby products are visible, products within the same part of the CN classification are not necessarily equally close to one another. Similarly, relatively close products can appear across different parts of the classification. This reflects that our measure of product distance is based on production capabilities rather than directly on the CN classification.[19]

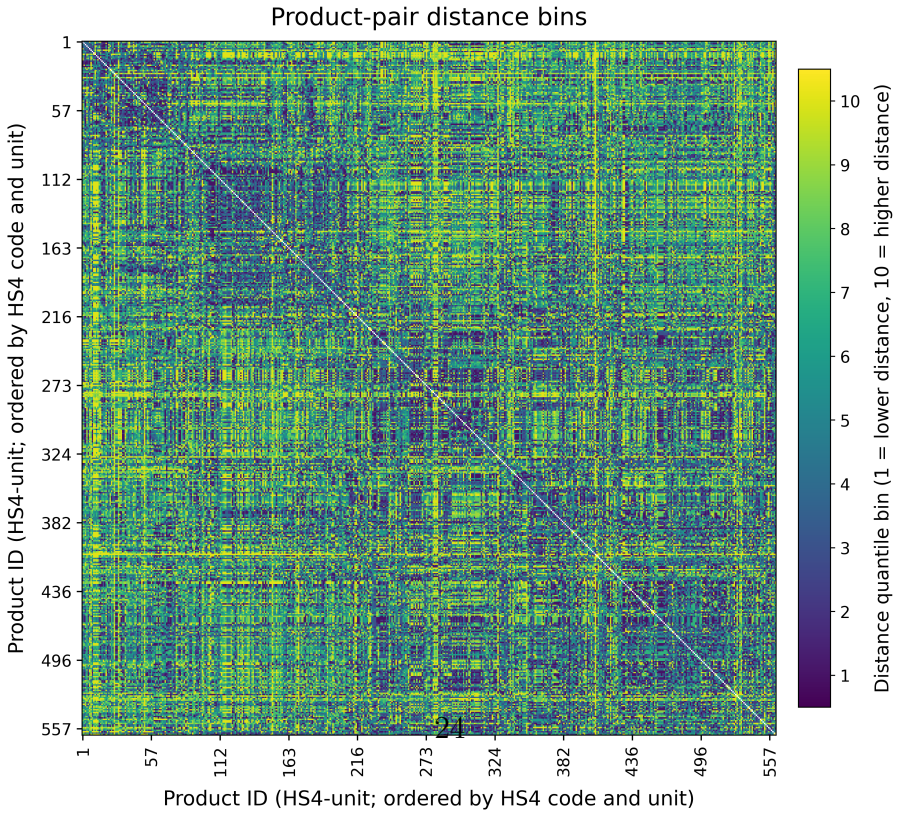
[19]The data underlying this representation is available here: download the distance data.

Figure 1: Distribution of Distances (2000)



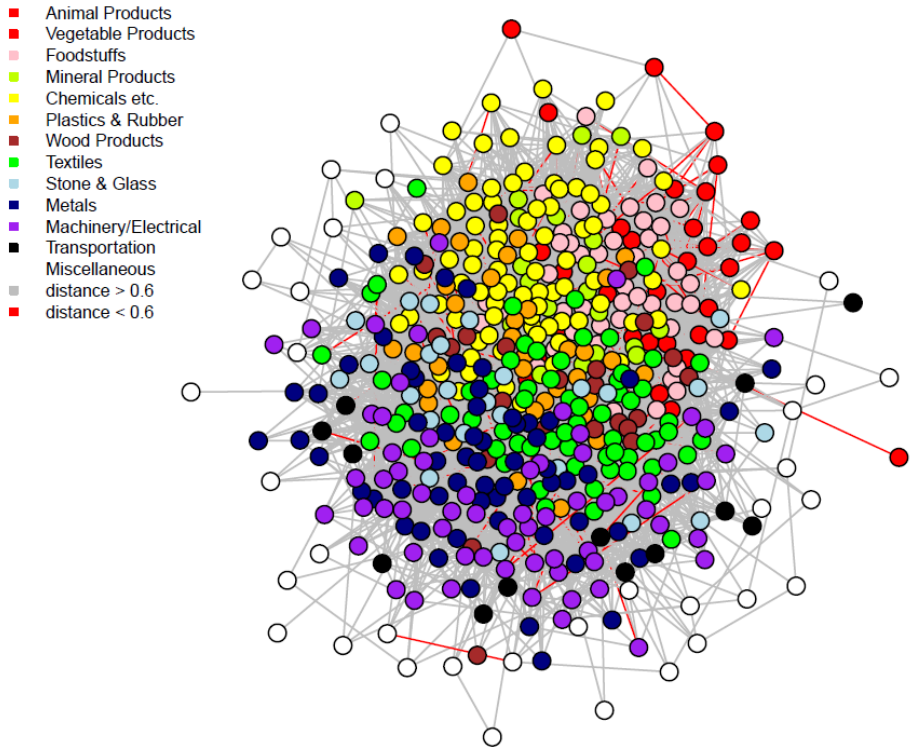

This figure illustrates the distribution of firm-product distances for the year 2000 using various approaches. For confidentiality, we have restricted the samples to ensure that each bar contains a sufficient number of observations. Consequently, the percentage values shown here are higher than those calculated from the full sample.

Figure 2: Product-to-Product Distance Bins




This figure depicts product-to-product distances among CN 4-digit product-units. We use the DGKP and baseline transformation. Marginal costs from DGKP are estimated at the CN 8-digit product-units which distinguish 8-digit products by unit of measure. We aggregate to the CN 4-digit product-unit with a quantity-weighted average across the CN 8-digit

Figure 3: Product Space




This figure depicts the product space with product-to-product distances estimated in a cluster of all directly and indirectly connected CN 4-digit products. We use the DGKP and baseline transformation. Marginal costs from DGKP are estimated at the CN 8-digit product-units which distinguish 8-digit products by unit of measure. We aggregate to the CN 4-digit product-unit with a quantity-weighted average across the CN 8-digit product-units within each 4-digit product-unit. Lines are drawn to create the smallest tree including all products, with some below a cutoff distance omitted for visibility.

Figure 3 provides a network representation of the product space constructed following methods similar to those in Hidalgo et al. (2007). To make the structure visible, we retain a subset of product links rather than displaying all pairwise distances. Products within similar sectors tend to cluster together, such as Chemical products alongside Plastic and Rubber, or Machinery in close proximity to Metals and Transportation. Additionally, the periphery is marked by an assortment of miscellaneous products. Together, Figures 2 and 3 illustrate that the product space contains substantial variation in distances across products while retaining broader similarities among related sectors.

We now move to our estimates of firm behavior with counterfactual advantages from the product space.

# 5 Product Adoption and Counterfactual Advantages

While product-level shocks, like rising demand or changes in tariffs, might prompt a firm to introduce or discontinue a product, the precise prediction of *which* firms will be involved

has largely been overlooked in the literature. In the data, multi-product firms account for 79.1 percent of all product introductions, although only 52.3 percent occur within the firm's best-selling CN 2-digit category and 5.1 percent within their best-selling CN 6-digit code. This capability of multi-product firms to transcend CN codes demonstrates the usefulness of a production based classification that can represent likely product scope choices. This section first quantifies the predictive capability of the product space in determining which products will be introduced. In Appendix C.2, we also examine the predictive capability regarding product discontinuations. The section continues with an IV strategy that shows that demand shocks at the product level lead firms with better counterfactual advantage to adopt them.

## 5.1 Product Adoption and Counterfactual Advantage

For each firm, we identify products not produced in its initial year of observation. For example, if a firm first appears in the data in 2000, we consider as candidate products all items in the firm's cluster not produced in that year. In subsequent years, we construct a product introduction indicator $\text{Intro}_{\omega,\nu,t}$ for firm $\omega$, product $\nu$, and year $t$, equal to 1 if the firm produces the product in that year, and 0 otherwise. We normalize this outcome by the average product introduction rate, which is approximately 0.4 percent. Estimated effects on this outcome can be interpreted as a percentage of the time in the panel that a newly introduced product is actually produced, relative to the base rate.

We estimate the following regression:

$$\text{Intro}_{\omega,\nu,t} = \beta \ln\left(1 + \tilde{a}_{\omega,\nu,t-1}\right) + \phi_{\omega,t} + \rho_{\nu,t} + \epsilon_{\omega,\nu,t}, \tag{27}$$

where the main independent variable is the lagged transformed counterfactual advantage $T(\tilde{a}_{\omega,\nu,t-1}) = \ln(1 + \tilde{a}_{\omega,\nu,t-1})$, as in our baseline product space. We lag this variable to avoid contemporaneous feedback from product adoption into the construction of the product space. For specifications using CN classification or product relatedness (RCA), we substitute $\ln(1+\tilde{a}_{\omega,\nu,t-1})$ with the corresponding lagged firm-product distance measure.

Following our theoretical framework, we include product-year fixed effects ($\rho_{\nu,t}$) to control for time-varying shocks to demand or technology at the product level. Firm-year fixed effects ($\phi_{\omega,t}$) absorb contemporaneous shocks such as changes in productivity, ensuring that identification comes from within-firm variation across products rather than across firms.[20] Throughout this section, standard errors are clustered at the product-year

[20] As discussed above, when a firm produces across multiple, highly distinct CN 2-digit clusters, we treat each firm-cluster pair as a distinct production unit. This ensures that the fixed effects also absorb cluster-year shocks.

level. The findings are illustrated in Table 5. The columns demonstrate that product adoption rates diminish as counterfactual advantage (CN distance, RCA) worsens.

Table 5: Product Introduction and Advantage (All)

| | Dependent Variable: $\text{Intro}_{\omega,\nu,t} > 0$ | | | |
|---|---|---|---|---|
| | (CES) | (DGKP) | (CN) | (RCA) |
| Lagged Advantage | -3.685*** | -1.543*** | -1.688*** | -13.474*** |
| | (0.469) | (0.200) | (0.081) | (0.980) |
| Product-Time FE | Yes | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.07 | 0.07 | 0.08 | 0.07 |
| # Obs. | 645290 | 645290 | 645290 | 645290 |

Results from OLS estimation of (27). Cluster standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

**Discrete versus Continuous Classifications.** As shown in Table 5, distance measures based on the CN classification are predictive of product introduction. This suggests that the CN classification captures meaningful proximity between products: firms tend to introduce products located in nearby CN categories. However, a key limitation of the CN classification is its discreteness. It cannot differentiate between products within the same category, whereas our continuous measures are more granular. For example, a firm whose core product is in cotton fabrics would, under the CN classification, appear equally likely to expand into apparel and into carpets, even though these outputs differ substantially in production processes and inputs. This is because both apparel and carpets lie in different CN 2-digit codes relative to cotton but are grouped within the same cluster, and therefore are at the same CN distance to the firm.

To illustrate the value of continuous classifications, the first three columns of Table 6 report results from estimating (27) while controlling for CN distance bins via fixed effects. The coefficients on advantage remain negative and statistically significant, indicating that our continuous measures retain predictive power even within CN-defined hierarchies.

**Advantage versus RCA** We perform a similar analysis using distance measures based on RCA. These metrics also exhibit predictive power for product introductions, with results comparable to those obtained using our continuous classification. However, a major limitation of the RCA approach is its lack of coverage: over 80% of firm-products have a relatedness of zero, which corresponds to distance taking the maximum value of one. The last three columns of Table 6 show that our advantage-based measures retain predictive power even when the RCA-based relatedness measure is uninformative. This

is reflected in the negative and statistically significant coefficients on advantage.[21]

Table 6: Product Introduction and Advantage: CN Controls and RCA Relatedness

| | Dependent Variable: $\text{Intro}_{\omega,\nu,t} > 0$ | | | | | |
|---|---|---|---|---|---|---|
| | CN Controls | | | RCA Relatedness = 0 | | |
| | (CES) | (DGKP) | (RCA) | (CES) | (DGKP) | (CN) |
| Lagged Advantage | -1.615*** | -0.737*** | -8.914*** | -1.631*** | -0.472*** | -1.194*** |
| | (0.466) | (0.199) | (0.943) | (0.413) | (0.172) | (0.089) |
| Product-Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| CN Distance FE | Yes | Yes | Yes | | | |
| Lagged Relatedness = 0 | | | | Yes | Yes | Yes |
| $R^2$ | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 |
| # Obs. | 645290 | 645290 | 645290 | 542445 | 542445 | 542445 |

Results from OLS estimation of (27). Columns 1–3 control for CN distance bins via fixed effects. Columns 4–6 restrict the sample to observations for which RCA-based relatedness equals zero. Clustered standard errors in parentheses; clustering is at the product-year level. *** significant at 99%, ** at 95%, * at 90%.

## 5.2 Product Adoption and Demand Shocks

Many reasons can prompt a firm to introduce a new product. Here we examine how a positive demand shock influences the likelihood of a product being introduced, and how this relationship is modulated by a firm's counterfactual advantage in that product.

To quantify a demand shock that is credibly exogenous to supply conditions, we instrument for positive export demand shocks on product introduction. We estimate

$$\text{Intro}_{\omega,\nu,t} = \beta \ln\left(1+\tilde{a}_{\omega,\nu,t-1}\right) + \gamma \text{Exports}_{\nu,t} \cdot \ln\left(1+\tilde{a}_{\omega,\nu,t-1}\right) + \phi_{\omega,t} + \rho_{\nu,t} + \kappa_{\omega,\nu} + \epsilon_{\omega,\nu,t}, \tag{28}$$

where the dependent variable $\text{Intro}_{\omega,\nu,t}$ and counterfactual advantages mirror those used in regression (27). $\text{Exports}_{\nu,t}$ denotes the log of the total exports from Denmark of product $\nu$ in year $t$. Equation (28) closely follows the linear probability model for product adoption predicted by our theory in Equation (11), where the probability of introducing a product is log-linear in advantage and depends on product-year, firm-year, and firm-product fixed effects. We restrict the sample to firm-products that have been introduced in the years considered, thereby excluding firm-products $\omega, \nu$ such that $\max_t \text{Intro}_{\omega,\nu,t} = 0$. Therefore, the identifying variation is in the switch of a product from being not produced to being

[21] As a robustness check, we estimate a specification based on ranking products by their distance rather than using the continuous values of distance. The results, reported in Appendix C.1, confirm that firms are more likely to introduce products that are closer in rank, even when including CN fixed effects or when RCA-based relatedness is zero.

produced within each firm.

We instrument $\text{Exports}_{\nu,t}$ by adhering to the approach by Hummels et al. (2014), similar to Dhyne et al. (2021), using either global exports or those from comparable nations to instrument for Danish exports by calculating $\text{Exports IV}_{\nu,t} = \text{Log} \sum_{k \neq DNK} \text{Exports}_{k,\nu,t}$, with the total exports of all countries except Denmark.[22] Notice that export data, sourced from BACI, is reported at the 6-digit level, less granular than our 8-digit level products. The first stage results are presented in Table 7, showing a strong first stage.

Table 7: Export Shocks - IV. All Countries (First Stage)

| | Dependent Variable: Lagged Advantage * Exports | | | |
|---|---|---|---|---|
| | (CES) | (DGKP) | (CN) | (RCA) |
| Lagged Advantage * Exports IV | 0.781*** | 0.792*** | 0.769*** | 0.034*** |
| | (0.003) | (0.003) | (0.004) | (0.002) |
| Firm-Time FE | Yes | Yes | Yes | Yes |
| Firm-Product FE | Yes | Yes | Yes | Yes |
| Prod-Time FE | Yes | Yes | Yes | Yes |
| # Obs. | 13782 | 13782 | 13782 | 13782 |
| $R^2$ | 1.00 | 1.00 | 1.00 | 0.93 |

Results from OLS estimation of the first stage of (28). Cluster standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%. Exports $\text{IV}_{\nu t} = \text{Log} \sum_{k \neq DNK} \text{Exports}_{k\nu t}$.

In the second stage of the IV Table 8, we show that our key parameter of interest, the interaction between distance and exports, is negative and statistically significant under the CES and DGKP cost specifications. This outcome shows that the impact of export demand shocks to increase product introduction is amplified for products where co-production capability is stronger and attenuated for products with less counterfactual advantage. When demand increases for a product, closer firms in the product space supply that demand.

Table 8 shows that both the CN-based and the RCA-based distance measures do not attain statistical significance in the IV specification. In contrast, our more granular advantage based method is less noisy in the IV estimation and can significantly predict adoption following a demand shock.

[22] As a robustness check, we consider $\text{Exports IV}_{\nu,t} = \text{Log} \sum_{k \in K} \text{Exports}_{k,\nu,t}$ where $K$ denotes two different sets of countries: the same set of countries as Autor et al. (2013), swapping Denmark for the USA: Australia, Finland, Germany, Japan, New Zealand, Spain, Switzerland, and USA, and a set of EU countries: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, and Sweden. Results are in appendix C.3, along with the results of the first stage regressions.

Table 8: Export Shocks and Product Adoption

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *Panel A: CES and DGKP advantage measures* | | | | |
| | CES | | DGKP | |
| | OLS | IV | OLS | IV |
| Advantage | 47.60 | 401.5*** | 25.54 | 96.81*** |
| | (66.15) | (121.4) | (17.99) | (31.23) |
| Advantage × Exports | −10.77* | −41.70*** | −3.429** | −9.555*** |
| | (5.650) | (10.52) | (1.522) | (2.659) |
| KP F-statistic | | 404.49 | | 634.29 |
| *Panel B: CN and RCA advantage measures* | | | | |
| | CN | | RCA | |
| | OLS | IV | OLS | IV |
| Advantage | 3.720 | 8.426 | −60.62*** | −52.54*** |
| | (5.447) | (10.03) | (4.094) | (19.20) |
| Advantage × Exports | −0.561 | −0.971 | −3.737*** | −20.43 |
| | (0.465) | (0.869) | (1.088) | (39.32) |
| KP F-statistic | | 484.58 | | 7.66 |
| Firm–Time FE | Yes | Yes | Yes | Yes |
| Firm–Product FE | Yes | Yes | Yes | Yes |
| Product–Time FE | Yes | Yes | Yes | Yes |
| Observations | 13,782 | 13,782 | 13,782 | 13,782 |

*Notes:* Results from OLS and IV estimation of (28). Standard errors clustered by product–year are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. We instrument Log Exports$_{\nu t} \times \ln(1+\tilde{a}_{\omega,\nu,t-1})$ with Log Exports IV$_{\nu t} \times \ln(1+\tilde{a}_{\omega,\nu,t-1})$, where Log Exports IV$_{\nu t} = \log\left(\sum_{k\neq \text{DNK}} \text{Exports}_{k\nu t}\right)$ is total exports by all countries except Denmark. KP F-statistic denotes the Kleibergen–Paap rank Wald F-statistic.

We next use counterfactual advantages and markups to analyze the extensive margin effects of cost shocks. As an application we then evaluate when policies targeting product entry can help stabilize prices better than status quo preserving subsidies to incumbents.

# 6 Cost Shocks and Price Stabilization

In light of the findings and methods above, we now illustrate the potential to address new questions with a policy exercise. Our results show that much entry is driven by multi-product firms expanding their scope across broad product categories. The capabilities required to produce a given good are therefore not necessarily confined to its conventional

industry, and firms in seemingly unrelated sectors may possess latent capabilities that allow them to introduce these products when shocks occur. This creates a targeting problem for industrial policy: support directed only toward existing producers or firms classified within the targeted industry may overlook firms that could enter at relatively low cost.

Existing industrial policies operate on both sides of this margin. Some instruments subsidize production by firms already supplying a targeted good, while others subsidize investments required to establish or expand domestic production capacity. For example, U.S. clean-energy policy combines production-based support under Section 45X with investment support under Section 48C, while semiconductor policies in the United States and Europe and European IPCEIs provide support for investments intended to establish or expand production in strategic value chains. By recovering counterfactual costs and benefits for products firms do not currently produce, our framework can identify potential entrants across conventional industry boundaries and quantify how reductions in their fixed or marginal costs affect entry and prices. We first illustrate these responses to cost reductions and then use the resulting counterfactuals to compare price-stabilization policies operating on the intensive and extensive margins.[23]

## 6.1 Counterfactual Advantages in Levels

Our first goal is to derive counterfactuals to quantify how firm production, entry, and price levels change when marginal or fixed costs decrease. Previously, we used observed advantages $a_{\omega,\nu,t}$ and the product space to recover counterfactual advantages $\tilde{a}_{\omega,\nu,t}$ relative to the firm. In the product adoption exercises, the firm-wide efficiency term $\varphi_{\omega,t}$ introduced in Section 2 is absorbed by firm-time fixed effects. Here, we instead require counterfactual advantages in levels in order to compute marginal costs, profits, and prices. Following Equation (8), these are given by

$$a^{cf}_{\omega,\nu,t} \equiv \tilde{a}_{\omega,\nu,t}/\varphi_{\omega,t}. \tag{29}$$

To recover $\varphi_{\omega,t}$, we use a product in each firm's observed portfolio for which both the observed advantage and the counterfactual relative advantage are available. For each firm $\omega$, let $\overline{\nu}_{\omega,t}$ denote the product in its portfolio that is produced by the largest number of firms. We use a widely produced product as the reference point to reduce sensitivity

[23] Here we take products in isolation for simplicity, ignoring potential interactions such as Input-Output linkages which could dampen or magnify effects.

to normalization using products produced by only a small number of firms. We define

$$\widehat{\varphi}_{\omega,t} \equiv \tilde{a}_{\omega,\overline{\nu}_{\omega,t},t}/a_{\omega,\overline{\nu}_{\omega,t},t}, \tag{30}$$

which aligns the counterfactual relative advantage of the reference product with its observed advantage. We then construct the level counterfactual advantage for every unproduced firm-product pair as

$$\widehat{a}^{cf}_{\omega,\nu,t} \equiv \tilde{a}_{\omega,\nu,t}/\widehat{\varphi}_{\omega,t}. \tag{31}$$

We validate this normalization in Appendix F. For firm-products where advantage is observed, the normalized counterfactual advantage is positively associated with observed advantage under both the CES and DGKP measures. We also find a positive relationship when using lagged counterfactual advantage for products that firms subsequently introduce. These results support using the normalization to map relative counterfactual advantages into levels for the quantitative exercises below.

These counterfactual advantages are in the same units as the observed advantages $a_{\omega,\nu,t}$ and therefore map directly into counterfactual marginal costs. We also require counterfactual markups. For constant markups (CES), we again set the elasticity of substitution to $\sigma = 5$ for a markup of 1.25. For heterogeneous markups, we estimate firm-product elasticities using the DGKP markups and predict markups for unproduced firm-product pairs. We next detail that estimation.

## 6.2 Counterfactual heterogeneous markups

Here we detail how to estimate counterfactual firm-product markups with a model of firm and product fixed effects. Given the markup $\mu_{\omega,\nu,t}$ for product $\nu$ produced by firm $\omega$, we propose a simple structure consistent with CoPaTh preferences. We estimate markups within each cluster according to an OLS estimate of $\beta_t = \left[ \begin{array}{cc} \beta_{\omega,t} & \beta_{\nu,t} \end{array} \right]$ in

$$\ln \mu_{\omega,\nu,t} = \left[ \begin{array}{cc} \chi_{\omega,t} & \chi_{\nu,t} \end{array} \right] \beta_t^T + \eta_{\omega,\nu,t}, \tag{32}$$

where $\chi_{\nu,t}$ and $\chi_{\omega,t}$ are row vectors corresponding to the fixed effects of variety $\nu$ and firm $\omega$. This specification allows us to estimate firm and product specific components of markups for every produced product. We then use estimated $\hat{\beta}_t$ by cluster to predict counterfactual markups $\hat{\mu}_{\omega,\nu,t}$ as

$$\hat{\mu}_{\omega,\nu,t} \equiv \exp\left( \left[ \begin{array}{cc} \tilde{\chi}_{\omega,t} & \tilde{\chi}_{\nu,t} \end{array} \right] \hat{\beta}_t^T \right), \tag{33}$$

where $\tilde{\chi}_{\nu,t}$ and $\tilde{\chi}_{\omega,t}$ are row vectors ranging over all unproduced firm-variety pairs in each cluster. While estimates of $\beta_t$ in Equation (32) are not uniquely identified due to different possible normalizations, all estimates obtained from this regression imply identical predicted counterfactual markups in Equation (33) within cluster.[24] This fact is a corollary of the standard connected-set identification result for two-way fixed-effects models (Abowd et al., 2002; Jochmans and Weidner, 2019). Within a connected component, the firm and product effects are identified up to offsetting additive constants, implying the predicted markup for any firm-product pair in a cluster is uniquely determined.

Summaries of the estimation results from Equation (32) are presented in Table 9. With the exception of the Miscellaneous cluster and the broad Textiles cluster, the fit is strong with $R^2$ over .7. We compare the average markup by cluster across unproduced and produced products on the right of the Table, weighting them by inverse advantage in columns 4 and 5 to capture differences in firm size. Differences across the unproduced and produced columns demonstrate that heterogeneous markups reveal additional extensive-margin gains that a constant markup framework would miss.

Table 9: Markup Fit and Average Markups by HS Product Cluster

| HS Group | N obs | $R^2$ | Advantage-Weighted $\mu$ | | Unweighted $\mu$ | |
|---|---|---|---|---|---|---|
| | | | Unproduced | Produced | Unproduced | Produced |
| Animal Products | 84,979 | 0.798 | 1.695 | 1.514 | 1.513 | 1.359 |
| Chemicals | 129,272 | 0.782 | 1.508 | 1.550 | 1.344 | 1.244 |
| Foodstuffs | 464,151 | 0.625 | 1.467 | 1.286 | 1.367 | 1.236 |
| Machinery and Electrical | 222,932 | 0.870 | 1.895 | 2.058 | 1.654 | 1.527 |
| Metals | 192,401 | 0.818 | 1.801 | 2.110 | 1.540 | 1.449 |
| Mineral Products | 6,466 | 0.935 | 1.479 | 1.951 | 1.438 | 1.236 |
| Miscellaneous | 81,716 | 0.506 | 1.735 | 1.522 | 1.543 | 1.394 |
| Plastics and Rubbers | 44,187 | 0.860 | 1.956 | 1.634 | 1.580 | 1.367 |
| Stone and Glass | 7,514 | 0.722 | 1.897 | 1.987 | 1.729 | 1.601 |
| Textiles | 143,606 | 0.642 | 1.698 | 1.584 | 1.468 | 1.302 |
| Transportation | 1,848 | 0.826 | 1.874 | 2.045 | 1.662 | 1.479 |
| Wood Products | 15,225 | 0.821 | 1.745 | 1.795 | 1.535 | 1.528 |

*Note*: Table reports markup-fit statistics and average predicted markups by HS product cluster. Cluster $R^2$ reports the regression fit within each cluster. The overall $R^2$ from a regression of log estimated DGKP markups on log predicted markups, pooling all clusters, is 0.752. Advantage-weighted $\mu$ is the advantage-weighted markup; Unweighted $\mu$ is the simple average markup. Produced refers to firm-product observations that are produced by the firm; Unproduced refers to non-produced firm-product observations.

With these counterfactuals in hand we can now populate the model to evaluate a wide variety of changes and policies. We begin with cost reduction shocks and then turn to price stabilization in the face of cost shock increases. To simplify notation in the exercises

[24]The predicted markups are trimmed to range between 1.1 and 2.5, by changing values outside this range to the closest in the range. Our positive results are robust to a wider potential range of markups, although the normative results are affected in the DGKP case. Adapting the markups estimated from this method for normative conclusions is an area for future work.

below, we suppress time subscripts.

### 6.3 Extensive Margin Responses to Cost Reduction Shocks

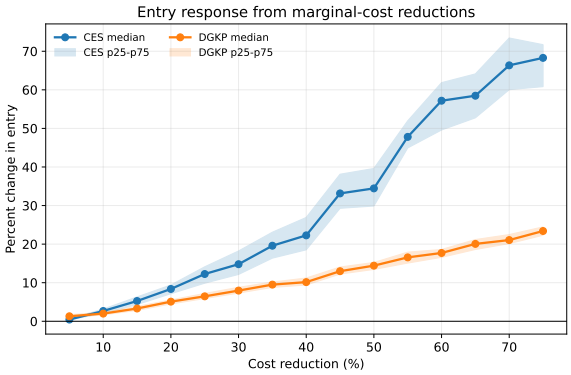
We now compute the effect of marginal and fixed cost shock reductions on firm entry using our counterfactuals. First we need to fix values for fixed costs of entry for every product. Directly measuring fixed costs is notoriously difficult, and quantitative estimates often rely on structural estimation (Das et al., 2007). A related literature uses minimum observed export revenues to infer the profitability threshold for exporting (Crozet et al., 2012), and Arkolakis et al. (2021) use variation in these minima to study economies of scope. Because our counterfactuals deliver operating profits directly, we instead follow the more closely related approach of Feenstra et al. (2026) and use the lower tail of model-implied operating profits to discipline product-level fixed costs in each year. Considering a definition of fixed costs as the minimum of observed variable profits, there are often firms with counterfactual variable profits higher than these fixed costs and therefore should have entered. In that instance, we instead increase fixed costs by defining product level fixed costs $f_\nu$ by looking at all firms producing $\nu$, taking the minimum of each firm's variable profits and define $f_\nu$ as the maximum of these. If there remain firms with counterfactual profits high enough to enter, we increase fixed costs $f_\nu$ further by defining them as the maximum of the minimum of all variable profits of all firms in the cluster. We summarize the frequencies across cases in Table E.1 of the Appendix.[25]

Having set fixed costs, we next predict aggregate product entry and price index reductions at successive levels of cost reduction to both fixed and marginal costs for potential entrants in Figure 4. Panels a and b show the effect of marginal cost reductions across the two markup settings, with considerably stronger entry responses under constant markups and similar but slightly stronger effects on the price index under heterogeneous markups. Unsurprisingly, fixed cost reductions induce entry quickly across both markup settings, but here price index reductions occur at a much faster rate under heterogeneous markups. The figure suggests that entry responses become more elastic as fixed costs fall, whereas the elasticity with respect to marginal-cost reductions is initially fairly constant and then increases somewhat. Both findings differ from theoretical benchmarks in which entry elasticities are constant (e.g., Chaney (2008)).

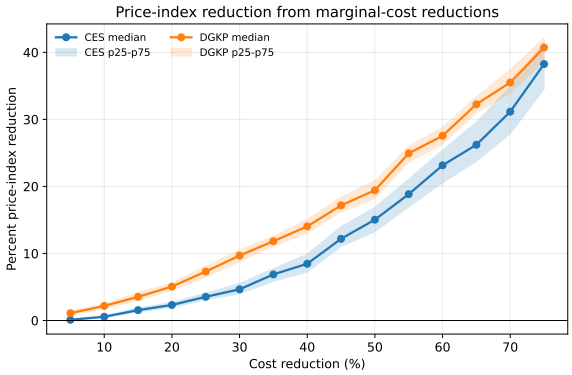
This cost reduction exercise provides insight into the underlying structure of the model counterfactuals, and we can go further to evaluate alternative policy instruments. A broad distinction in current industrial policy is between support tied to production by existing suppliers and support for the investments required to establish new production

[25]This leaves very few cases of firms who should have entered a product but did not. For these we treat them as having entered at the counterfactual quantities implied.

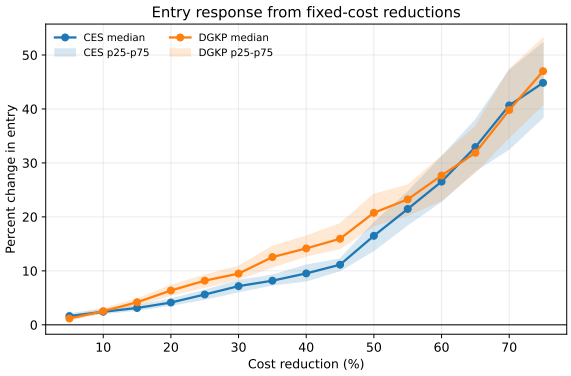
Figure 4: Counterfactual entry and price-index responses to cost reductions



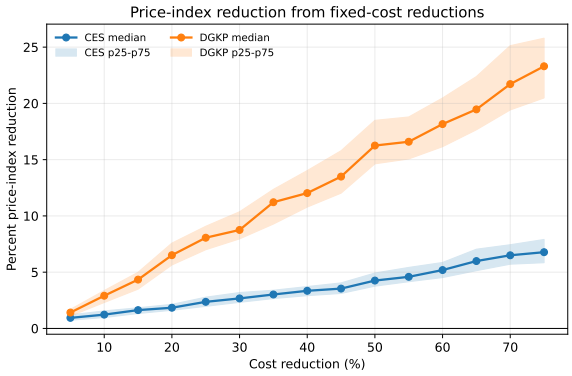

(a) Marginal-cost reduction: entry




(b) Marginal-cost reduction: price index




(c) Fixed-cost reduction: entry




(d) Fixed-cost reduction: price index

*Notes:* The figure reports bootstrap estimates of product-level counterfactual responses to marginal-cost and fixed-cost reductions. Lines show the median bootstrap estimate. Shaded bands show the 25th and 75th percentiles across product-level bootstrap replications. Entry responses are proportional changes in the number of active firm–product varieties. Price-index responses are measured as reductions in the product price index, so positive values indicate lower prices.

capacity. The U.S. 45X production credit and 48C investment credit provide examples of instruments operating primarily on these different margins, although our exercise is intended as a stylized comparison rather than a representation of either program. We next evaluate whether following a cost shock to incumbent producers, a policymaker can attain a given price objective more cheaply by subsidizing incumbent production or by reducing the fixed costs of entry for firms with latent capabilities in the affected product.

## 6.4 Price Stabilization Through Entry

We now consider a policymaker who faces a marginal cost shock to incumbent firms and wishes to maintain the price index $P\left(\boldsymbol{p_\nu}^{\text{mkt}}\right)$ where the superscript $mkt$ indicates the original market equilibrium value.[26] Whether this is for strategic capability, political stability, dynamic efficiency or other rationales, we take the objective as given. Consider a setting at the market equilibrium in which a cost shock $\tau_\nu$ occurs to marginal costs, so that firms face marginal costs $(1+\tau_\nu)\, m_{\omega,\nu}$. One immediate policy option is to implement a status quo subsidy $s_\nu^{\text{quo}} \equiv \tau_\nu$ so that firms face marginal costs $(1+\tau_\nu - s_\nu^{\text{quo}})\, m_{\omega,\nu} = m_{\omega,\nu}$ returning to market quantities and prices at a cost of

$$C_\nu^{\text{quo}} \equiv s_\nu^{\text{quo}} \int_{\Omega_\nu^{\text{mkt}}} m_{\omega,\nu} q_{\omega,\nu}^{\text{mkt}} d\omega$$

where the integral is over all firms producing product $\nu$ in the market equilibrium, $\Omega_\nu^{\text{mkt}}$. While we do not have a formal proof, following the logic of Dhingra and Morrow (2019) this allocation should be efficient at pre-shock levels due to constant (though heterogeneous) markups, so the effects here are not capturing gains from correcting distortions.

However, since fixed costs are now smaller compared to marginal costs, this opens the possibility that inducing product entry can attain the same price-index target at lower fiscal cost than offsetting the incumbent shock. This can be done with an entry subsidy that reduces fixed costs by a percentage $e_\nu$. This must achieve a price index of $P\left(\boldsymbol{p_\nu}^{\text{ent}}\right) = P\left(\boldsymbol{p_\nu}^{\text{mkt}}\right)$ where the superscript $ent$ indicates market prices under the entry subsidy. This would cost a total of

$$C_\nu^{\text{ent}} \equiv e_\nu \int_{\Omega_\nu^{\text{ent}}} f_\nu d\omega$$

where the integral is over all new firms producing product $\nu$ under the entry subsidy policy, $\Omega_\nu^{\text{ent}}$.

We next compare this price stabilizing entry subsidy with the previous status quo

[26] Note that since product revenue $R_\nu$ is unchanged, this is equivalent to maintaining the same aggregate level of product output.

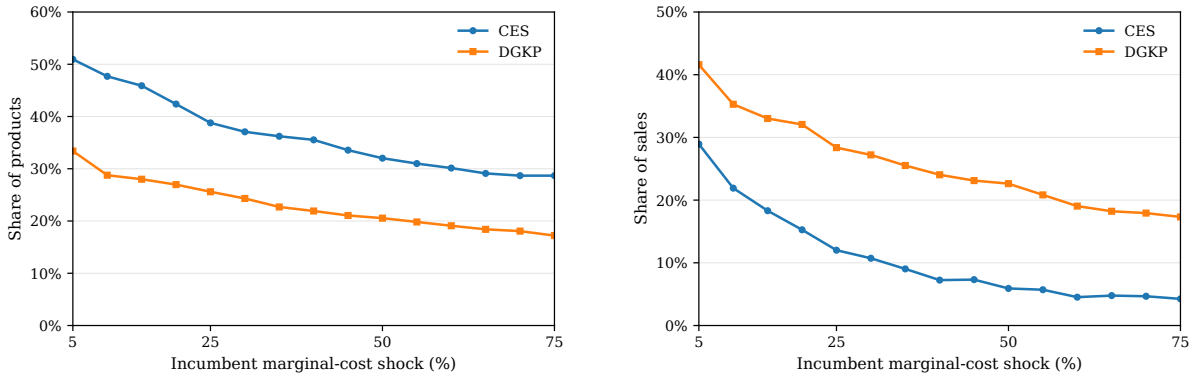
policy on a total cost basis. This is enabled by using our counterfactuals to identify firms that can enter products across broad product classifications. The entry policy operates through a subsidy to the fixed cost of producing the product, while the status quo policy subsidizes the marginal costs of incumbent firms.

Figures 5a and 5b display the share of products and the share of sales for which the entry subsidy is cheaper than the status quo subsidy across marginal cost shocks ranging from 5 to 75 percent. Under CES demand, entry subsidies are cheaper for approximately half of products at small shocks. At a 5 percent marginal cost shock, the entry subsidy is cheaper for 50.9 percent of products, representing 28.9 percent of sales. The importance of the entry policy declines as the shock grows: at a 50 percent shock it is cheaper for 32.0 percent of products, representing only 5.9 percent of sales, and at a 75 percent shock the corresponding shares fall to 28.7 and 4.2 percent. This pattern is consistent with the exhaustion of cost-effective potential entrants as the price reduction required to offset the incumbent cost shock becomes larger.[27]

Under heterogeneous DGKP markups, the same declining pattern appears, but the composition of products for which entry is preferred differs substantially. At a 5 percent shock, the entry subsidy is cheaper for 33.4 percent of products accounting for 41.6 percent of sales. At a 50 percent shock, the corresponding shares are 20.5 and 22.6 percent. Thus, relative to CES, accounting for heterogeneous markups leads the entry subsidy to be preferred for fewer products, but those products account for a substantially larger share of sales. In this sense, accounting for heterogeneous markups targets entry subsidies on fewer, higher value products.

[27]Firm entry reduces the price index discretely, unlike a production subsidy. As a result, the entry subsidy will generally overshoot the target: allowing the marginal firm to enter pushes the price index below the original level, while preventing its entry would leave the target unmet. For comparability, we therefore set the Status Quo subsidy so that it reproduces the price index generated by the entry subsidy.

Figure 5: Entry Subsidy Cheaper than Status Quo Subsidy under CES and DGKP




(a) Share of products

(b) Share of sales

*Notes:*Panel (a) reports the share of products for which subsidizing entry through fixed costs is cheaper than subsidizing the marginal costs of incumbent firms while attaining the same price-index target. Panel (b) reports the share of sales accounted for by those products. Marginal cost shocks range from 5 to 75 percent.

These results highlight that counterfactual information about latent producers can therefore affect both whether an extensive-margin intervention is preferred and where such an intervention should be targeted. This approach could also be used to evaluate a wide variety of policy proposals which we leave for future work.

# 7 Conclusion

This paper develops a framework to predict which products firms will adopt by recovering counterfactual advantages over unproduced goods. We construct a data-driven product space from observed co-production patterns to map firms' latent capabilities and embed it in a multi-product model with heterogeneous markups. Firm–product proximity in this product space significantly predicts subsequent product adoption, including within conventional product classifications. Using Danish data and an identification strategy based on export-demand shocks, we further show that firms closer in the product space are more likely to enter a product market following positive demand shocks. These results suggest that observed patterns of co-production contain information about firms' latent ability to expand into products they do not currently produce.

The counterfactuals show how these latent capabilities matter for the response to shocks and for industrial policy. Reductions in marginal and fixed costs generate extensive-margin entry responses whose magnitude and price effects depend on the structure of markups. More broadly, the product space changes the relevant set of firms for policy: rather than considering only incumbent producers in a targeted industry, it identifies firms elsewhere in the economy that could profitably become suppliers. In our price stabilization exercise, subsidizing entry can often attain a given price objective at lower fiscal

cost than subsidizing incumbent production, particularly following smaller cost shocks. Accounting for heterogeneous markups also changes which products are targeted: entry subsidies are preferred for fewer products than under constant markups, but those products account for a larger share of sales. The effects of industrial policy therefore depend not only on which products are targeted, but also on which firms possess the latent capabilities to supply them.

This framework opens several avenues for future research. First, it uses ideas from the recommender system literature to embed rich patterns of firm activity into well understood economic models. This provides a way to predict responses to policies such as taxes, subsidies, or capability building while allowing adjustment to occur through both incumbent production and entry. Second, the product space could be extended to settings where conventional product classifications provide less information about productive similarity, including services. Finally, extending the approach to study firm entry and exit, interactions across products, and other sources of adjustment could further help quantify competitive responses and the policy levers that shape who can produce what.

# References


Abowd, J. M., R. H. Creecy, and F. Kramarz (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," Tech. Rep. TP-2002-06, Center for Economic Studies, U.S. Census Bureau.

Arkolakis, C., S. Ganapati, and M.-A. Muendler (2021): "The Extensive Margin of Exporting Products: A Firm-level Analysis," *American Economic Journal: Macroeconomics*, 13.

Autor, D., D. Dorn, and G. H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103, 2121–2168.

Bengio, Y., A. Courville, and P. Vincent (2014): "Representation Learning: A Review and New Perspectives," ArXiv:1206.5538 [cs].

Bernard, A. B., E. J. Blanchard, I. Van Beveren, and H. Vandenbussche (2018): "Carry-Along Trade," *Review of Economic Studies*, 86, 526–563.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010): "Multiple-Product Firms and Product Switching," *American Economic Review*, 100, 70–97.

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

Bloom, N., M. Schankerman, and J. Van Reenen (2013): "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 81, 1347–1393.

Boehm, J., S. Dhingra, and J. Morrow (2022): "The comparative advantage of firms," *Journal of Political Economy*, 130, 3025–3100.

Buus, M. T., J. R. Munch, J. Rodrigue, and G. Schaur (2022): "Do export support programs affect prices, quality, markups and marginal costs? Evidence from a natural policy experiment," *Review of Economics and Statistics*, 1–45.

Chaney, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98, 1707–1721.

Chen, J. and J. Roth (2024): "Logs with zeros? Some problems and solutions," *Quarterly Journal of Economics*, 139, 891–936.

Crozet, M., K. Head, and T. Mayer (2012): "Quality Sorting and Trade: Firm-Level Evidence for French Wine," *The Review of Economic Studies*, 79, 609–644.

Das, S., M. J. Roberts, and J. R. Tybout (2007): "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, 75, 837–873.

De Loecker, J. (2021): "Comment on *(Un)pleasant...* by Bond et al (2020)," *Journal of Monetary Economics*.

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

De Loecker, J. and F. Warzynski (2012): "Markups and firm-level export status," *American economic review*, 102, 2437–2471.

De Roux, N., M. Eslava, S. Franco, and E. Verhoogen (2021): "Estimating production functions in differentiated-product industries with quantity information and external instruments," Tech. rep., National Bureau of Economic Research.

Dhingra, S. (2013): "Trading Away Wide Brands for Cheap Brands," *American Economic Review*, 103, 2554–84.

Dhingra, S. and J. Morrow (2019): "Monopolistic competition and optimum product diversity under firm heterogeneity," *Journal of Political Economy*, 127, 196–232.

Dhyne, E., A. K. Kikkawa, M. Mogstad, and F. Tintelnot (2021): "Trade and Domestic Production Networks," *Review of Economic Studies*, 88, 643–668.

Dhyne, E., A. Petrin, V. Smeets, and F. Warzynski (2022): "Theory for Extending Single-Product Production Function Estimation to Multi-Product Settings," *National Bureau of Economic Research Working Paper*.

Ding, X. (2023): "Industry Linkages from Joint Production," *Working Paper*.

Eckel, C., L. Flach, and N. Meng (2023): "Demand and Supply Side Linkages in Exporting Multiproduct Firms," Tech. rep., Discussion Paper.

Eckel, C., L. Iacovone, B. Javorcik, and J. P. Neary (2015): "Multi-product firms at home and away: Cost- versus quality-based competence," *Journal of International Economics*, 95, 216–232.

Eckel, C. and P. J. Neary (2010): "Multi-Product Firms and Flexible Manufacturing in the Global Economy," *Review of Economic Studies*, 77, 188–217.

Escolar, E. G., Y. Hiraoka, M. Igami, and Y. Ozcan (2023): "Mapping firms locations in technological space: A topological analysis of patent statistics," *Research Policy*, 52, 104821.

Feenstra, R. and H. Ma (2007): "Optimal Choice of Product Scope for Multiproduct Firms under Monopolistic Competition," *in E. Helpman, D. Marin and T. Verdier, eds., The Organization of Firms in a Global Economy, Harvard University Press*.

Feenstra, R. C. and J. A. Levinsohn (1995): "Estimating Markups and Market Conduct with Multidimensional Product Attributes," *The Review of Economic Studies*, 62, 19.

Feenstra, R. C., L. Macedoni, and M. Xu (2026): "Large Firms, Consumer Heterogeneity, and the Profit Share," *American Economic Journal: Macroeconomics*, 18, 260–291.

Flach, L. and M. Irlacher (2018): "Product versus Process: Innovation Strategies of Multiproduct Firms," *American Economic Journal: Microeconomics*, 10, 236–77.

Fontagne, L., A. Secchi, and C. Tomasi (2018): "Exporters product vectors across markets," *European Economic Review*, 110, 150–180.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010): "Multiproduct firms and product turnover in the developing world: Evidence from india," *Review of Economics and Statistics*, 92, 1042–1049.

Grant, M. (2023): "Where Do the Data Come From? Endogenous Classification in Administrative Data," *Working Paper*.

Guevara, M. R., D. Hartmann, M. Aristaran, M. Mendoza, and C. A. Hidalgo (2016): "The Research Space: Using Career Paths to Predict the Evolution of the Research Output of Individuals, Institutions, and Nations," *Scientometrics*, 109, 1695–1709.

Hausmann, R. and C. A. Hidalgo (2011): "The Network Structure of Economic Output," *Journal of Economic Growth*, 16, 309–342.

Hausmann, R., J. Hwang, and D. Rodrik (2007): "What you export matters," *Journal of Economic Growth*, 12, 1–25.

Hausmann, R. and B. Klinger (2006): "Structural Transformation and Patterns of Comparative Advantage in the Product Space," Faculty Research Working Paper RWP06-041, Harvard University, John F. Kennedy School of Government, also published as Center for International Development Working Paper No. 128.

Hidalgo, C. A. (2021): "Economic complexity theory and applications," *Nature Reviews Physics*, 3, 92–113.

Hidalgo, C. A., B. Klinger, A.-L. Barabàsi, and R. Hausmann (2007): "The Product Space Conditions the Development of Nations," *Science*, 317, 482–487.

Hoberg, G. and G. Phillips (2016): "Text-based network industries and endogenous product differentiation," *Journal of Political Economy*, 124, 1423–1465.

Hummels, D., R. Jorgensen, J. Munch, and C. Xiang (2014): "The wage effects of offshoring: Evidence from Danish matched worker-firm data," *American Economic Review*, 104, 1597–1629.

Irlacher, M. (2024): "Multiproduct Firms in International Trade," in *Oxford Research Encyclopedia of Economics and Finance*, ed. by A. Banerjee, Oxford University Press.

Jacobs, G. and C. ONeill (2003): "On the reliability (or otherwise) of SIC codes," *European Business Review*, 15, 164–169.

Jaffe, A. (1986): "Technological Opportunity and Spillovers of R&D: Evidence from Firms, Patents, Profits and Market Value," *American Economic Review*, 76, 984–1001.

Jakel, I. C., C. Eckel, L. Macedoni, and R. Riezman (2023): "Explorations in Product Space: Multi-Product Firms and their Product Portfolios," *Working Paper*.

Jara-Figueroa, C., B. Jun, E. L. Glaeser, and C. A. Hidalgo (2018): "The Role of Industry-Specific, Occupation-Specific, and Location-specific Knowledge in the Growth and Survival of New Firms," *Proceedings of the National Academy of Sciences*, 115, 12646–12653.

Jochmans, K. and M. Weidner (2019): "Fixed-Effect Regressions on Network Data," *Econometrica*, 87, 1543–1560.

Koren, Y., R. Bell, and C. Volinsky (2009): "Matrix Factorization Techniques for Recommender Systems," *Computer*, 42, 30–37.

Lancaster, K. J. (1966): "A New Approach to Consumer Theory," *Journal of Political Economy*, 74, 132–157.

Macedoni, L. (2022): "Large multiproduct exporters across rich and poor countries: Theory and evidence," *Journal of Development Economics*, 156, 102835.

Macedoni, L. and M. J. Xu (2022): "Flexibility and Productivity: Towards the Understanding of Firm Heterogeneity," *International Economic Review*, 63, 1055–1108.

Matsuyama, K. and P. Ushchev (2017): "Beyond CES: Three Alternative Classes of Flexible Homothetic Demand Systems," *SSRN Electronic Journal*.

——— (2020): "Constant Pass-Through," *Working Paper*.

Mayer, T., M. J. Melitz, and G. I. Ottaviano (2014): "Market Size, Competition, and the Product Mix of Exporters," *American Economic Review*, 104, 495–536.

Mealy, P. and A. Teytelboym (2022): "Economic complexity and the green economy," *Research Policy*, 51, 103948.

Neffke, F., M. Henning, and R. Boschma (2011): "How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions," *Economic Geography*, 87, 237–265.

Nocke, V. and S. Yeaple (2014): "Globalization and multiproduct firms," *International Economic Review*, 55, 993–1018.

Orr, S. (2022): "Within-Firm Productivity Dispersion: Estimates and Implications," *Journal of Political Economy*, 130, 2771–2828.

Pellegrino, B. (2025): "Product Differentiation and Oligopoly: A Network Approach," *American Economic Review*, 115, 1170–1225.

Rosen, S. (1974): "Hedonic prices and implicit markets: product differentiation in pure competition," *Journal of political economy*, 82, 34–55.

Sarwar, B., G. Karypis, J. Konstan, and J. Riedl (2001): "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, Hong Kong Hong Kong: ACM, 285–295.

Van Beveren, I., A. B. Bernard, and H. Vandenbussche (2012): "Concording EU Trade and Production Data over Time," NBER Working Paper 18604, National Bureau of Economic Research, Cambridge, MA.

# A Summary of Estimated Elasticities and Markups

Table A.1: Average output elasticities and markups by HS groups

| Sector | N obs | $\theta_l$ | $\theta_k$ | $\theta_m$ | RTS | $\mu$ |
|---|---|---|---|---|---|---|
| Animal Products | 6963 | -0.015 | -0.021 | 0.662 | 0.625 | 1.441 |
| | | (0.196) | (0.134) | (0.162) | (0.383) | (2.236) |
| Chemicals | 5602 | 0.555 | -0.005 | 0.522 | 1.072 | 0.918 |
| | | (0.243) | (0.110) | (0.056) | (0.129) | (1.730) |
| Foodstuffs | 12075 | 0.176 | 0.079 | 0.592 | 0.847 | 0.938 |
| | | (0.148) | (0.043) | (0.097) | (0.252) | (1.169) |
| Machinery & Electrical | 6020 | 0.458 | 0.003 | 0.533 | 0.994 | 2.059 |
| | | (0.018) | (0.010) | (0.034) | (0.017) | (3.108) |
| Metals | 5368 | 0.464 | 0.036 | 0.500 | 1.000 | 1.610 |
| | | (0.031) | (0.018) | (0.032) | (0.013) | (2.151) |
| Mineral Products | 1337 | 0.730 | 0.029 | 0.206 | 0.965 | 1.139 |
| | | (0.151) | (0.104) | (0.082) | (0.082) | (2.134) |
| Miscellaneous | 6275 | 0.452 | -0.003 | 0.541 | 0.990 | 1.558 |
| | | (0.117) | (0.033) | (0.092) | (0.094) | (1.823) |
| Plastics & Rubbers | 2577 | 0.458 | 0.068 | 0.484 | 1.010 | 1.187 |
| | | (0.157) | (0.056) | (0.090) | (0.054) | (1.672) |
| Stone & Glass | 2256 | 0.549 | 0.033 | 0.477 | 1.059 | 2.538 |
| | | (0.096) | (0.027) | (0.095) | (0.078) | (3.791) |
| Textiles | 9718 | 0.286 | 0.033 | 0.679 | 0.998 | 1.187 |
| | | (0.038) | (0.041) | (0.035) | (0.036) | (1.460) |
| Transportation | 593 | 0.386 | -0.009 | 0.631 | 1.008 | 1.921 |
| | | (0.082) | (0.021) | (0.102) | (0.040) | (2.569) |
| Wood Products | 3338 | 0.473 | 0.025 | 0.505 | 1.004 | 1.756 |
| | | (0.123) | (0.026) | (0.107) | (0.025) | (2.067) |

*Note*: the table reports sample means of key outputs from the De Loecker et al. (2016) estimation procedure by HS product sections. Standard deviations (not standard errors) are reported in brackets. Column 1 reports the number of observations. Columns 2–4 report the average estimated output elasticity with respect to each factor of production for the translog production function. Column 5 reports the average returns to scale, which is the sum of the preceding three elasticities. Column 6 reports average estimated markups.

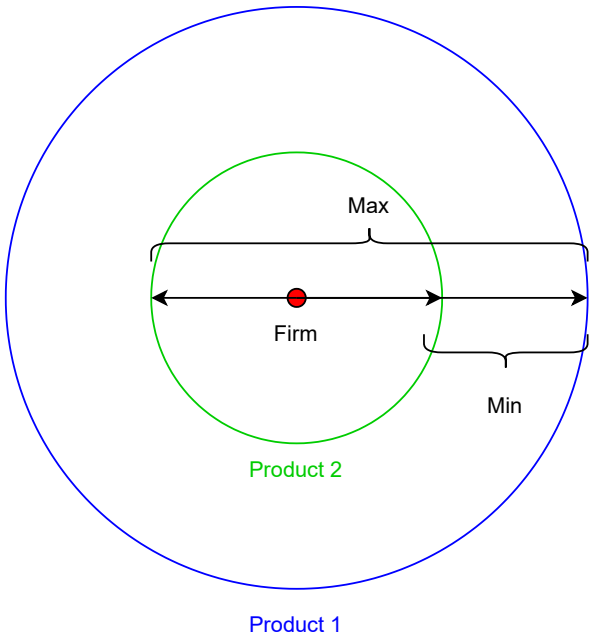
# Supplemental Appendix

## B Geometric interpretation of distance measures

To illustrate the first step in our algorithm geometrically, consider a firm $\omega$ producing two products 1 and 2 with advantages $a_{\omega,1}$ and $a_{\omega,2}$. The distance of the firm to the two products equal these advantages and they can be visualized spatially by placing each product around the firm as in Figure B.1, where product 1 is on the circle of radius $a_{\omega,1}$ and product 2 is on the circle of radius $a_{\omega,2}$. One can imagine if the firm produced product 2 perfectly with $a_{\omega,2} = 0$ (i.e., with an advantage at the zero-th percentile of costs), the distance from the firm to product 1 would be the distance between products 1 and 2, labeled $d_{1,2}$. An application of the triangle inequality in this example provides numerical bounds on this distance

$$|a_{\omega,1} - a_{\omega,2}| \leq d_{1,2} \leq a_{\omega,1} + a_{\omega,2}. \tag{B.1}$$

Figure B.1: Product Distance Bounds in Co-Production




To illustrate this, consider the example in Figure B.1. Conditional on observing the firm's advantages in products 1 and 2, we can apply the triangle inequality to recover bounds of the possible distance between the two products, $d_{1,2}$. The minimum distance arises when both products lie on the same line extending in one direction from the firm's location where both products are located to the right. In this case, the distance between the products equals the absolute difference in their distances to the firm, i.e., $|a_{\omega,1} -$

$a_{\omega,2}|$. Conversely, the maximum distance occurs when the two products lie on opposite directions through the firm with product 1 to the right and product 2 to the left, so that the distance between them equals the sum of the two radii, $a_{\omega,1} + a_{\omega,2}$.

$$d_{\nu,\tau} \equiv \sum_{\omega} \frac{a_{\omega,\nu} + a_{\omega,\tau} + |a_{\omega,\nu} - a_{\omega,\tau}|}{2N_{\nu,\tau}} = \sum_{\omega} \frac{\max\{a_{\omega,\nu}, a_{\omega,\tau}\}}{N_{\nu,\tau}}. \tag{B.2}$$

Equation (B.2) is an average between the upper and lower distance bounds of Equation (B.1) and therefore consistent with Equation (B.1) taken as an average.[28] It is mathematically equivalent to the average of the maximum of the two advantages.

# C Estimation

## C.1 Product Adoption by Proximity

As a robustness check, we estimate a version of the product adoption equation that uses rank-based measures of distance. Specifically, we estimate

$$\text{Intro}_{\omega,\nu,t} = \beta \cdot \text{Rank}_{\omega,\nu,t-1} + \phi_{\omega,t} + \rho_{\nu,t} + \epsilon_{\omega,\nu,t}, \tag{C.1}$$

where $\text{Rank}_{\omega,\nu,t-1}$ denotes the rank of products not produced by firm $\omega$ based on counterfactual advantage (CN distance or product relatedness) in year $t-1$. The closest product receives rank 1.

Table C.1 presents the baseline estimates. Across all specifications, we find that product introduction rates decline with distance rank.[29]

The results are robust to controlling for CN distance fixed effects (Table C.2) and to restricting the sample to firm-product pairs with zero RCA-based relatedness (Table C.3).

[28] Given the bounds in Equation (B.1), one might be tempted to use order statistics to obtain a tighter fit. For example, instead of averaging, one could take the *minimum* of the upper bounds and the *maximum* of the lower bounds across observations to construct a narrower range of potential distances. In the absence of measurement error, this approach would indeed be preferable. However, in the presence of errors, the minimum of the upper bounds can be smaller than the maximum of the lower bounds, especially as the number of observations grows. Empirically, we observe such a pattern in a non-negligible fraction of our sample. For this reason, we adopt a more robust approach that averages the distance bounds across firms to mitigate the influence of errors and outliers.

[29] For the CN classification, product pairs within the same distance bin are randomly ranked due to the discrete nature of the taxonomy. Despite this limitation, the rank-based approach still captures meaningful variation in product proximity.

Table C.1: Product Introduction and Product Rankings

| | Dependent Variable: Dummy=1 for Product Introduction | | | |
|---|---|---|---|---|
| | (CES) | (DGKP) | (CN) | (RCA) |
| Lagged Rank (=1 closest) | -0.005*** | -0.005*** | -0.008*** | -0.009*** |
| | (0.001) | (0.001) | (0.000) | (0.000) |
| Product-Time FE | Yes | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.07 | 0.07 | 0.07 | 0.07 |
| # Obs. | 645290 | 645290 | 645290 | 645290 |

Results from OLS estimation of (C.1). Cluster standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

Table C.2: Product Introduction and Product Rankings

| | Dependent Variable: Dummy=1 for Product Introduction | | |
|---|---|---|---|
| | (CES) | (DGKP) | (RCA) |
| Lagged Rank (=1 closest) | -0.002*** | -0.003*** | -0.006*** |
| | (0.001) | (0.001) | (0.000) |
| Product-Time FE | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes |
| CN Distance FE | Yes | Yes | Yes |
| $R^2$ | 0.08 | 0.08 | 0.08 |
| # Obs. | 645290 | 645290 | 645290 |

Results from OLS estimation of (C.1). Cluster standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

Table C.3: Product Introduction and Product Rankings

| | Dependent Variable: Dummy=1 for Product Introduction | | |
|---|---|---|---|
| | (CES) | (DGKP) | (CN) |
| Lagged Rank (=1 closest) | -0.002*** | -0.002*** | -0.005*** |
| | (0.000) | (0.000) | (0.000) |
| Product-Time FE | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes |
| Lagged Relatedness = | 0 | 0 | 0 |
| $R^2$ | 0.08 | 0.08 | 0.08 |
| # Obs. | 542445 | 542445 | 542445 |

Results from OLS estimation of (C.1). Cluster standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

## C.2 Product Drops

In this section, we quantify the ability of the product space to predict product drops, which we define as non-production following production at entry. As the measure of distance within a firm is based on advantages that we observe, the estimation of the full

product space is not necessary. As a result, we consider this exercise as a sanity check for our measure of absolute advantages. For each firm, we select the products that the firm produces in its first year in the data. For instance, if a firm enters the data set in 2000, the sample of products we consider are the products in the cluster of the firm that the firm produces in 2000. For each product in the years after entry, we compute a production indicator $\text{Drop}_{\omega,\nu,t}$ for product $\nu$, firm $\omega$, and year $t$, which equals 1 if the product is not produced (i.e., dropped) and zero otherwise.

We estimate the following regression

$$\text{Drop}_{\omega,\nu,t} = \beta \cdot \ln\left(1 + \tilde{a}_{\omega,\nu,t-1}\right) + \phi_{\omega,t} + \rho_{\nu,t} + \varepsilon_{\omega,\nu,t} \tag{C.2}$$

where the main independent variable is the lagged transformed counterfactual advantage $T(\tilde{a}_{\omega,\nu,t-1}) = \ln(1 + \tilde{a}_{\omega,\nu,t-1})$, as in our baseline product space.

Results are shown in Table C.4 and C.5. Products with less advantage are more likely to be dropped, as the coefficient on the lagged advantage is positive and statistically significant in each specification. As shown for the case of product introduction, using a discrete classification cannot shed light on which products, within a certain aggregation of codes, are more likely to get dropped, while our measure based on advantage is able to distinguish between these products.

Table C.4: Product Drop and Advantage

| | Dependent Variable: Dummy=1 for Product Drop | | | |
|---|---|---|---|---|
| | (CES) | (DGKP) | (CN) | (RCA) |
| Lagged Advantage | 6.419*** | 1.550*** | 0.155*** | 1.918*** |
| | (0.571) | (0.139) | (0.032) | (0.249) |
| Product-Time FE | Yes | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.60 | 0.60 | 0.60 | 0.60 |
| # Obs. | 22642 | 22642 | 22642 | 22642 |

Results from OLS estimation of (C.2). Cluster standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

Table C.5: Product Drop and Advantage

| | Dependent Variable: Dummy=1 for Product Drop | | |
|---|---|---|---|
| | (CES) | (DGKP) | (RCA) |
| Lagged Advantage | 4.737*** | 1.195*** | 1.589*** |
| | (0.619) | (0.149) | (0.248) |
| Product-Time FE | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes |
| CN Distance FE | Yes | Yes | |
| $R^2$ | 0.60 | 0.60 | 0.60 |
| # Obs. | 22642 | 22642 | 22642 |

Results from OLS estimation of (C.2). Cluster standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

## C.3 Export Shocks and Product Adoption

Table C.6: Export Shocks - IV = Autor et al. (2013) Countries

| | Dependent Variable: Dummy=1 for Product Introduction | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (CES) | (CES) | (DGKP) | (DGKP) | (CN) | (CN) | (RCA) | (RCA) |
| Lagged Advantage | 47.600 | 546.423*** | 25.540 | 101.694*** | 3.720 | 18.018 | -60.616*** | -50.157*** |
| | (66.152) | (158.317) | (17.992) | (38.854) | (5.447) | (12.369) | (4.094) | (11.136) |
| Lagged Advantage | -10.772* | -54.372*** | -3.429** | -9.974*** | -0.561 | -1.806* | -3.737*** | -25.362 |
| X Exports | (5.650) | (13.698) | (1.522) | (3.309) | (0.465) | (1.067) | (1.088) | (21.181) |
| Firm-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm-Product FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Prod-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Spec. | OLS | IV | OLS | IV | OLS | IV | OLS | IV |
| # Obs. | 13782 | 13782 | 13782 | 13782 | 13782 | 13782 | 13782 | 13782 |
| F-Stat | | 191.63 | | 288.95 | | 228.46 | | 23.54 |

Results from OLS and IV estimation of (28). Cluster standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%. We instrument Log Exports$_{\nu t}$ $\times \ln\left(1+\tilde{a}_{\omega,\nu,t-1}\right)$ with Log Exports IV$_{\nu t}$ $\times \ln\left(1+\tilde{a}_{\omega,\nu,t-1}\right)$ where Log Exports IV$_{\nu t}$ = Log $\sum_{k\in K}$ Exports$_{k\nu t}$ where $K$ denotes the same set of countries as Autor et al. (2013): Australia, Finland, Germany, Japan, New Zealand, Spain, Switzerland, and USA. F-Stat denotes the Kleibergen-Paap (KP) rank Wald F-statistic.

Table C.7: Export Shocks - IV = EU Countries

| | Dependent Variable: Dummy=1 for Product Introduction | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (CES) | (CES) | (DGKP) | (DGKP) | (CN) | (CN) | (RCA) | (RCA) |
| Lagged Advantage | 47.600 | 509.799*** | 25.540 | 116.219*** | 3.720 | 14.576 | -60.616*** | -43.784*** |
| | (66.152) | (140.652) | (17.992) | (35.409) | (5.447) | (11.370) | (4.094) | (15.438) |
| Lagged Advantage | -10.772* | -51.171*** | -3.429** | -11.222*** | -0.561 | -1.506 | -3.737*** | -38.540 |
| X Exports | (5.650) | (12.167) | (1.522) | (3.016) | (0.465) | (0.983) | (1.088) | (30.896) |
| Firm-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm-Product FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Prod-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Spec. | OLS | IV | OLS | IV | OLS | IV | OLS | IV |
| # Obs. | 13782 | 13782 | 13782 | 13782 | 13782 | 13782 | 13782 | 13782 |
| F-Stat | | 257.25 | | 419.77 | | 304.73 | | 11.75 |

Results from OLS and IV estimation of (28). Cluster standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%. We instrument Log Exports$_{\nu t}$ $\times \ln\left(1+\tilde{a}_{\omega,\nu,t-1}\right)$ with Log Exports IV$_{\nu t}$ $\times \ln\left(1+\tilde{a}_{\omega,\nu,t-1}\right)$ where Log Exports IV$_{\nu t}$ = Log $\sum_{k\in K}$ Exports$_{k\nu t}$ where $K$ denotes the set of EU countries: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, and Sweden. F-Stat denotes the Kleibergen-Paap (KP) rank Wald F-statistic.

## C.4 First Stage Regressions

Table C.8: Export Shocks - IV. Autor et al. (2013) Countries (First Stage)

| | Dependent Variable: Lagged Advantage X Exports | | | |
|---|---|---|---|---|
| | (CES) | (DGKP) | (CN) | (RCA) |
| Lagged Advantage X Exports IV | 0.889*** | 0.908*** | 0.872*** | 0.040*** |
| | (0.004) | (0.004) | (0.006) | (0.003) |
| Firm-Time FE | Yes | Yes | Yes | Yes |
| Firm-Product FE | Yes | Yes | Yes | Yes |
| Prod-Time FE | Yes | Yes | Yes | Yes |
| # Obs. | 13782 | 13782 | 13782 | 13782 |
| $R^2$ | 1.00 | 0.99 | 1.00 | 0.93 |

Results from OLS estimation of the first stage of (28). Cluster standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%. Log Exports IV$_{\nu t}$ = Log $\sum_{k\in K}$ Exports$_{k\nu t}$ where $K$ denotes the same set of countries as Autor et al. (2013): Australia, Finland, Germany, Japan, New Zealand, Spain, Switzerland, and USA.

Table C.9: Export Shocks - IV. EU Countries (First Stage)

| | Dependent Variable: Lagged Advantage X Exports | | | |
|---|---|---|---|---|
| | (CES) | (DGKP) | (CN) | (RCA) |
| Lagged Advantage X Exports IV | 0.857*** | 0.871*** | 0.838*** | 0.038*** |
| | (0.004) | (0.004) | (0.005) | (0.003) |
| Firm-Time FE | Yes | Yes | Yes | Yes |
| Firm-Product FE | Yes | Yes | Yes | Yes |
| Prod-Time FE | Yes | Yes | Yes | Yes |
| # Obs. | 13782 | 13782 | 13782 | 13782 |
| $R^2$ | 1.00 | 0.99 | 1.00 | 0.93 |

Results from OLS estimation of the first stage of (28). Cluster standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%. Log Exports IV$_{\nu t}$ = Log $\sum_{k\in K}$ Exports$_{k\nu t}$ where $K$ denotes the set of EU countries: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, and Sweden.

## C.5 Advantage transformed in Levels and IHS

In this section, we replicate the baseline results presented in the main text but using transformations corresponding to Levels or Inverse Hyperbolic Sine (IHS) for robustness. Summary statistics for the distances measured from product-to-product and firm-to-product are provided in Table C.10 for Levels and Table C.11 for IHS. The results for product introduction with lagged distance modulating various fixed effects and sub-samples are all qualitatively similar to those above, so we omit them for brevity.

Table C.10: Product Space Summary Statistics (Levels Transformation, Year 2000)

| | **Product-to-Product** | | **Firm-to-Product** | |
|---|---|---|---|---|
| | CES | DGKP | CES | DGKP |
| Average | 0.87 | 2.74 | 0.53 | 1.39 |
| Std. Dev. | 0.49 | 2.16 | 0.22 | 0.71 |
| 5th Perc. | 0.21 | 0.52 | 0.22 | 0.38 |
| 10th Perc. | 0.30 | 0.79 | 0.27 | 0.59 |
| 25th Perc. | 0.48 | 1.00 | 0.37 | 0.91 |
| 50th Perc. | 0.80 | 1.92 | 0.50 | 1.31 |
| 75th Perc. | 1.11 | 3.99 | 0.65 | 1.76 |
| 90th Perc. | 1.53 | 5.93 | 0.81 | 2.19 |
| 95th Perc. | 1.75 | 6.91 | 0.91 | 2.51 |

The table reports the distribution of product-to-product and firm-to-product distances in the year 2000. The sample comprises all product-to-product and all firm-to-product distances across all clusters. For confidentiality reasons, we divide distances data in 100 bins, compute the average distance within bin, and report these values in the rows with percentiles.

Table C.11: Product Space Summary Statistics (IHS Transformation, Year 2000)

| | **Product-to-Product** | | **Firm-to-Product** | |
|---|---|---|---|---|
| | CES | DGKP | CES | DGKP |
| Average | 0.53 | 0.97 | 0.32 | 0.75 |
| Std. Dev. | 0.29 | 0.40 | 0.12 | 0.27 |
| 5th Perc. | 0.15 | 0.36 | 0.15 | 0.26 |
| 10th Perc. | 0.21 | 0.50 | 0.18 | 0.40 |
| 25th Perc. | 0.31 | 0.72 | 0.24 | 0.59 |
| 50th Perc. | 0.45 | 0.88 | 0.31 | 0.75 |
| 75th Perc. | 0.68 | 1.16 | 0.39 | 0.90 |
| 90th Perc. | 0.91 | 1.43 | 0.47 | 1.04 |
| 95th Perc. | 1.06 | 1.66 | 0.52 | 1.13 |

The table reports the distribution of product-to-product and firm-to-product distances in the year 2000. The sample comprises all product-to-product and all firm-to-product distances across all clusters. For confidentiality reasons, we divide distances data in 100 bins, compute the average distance within bin, and report these values in the rows with percentiles.

Tables C.12 and C.13 examine the impact of export demand shocks on product introduction, with the respective first stage regressions detailed in Tables C.14 and C.15. The

estimates show that the qualitative conclusions are the same as in the baseline specification.

Table C.12: Export Shocks and Product Adoption (Levels Transformation)

| | Dependent Variable: Dummy=1 for Product Introduction | | | |
|---|---|---|---|---|
| | (CES) | (CES) | (DGKP) | (DGKP) |
| Lagged Advantage | 10.117 | 59.895*** | 1.462 | 11.970* |
| | (10.153) | (19.114) | (4.096) | (6.356) |
| Lagged Advantage X Exports | -1.889** | -6.320*** | -0.383 | -1.271** |
| | (0.878) | (1.689) | (0.335) | (0.530) |
| Firm-Time FE | Yes | Yes | Yes | Yes |
| Firm-Product FE | Yes | Yes | Yes | Yes |
| Prod-Time FE | Yes | Yes | Yes | Yes |
| Spec. | OLS | IV | OLS | IV |
| # Obs. | 13782 | 13782 | 13782 | 13782 |
| F-Stat | | 526.48 | | 595.68 |

Results from OLS and IV estimation of (28). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%. We instrument Log Exports$_{\nu t} \times \tilde{a}_{\omega,\nu,t-1}$ with Exports IV$_{\nu t} \times \tilde{a}_{\omega,\nu,t-1}$ where Exports IV$_{\nu t}$ = Log $\sum_{k \neq DNK}$ Exports$_{k\nu t}$ is the total exports of all countries except Denmark. F-Stat denotes the Kleibergen-Paap (KP) rank Wald F-statistic.

Table C.13: Export Shocks and Product Adoption (IHS Transformation)

| | Dependent Variable: Dummy=1 for Product Introduction | | | |
|---|---|---|---|---|
| | (CES) | (CES) | (DGKP) | (DGKP) |
| Lagged Advantage | 69.421 | 302.434*** | 15.288 | 70.766*** |
| | (47.155) | (84.589) | (13.665) | (23.694) |
| Lagged Advantage X Exports | -10.349** | -30.700*** | -2.203* | -6.968*** |
| | (4.023) | (7.319) | (1.155) | (2.017) |
| Firm-Time FE | Yes | Yes | Yes | Yes |
| Firm-Product FE | Yes | Yes | Yes | Yes |
| Prod-Time FE | Yes | Yes | Yes | Yes |
| Spec. | OLS | IV | OLS | IV |
| # Obs. | 13782 | 13782 | 13782 | 13782 |
| F-Stat | | 456.18 | | 577.29 |

Results from OLS and IV estimation of (28). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%. We instrument Log Exports$_{\nu t} \times \ln\left(\tilde{a}_{\omega,\nu,t-1} + \sqrt{1+\tilde{a}^2_{\omega,\nu,t-1}}\right)$ with Exports IV$_{\nu t} \times \ln\left(\tilde{a}_{\omega,\nu,t-1} + \sqrt{1+\tilde{a}^2_{\omega,\nu,t-1}}\right)$ where Exports IV$_{\nu t}$ = Log $\sum_{k \neq DNK}$ Exports$_{k\nu t}$ is the total exports of all countries except Denmark. F-Stat denotes the Kleibergen-Paap (KP) rank Wald F-statistic.

Table C.14: Export Shocks - IV = All Countries (Levels, First Stage)

| | Dependent Variable: Lagged Advantage X Exports | |
|---|---|---|
| | (CES) | (DGKP) |
| Lagged Advantage X Exports IV | 0.768*** | 0.799*** |
| | (0.004) | (0.003) |
| Firm-Time FE | Yes | Yes |
| Firm-Product FE | Yes | Yes |
| Prod-Time FE | Yes | Yes |
| # Obs. | 13782 | 13782 |
| $R^2$ | 1.00 | 1.00 |

Results from OLS estimation of the first stage of (28). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%. Exports $\text{IV}_{\nu t} = \text{Log} \sum_{k \neq DNK} \text{Exports}_{k\nu t}$.

Table C.15: Export Shocks - IV = All Countries (IHS, First Stage)

| | Dependent Variable: Lagged Advantage X Exports | |
|---|---|---|
| | (CES) | (DGKP) |
| Lagged Advantage X Exports IV | 0.782*** | 0.792*** |
| | (0.003) | (0.003) |
| Firm-Time FE | Yes | Yes |
| Firm-Product FE | Yes | Yes |
| Prod-Time FE | Yes | Yes |
| # Obs. | 13782 | 13782 |
| $R^2$ | 1.00 | 1.00 |

Results from OLS estimation of the first stage of (28). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%. Exports $\text{IV}_{\nu t} = \text{Log} \sum_{k \neq DNK} \text{Exports}_{k\nu t}$.

# D Larger Cluster Analysis

In this section, we replicate the baseline results presented in the main text, using a more aggregate definition of sectors. Specifically, we consider six sectors: Animals/Vegetables/Food (CN 2-digit 01-24), Minerals/Chemicals/Plastics (25-40), Textiles/Footwear (41-43, 50-67), Stone/Metals (68-83), Machinery/Transportation (84-89), and Miscellaneous (44-49, 90-97). This replication uses the same distance transformation as the baseline. Tables D.1 and D.2 present the descriptive statistics for the cluster characteristics. Table D.3 provides summary statistics for the distances measured from product-to-product and firm-to-product. Notice that relative to our baseline results, the average distances increase by little, with the larger increases concentrated in the higher percentiles.

Tables D.4, D.5, and D.6 detail the correlation between the likelihood of product introduction and the rankings based on the firm's product distance. Tables D.7, D.8, and D.9 explore the relationship between the probability of introducing a product and its lagged distance to the firm. Table D.10 examines the impact of export demand shocks

on product introduction, with the first stage regression detailed in Table D.11.

Table D.1: Cluster Descriptive Statistics

| | Number of Products | | | | | Number of Firms | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Year | Avg. | Std. | Med. | 25P. | 75P. | Avg. | Std. | Med. | 25P. | 75P. |
| 2000 | 207 | 141 | 191 | 84 | 260 | 134 | 64 | 111 | 90 | 196 |
| 2001 | 211 | 126 | 189 | 113 | 258 | 150 | 66 | 129 | 115 | 190 |
| 2002 | 208 | 141 | 174 | 91 | 260 | 143 | 72 | 129 | 83 | 190 |
| 2003 | 200 | 137 | 171 | 85 | 247 | 135 | 68 | 121 | 77 | 183 |
| 2004 | 205 | 123 | 175 | 118 | 250 | 144 | 56 | 133 | 128 | 182 |
| 2005 | 205 | 124 | 166 | 117 | 249 | 141 | 63 | 127 | 120 | 194 |
| 2006 | 211 | 128 | 174 | 124 | 258 | 136 | 60 | 130 | 105 | 192 |
| 2007 | 160 | 102 | 128 | 95 | 203 | 87 | 33 | 88 | 78 | 106 |
| 2008 | 163 | 107 | 130 | 111 | 208 | 84 | 34 | 88 | 79 | 93 |
| 2009 | 163 | 110 | 136 | 108 | 209 | 86 | 37 | 88 | 78 | 100 |
| 2010 | 173 | 97 | 143 | 123 | 203 | 100 | 41 | 111 | 87 | 131 |
| 2011 | 190 | 106 | 156 | 133 | 210 | 108 | 48 | 129 | 91 | 140 |
| 2012 | 195 | 103 | 162 | 135 | 225 | 100 | 40 | 112 | 82 | 133 |
| 2013 | 205 | 104 | 169 | 154 | 239 | 104 | 43 | 118 | 85 | 129 |
| 2014 | 212 | 108 | 167 | 166 | 267 | 105 | 44 | 117 | 94 | 129 |
| 2015 | 214 | 114 | 171 | 156 | 286 | 106 | 43 | 115 | 95 | 139 |
| 2016 | 208 | 120 | 160 | 154 | 288 | 100 | 46 | 109 | 80 | 140 |
| 2017 | 217 | 119 | 171 | 150 | 290 | 106 | 45 | 115 | 107 | 137 |
| 2018 | 219 | 123 | 158 | 151 | 292 | 107 | 45 | 118 | 100 | 140 |
| **Average** | **198** | **118** | **162** | **125** | **247** | **114** | **50** | **115** | **93** | **150** |

In each year, we compute average (Avg.), standard deviation (Std.), median (Med.), and 25th and 75th percentiles (25P. and 75P.) of the number of products (first five columns) and number of firms (last five columns) across clusters. In each year, there are 6 clusters (for 6 sectors defined as groups of CN 2-digit codes). The last row (Average) reports the average of the statistics across years.

Table D.2: Cluster Descriptive Statistics

| | Number of Products | | Number of Firms | |
|---|---|---|---|---|
| Cluster | Avg. | Std. | Avg. | Std. |
| Animals/Vegetables/Food | 407 | 36 | 158 | 25 |
| Machinery/Transportation | 145 | 17 | 109 | 12 |
| Minerals/Chemicals/Plastics | 247 | 31 | 107 | 19 |
| Miscellaneous | 84 | 20 | 158 | 66 |
| Stone/Metals | 127 | 29 | 116 | 24 |
| Textiles/Footwear | 178 | 33 | 39 | 22 |

In each cluster, we compute average (Avg.) and standard deviation (Std.) of the number of products (first two columns) and number of firms (last two columns) across years.

Table D.3: Product Space Summary Statistics (Baseline Transformation, Year 2000)

| | **Product-to-Product** | | **Firm-to-Product** | |
|---|---|---|---|---|
| | CES | DGKP | CES | DGKP |
| Average | 0.39 | 0.82 | 0.25 | 0.65 |
| Std. Dev. | 0.20 | 0.38 | 0.09 | 0.23 |
| 5th Perc. | 0.12 | 0.32 | 0.11 | 0.29 |
| 10th Perc. | 0.16 | 0.41 | 0.14 | 0.37 |
| 25th Perc. | 0.24 | 0.59 | 0.18 | 0.50 |
| 50th Perc. | 0.35 | 0.75 | 0.24 | 0.62 |
| 75th Perc. | 0.48 | 0.99 | 0.29 | 0.77 |
| 90th Perc. | 0.69 | 1.23 | 0.35 | 0.94 |
| 95th Perc. | 0.74 | 1.42 | 0.40 | 1.05 |

The table reports the distribution of product-to-product and firm-to-product distances in the year 2000. The sample comprises all product-to-product and all firm-to-product distances across all clusters. The product space is estimated using marginal costs in logs. For confidentiality reasons, we divide distances data in 100 bins, compute the average distance within bin, and report these values in the rows with percentiles.

Table D.4: Product Introduction and Product Rankings

| | Dependent Variable: Dummy=1 for Product Introduction | | | |
|---|---|---|---|---|
| | (CES) | (DGKP) | (CN) | (RCA) |
| Lagged Rank (=1 closest) | -0.004*** | -0.004*** | -0.004*** | -0.006*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Product-Time FE | Yes | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.05 | 0.05 | 0.05 | 0.05 |
| # Obs. | 1332970 | 1332970 | 1332970 | 1332970 |

Results from OLS estimation of (C.1). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

Table D.5: Product Introduction and Product Rankings

| | Dependent Variable: Dummy=1 for Product Introduction | | |
|---|---|---|---|
| | (CES) | (DGKP) | (RCA) |
| Lagged Rank (=1 closest) | -0.002*** | -0.002*** | -0.004*** |
| | (0.000) | (0.000) | (0.000) |
| Product-Time FE | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes |
| CN Distance FE | Yes | Yes | Yes |
| $R^2$ | 0.06 | 0.06 | 0.06 |
| # Obs. | 1332970 | 1332970 | 1332970 |

Results from OLS estimation of (C.1). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

Table D.6: Product Introduction and Product Rankings

| | Dependent Variable: Dummy=1 for Product Introduction | | |
|---|---|---|---|
| | (CES) | (DGKP) | (CN) |
| Lagged Rank (=1 closest) | -0.001*** | -0.001*** | -0.002*** |
| | (0.000) | (0.000) | (0.000) |
| Product-Time FE | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes |
| Lagged Relatedness = | 0 | 0 | 0 |
| $R^2$ | 0.06 | 0.06 | 0.06 |
| # Obs. | 1173802 | 1173802 | 1173802 |

Results from OLS estimation of (C.1). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

Table D.7: Product Introduction and Advantage

| | Dependent Variable: Dummy=1 for Product Introduction | | | |
|---|---|---|---|---|
| | (CES) | (DGKP) | (CN) | (RCA) |
| Lagged Advantage | -6.166*** | -2.510*** | -2.555*** | -23.188*** |
| | (0.468) | (0.195) | (0.103) | (1.457) |
| Product-Time FE | Yes | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.05 | 0.05 | 0.05 | 0.05 |
| # Obs. | 1332970 | 1332970 | 1332970 | 1332970 |

Results from OLS estimation of (27). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

Table D.8: Product Introduction and Advantage

| | Dependent Variable: Dummy=1 for Product Introduction | | |
|---|---|---|---|
| | (CES) | (DGKP) | (RCA) |
| Lagged Advantage | -2.886*** | -1.196*** | -15.649*** |
| | (0.453) | (0.192) | (1.358) |
| Product-Time FE | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes |
| CN Distance FE | Yes | Yes | Yes |
| $R^2$ | 0.06 | 0.06 | 0.06 |
| # Obs. | 1332970 | 1332970 | 1332970 |

Results from OLS estimation of (27). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

Table D.9: Product Introduction and Advantage

| | Dependent Variable: Dummy=1 for Product Introduction | | |
|---|---|---|---|
| | (CES) | (DGKP) | (CN) |
| Lagged Advantage | -1.700*** | -0.650*** | -1.712*** |
| | (0.343) | (0.150) | (0.107) |
| Product-Time FE | Yes | Yes | Yes |
| Firm-Time FE | Yes | Yes | Yes |
| Relatedness = | 0 | 0 | 0 |
| $R^2$ | 0.06 | 0.06 | 0.06 |
| # Obs. | 1173802 | 1173802 | 1173802 |

Results from OLS estimation of (27). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%.

Table D.10: Export Shocks and Product Adoption

| | Dependent Variable: Dummy=1 for Product Introduction | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (CES) | (CES) | (DGKP) | (DGKP) | (CN) | (CN) | (RCA) | (RCA) |
| Lagged Advantage | -53.495 | 430.422 | 48.417* | 125.458** | 6.846 | 17.724 | -112.107*** | -26.161 |
| | (114.169) | (266.203) | (29.045) | (55.870) | (8.115) | (16.099) | (6.531) | (145.616) |
| Lagged Advantage | -6.754 | -48.850** | -5.755** | -12.347*** | -1.022 | -1.963 | -5.474*** | -173.642 |
| X Exports | (9.801) | (23.082) | (2.455) | (4.759) | (0.687) | (1.387) | (1.741) | (284.484) |
| Firm-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm-Product FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Prod-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| # Obs. | 16266 | 16266 | 16266 | 16266 | 16266 | 16266 | 16266 | 16266 |
| F-Stat | | 1882.60 | | 3624.34 | | 2964.54 | | 0.68 |

Results from OLS and IV estimation of (28). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%. We instrument $\text{Exports}_{\nu t} \times \ln\left(1+\tilde{a}_{\omega,\nu,t-1}\right)$ with $\text{Exports IV}_{\nu t} \times \ln\left(1+\tilde{a}_{\omega,\nu,t-1}\right)$ where $\text{Exports IV}_{\nu t} = \text{Log} \sum_{k \neq DNK} \text{Exports}_{k\nu t}$ is the total exports of all countries except Denmark. F-Stat denotes the Kleibergen-Paap (KP) rank Wald F-statistic.

Table D.11: Export Shocks - IV = All Countries (First Stage)

| | Dependent Variable: Lagged Advantage X Exports | | | |
|---|---|---|---|---|
| | (CES) | (DGKP) | (CN) | (RCA) |
| Lagged Advantage X Exports IV | 0.788*** | 0.797*** | 0.777*** | 0.035*** |
| | (0.003) | (0.003) | (0.004) | (0.002) |
| Firm-Time FE | Yes | Yes | Yes | Yes |
| Firm-Product FE | Yes | Yes | Yes | Yes |
| Prod-Time FE | Yes | Yes | Yes | Yes |
| # Obs. | 16266 | 16266 | 16266 | 16266 |
| $R^2$ | 1.00 | 1.00 | 1.00 | 0.92 |

Results from OLS estimation of the first stage of (28). Clustered standard errors in parenthesis. Cluster: product-year. ***: significant at 99%, ** at 95%, * at 90%. $\text{Exports IV}_{\nu t} = \text{Log} \sum_{k \neq DNK} \text{Exports}_{k\nu t}$.

# E Statistical Summary of Fixed Cost Classifications

Table E.1: Fixed-cost thresholds used in the counterfactuals

| | Aggregate | | Simple average across clusters | |
|---|---|---|---|---|
| Fixed-cost threshold | Sales Share | Product Share | Sales Share | Product Share |
| *Panel A: CES* | | | | |
| Minimum profit | 0.213 | 0.488 | 0.181 | 0.482 |
| Producer correction | 0.373 | 0.127 | 0.371 | 0.174 |
| Cluster correction | 0.414 | 0.385 | 0.449 | 0.344 |
| *Panel B: DGKP* | | | | |
| Minimum profit | 0.095 | 0.129 | 0.143 | 0.142 |
| Producer correction | 0.091 | 0.036 | 0.140 | 0.075 |
| Cluster correction | 0.814 | 0.835 | 0.718 | 0.783 |

*Notes:* The table reports the share of sales and the share of products for which the final fixed-cost threshold is determined by each step of the procedure. The first threshold is the minimum operating profit for each product. The second threshold raises the product fixed cost using the largest firm-level minimum operating profit among producers of the product when counterfactual participation violations remain. The third threshold applies the final cluster correction using firm-level minimum operating profits from potential entrants in the product-space cluster. Aggregate columns weight products by sales or product counts. The simple-average columns report the mean across product-space clusters.

# F Advantage Normalization Validation

In Section 6.1, we recover counterfactual advantages in levels by normalizing the relative counterfactual advantage of each firm using an observed product in its portfolio. For each firm we select the product produced by the largest number of firms and choose $\widehat{\varphi}_{\omega,t}$ so that the counterfactual advantage for this reference product equals its observed advantage. We choose a widely produced product to help stabilize the normalization.

Here we provide a validation of this normalization. For firm-product observations for which advantage is observed, we compare observed advantage $a_{\omega,\nu,t}$ with the counterfactual advantage obtained using the same normalization as in Section 6.1. We estimate

$$\ln a_{\omega,\nu,t} = \beta \ln \widehat{a}^{cf}_{\omega,\nu,t} + \rho_{c,t} + \epsilon_{\omega,\nu,t}, \tag{F.1}$$

where $\rho_{c,t}$ denotes cluster-year fixed effects. A positive estimate of $\beta$ implies that firm-products predicted to have higher counterfactual advantage indices also have higher observed advantage indices. Panel A of Table F.1 reports this relationship for observed firm-product pairs. Using the baseline log product space, the estimates are significant at 0.255 under CES and 0.407 under DGKP. The relationship is similar when constructing the product space using advantages in levels or the inverse hyperbolic sine transformation.

Panel B provides a validation based on product introductions. We restrict the sample

to firm-products that are subsequently introduced and relate their observed advantage when produced to the counterfactual advantage predicted before introduction. Specifically, we replace the contemporaneous counterfactual advantage in Equation (F.1) with its lag. Under the baseline log product space, the estimated coefficient is 0.094 under CES and 0.084 under DGKP. Both are statistically significant at the 1% level. Similar results obtain using the alternative transformations. Thus, the normalization used to recover counterfactual advantages in levels is positively associated with observed advantage both for contemporaneously observed firm-products and for introduced products.

Table F.1: Observed and Counterfactual Advantage: Reference-Product Normalization

| | Dependent Variable: Observed Absolute Advantage | | | | | |
|---|---|---|---|---|---|---|
| | Log | | Linear | | IHS | |
| | (CES) | (DGKP) | (CES) | (DGKP) | (CES) | (DGKP) |
| *Panel A: Contemporaneous Counterfactual Advantage* | | | | | | |
| Counterfactual Advantage | 0.255*** | 0.407*** | 0.284*** | 0.388*** | 0.278*** | 0.402*** |
| | (0.009) | (0.005) | (0.008) | (0.005) | (0.009) | (0.005) |
| Cluster-Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.21 | 0.50 | 0.27 | 0.49 | 0.24 | 0.50 |
| # Obs. | 52995 | 52995 | 52995 | 52995 | 52995 | 52995 |
| *Panel B: Product Introductions* | | | | | | |
| Lagged Advantage | 0.094*** | 0.084*** | 0.089*** | 0.065*** | 0.091*** | 0.077*** |
| | (0.018) | (0.021) | (0.019) | (0.020) | (0.019) | (0.021) |
| Cluster-Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.08 | 0.07 | 0.08 | 0.07 | 0.08 | 0.07 |
| # Obs. | 3706 | 3706 | 3706 | 3706 | 3706 | 3706 |

*Notes:* The table reports log–log regressions of observed absolute advantage on counterfactual absolute advantage, corresponding to the normalization used in Section 6.1. Panel A uses contemporaneous advantage. Panel B restricts the sample to product introductions and uses lagged advantage. Log, Linear, and IHS denote the transformation used to construct the product space. All regressions include cluster-year fixed effects. Standard errors clustered at the firm level are reported in parentheses. *** significant at 99%, ** at 95%, * at 90%.