The Perfect Match:

Assortative Matching in Mergers and Acquisitions.

Maria Guadalupe — INSEAD
Veronica Rappoport — LSE
Bernard Salanié — Columbia
Catherine Thomas* — LSE

November 22, 2024

Abstract

We interpret M&A deals in Western Europe during the 2010s as the equilibrium of a matching model. Merger surplus arises from complementarities between multiple firm pre-merger characteristics. Large, productive firms prefer to merge with similarly productive but smaller partners, suggesting positive complementarity in productivities and negative cross complementarity between productivity and scale. We use post-merger data to show that estimated complementarities are strong predictors of merged firm performance. Our results inform the empirical relevance of different theories of mergers.

*E-mail: maria.guadalupe@insead.edu, v.e.rappoport@lse.ac.uk, bsalanie@columbia.edu, c.m.thomas@lse.ac.uk.
We are grateful to Alfred Galichon, Chad Syverson, Jon Vogel, and seminar participants at Sciences Po, ESCP, HBS, HEC, IFN, Nottingham, the NBER, Yale, and LSE for helpful comments. We thank Vaidehi Paramasweran and Siqin Ding for excellent research assistance and Paul Horsler for helpful advice about data availability. All errors are our own.

1 Introduction

Merger and acquisition activity redraws firm boundaries, bringing different firms' assets and operations under joint management. This market for corporate control is substantial. Worldwide, the annual value of M&A transactions is consistently around USD 4.3 trillion—over 4% of global GDP. Different economic theories explaining mergers have two features in common, the owners of each party anticipate being at least as well-off post-merger as they were before and are free to decide with whom (and whether) to merge. These features provide the starting point of this paper: We interpret observed M&A as the endogenous equilibrium outcome of firms competing in partner choice and engaging in voluntary mergers. The sources of merger surplus to be divided between merging parties arise from complementarities between firms' pre-merger characteristics.

We develop and estimate a novel matching model that accounts for key features of the global merger market. Asserting that observed M&A deals are the equilibrium (the stable matching) of the model allows us to estimate the sources of anticipated merger surplus using data on who merges with whom. Each firm's observed choice of merger partner reflects trade-offs: how potential partners' characteristics complement their own, and the transfers they have to pay to support the equilibrium. These transfers reflect the desirability and scarcity of firms with different characteristics. This structural approach has two advantages. First, as in Akkus, Cookson, and Hortacsu (2016), it explicitly incorporates the endogenous selection of partners, which is very challenging in reduced form estimates of merger gains in the absence of credible instruments.[1] The second advantage is that it allows us to run counterfactuals; we will show that the merger surplus function that rationalizes the observed mergers has predictive power for post-merger outcomes.

The data come from Moody's Zephyr database. We focus on bilateral deals that were completed during the eleven-year period 2008 to 2018 among firms headquartered in one of six large Western European countries—Belgium, France, Germany, Italy, Spain and the UK. The data include deals that feature both public and private firms. We investigate the roles played by two discrete pre-merger firm characteristics, industry and country, and two continuous characteristics that together determine firm revenues, revenue productivity (TFPr) and input scale (a production function coefficient-weighted aggregate of the firm's assets and employment).

[1] Akkus, Cookson, and Hortacsu (2016) study M&A activity in the U.S. banking sector, and show the role of the size and overlap of branch networks.

In our results, the primary motive for merging is complementarity between merger partners' pre-merger productivities. On the other hand, larger firms benefit less from matching with productive counterparts. These results are driven by within-industry deals, which account for 75% of the deals in our sample, and are similar for domestic and cross-border deals. Matching patterns are less well explained by complementarities in observable characteristics for cross-industry deals. We also show that complementarity between the input scales of the merging firms plays a smaller role in anticipated merger gains; it is slightly positive for same-industry deals and slightly negative for different-industry deals.

The complementarity between revenue productivities that we find could be related to firm efficiency or markups. To further investigate the motives behind the mergers, we follow the approach in De Loecker, Eeckhout, and Unger (2020) to estimate firm-specific markups, and then further decompose revenue productivity into quantity productivity (TFPq) and markup. Our findings suggest that merging firms expect technological productivity (TFPq) to be the main source of complementarity in revenue productivity, especially for same-industry deals. Moreover, the explanatory power of the matching model substantially improves when firms are characterized by TFPq. We also find positive assortative matching in markups.[2] It is worth noting, however, that the deals we observe took place within the regulatory remit of EU and national-level competition authorities. Any proposed mergers that were clearly anti-competitive may well have been blocked preemptively.

To validate our empirical approach, we examine the post-merger performance of the firms in our data. If the parameters of the matching model are informative about firms' merger motives, and if firms are able to realize some of the merger gains they anticipated, we expect to see a positive correlation between the surplus function and the performance of merged firms vis-à-vis a no-merger counterfactual. To test this, we construct counterfactual estimates for firm-level post-merger revenues, employment, and total assets had the firms not merged. We observe large variation in the evolution of the merged entity relative to the no-merger counterfactual. Some of this variation is explained by predicted merger surplus: we find that the merger premia in terms of revenues, input scale, and productivities are positively correlated with the estimates of merger surplus delivered

[2] In industry-level models, the equilibrium price post-merger is increasing in the pre-merger market power (Williamson, 1968, Perry and Porter, 1985, Nocke and Whinston, 2022).

by the matching model. These correlations increase monotonically over time, possibly because it takes time for the surplus gains to materialize, and the performance benefits are larger for the firm designated as the acquirer in the data. An estimated merger surplus that is higher by one standard deviation gives predicted log revenues after four years that are 10% higher than in the counterfactual, equivalent to around one fifth of the standard deviation in this variable.

Our matching model builds on the pioneering contribution by Choo and Siow (2006), which introduces a tractable model of matching on unobservable as well as observable characteristics, and where the relative probabilities of different mergers are unaffected by the presence of the outside option of not merging.[3] To allow for both continuous and discrete firm characteristics, we apply methods from Dupuy and Galichon (2014). Our estimation method adapts the moment matching estimator proposed in Galichon and Salanié (2022), as well as their iterative proportional fitting procedure (IPFP) algorithm to compute the equilibrium.[4]

Most of this literature estimates matching models of marriage, as in the canonical theory set out by Becker (1973). Since a large majority of marriages consist of two partners of different genders, most empirical work on marriage models matching as *bipartite*, in the jargon of the field.[5] Our work instead conceives of merging firms as *roommates*.[6] In particular, we do not use the "acquirer" and "target" labels in our data as a pre-existing distinction. We take the stance that the identity of the acquirer, and the target, in a merger is an endogenous outcome that maximizes the joint surplus. We show that since in our data acquirers are on average larger than target firms, imposing a bipartite market structure would over-estimate the extent of positive assortative matching in input scale.

The classic *q-theory* of mergers in Jovanovic and Rousseau (2002), and in Nocke and Yeaple (2007) in the context of cross-border M&A, models merged firms as the combination of firms' pre-merger characteristics, as we do. In this framework, the assumption about technology trans-

[3]As our specification builds on Choo and Siow (2006)'s logit structure, it inherits a version of the usual independence of irrelevant alternatives property. See Galichon and Salanié (2019). This property is useful in our setting in that it allows us to estimate the relative importance of different sources of merger surplus from data on observed mergers without also requiring data on firms that opt to remain standalone.

[4]Fox (2010, 2018) introduced alternative methods that have been applied to mergers by Akkus, Cookson, and Hortacsu (2016) and to inter-firms alliances by Mindruta, Moeen, and Agarwal (2016).

[5]For a recent exception, see Ciscato, Galichon, and Goussé (2020).

[6]The name comes from Gale and Shapley (1962)'s seminal paper: "An even number of boys wish to divide up into pairs of roommates" (Example 3, page 12).

mission is that a merger allows a high-productivity firm to leverage its superior technology within the combined size of the merged corporation. The resulting matching equilibrium of a merger surplus function in this theory results in a "high-buys-low" productivity pattern and positive cross-characteristic complementarity between input scale and productivity. Our results contradict these two predictions: we find that more productive firms are more likely to match with smaller, and also productive, counterparts.

The existing empirical literature studying selection into mergers often finds "like buys like", as in our results on productivity. However, much of this literature characterizes firms along a single dimension of heterogeneity—for example, in market-to-book ratios (Rhodes-Kropf and Robinson, 2008), quantity-based productivity (Braguinsky, Ohyama, Okazaki, and Syverson, 2015), profitability (David, 2021), and patenting activity (Ozcan, 2015). The assumption of one-dimensional heterogeneity arises naturally from theories of the firm where the scale of production adjusts to its optimal level and size (Hopenhayn, 1992)—and where revenues, profitability and productivity all signal the same underlying ranking across firms. In our sample of merging firms, however, revenue productivity and size are negatively correlated.[7] There is strong evidence in our data that the mergers we study reflect matching on both dimensions.[8]

The positive assortative matching in productivities that we find suggests that technology transmission within a merged firm depends on both firms' original productivities. This feature coincides with the finding in Bilir and Morales (2020) for the case of foreign affiliates within a multinational corporation: new technologies in the headquarters increase the productivity of affiliates if the affiliate is also innovative.[9] It is also consistent with Javorcik and Arnold (2009) and Guadalupe, Kuzmina, and Thomas (2012), who show that already high-performing local firms are acquired by multinational parents. We establish, more generally, that the value of expansion via M&A depends on the existing capabilities of both firms involved.

[7]Input scale and revenue productivity are also negatively correlated in the ORBIS data for all firms for the six countries studied over this period. We estimate the production function coefficients used to decompose revenue into input scale and productivity using these data on all firms.

[8]The fact that we don't find much matching on input scale also undermines the idea that firms in our sample were merging purely for the sake of empire building. See Roll (1986), Jensen (1986), and Shleifer and Vishny (1989) for an analysis of the empire building motive for M&A.

[9]This finding relates to Cohen and Levinthal (1990), who define a firm's absorptive capacity as its prior knowledge. Mindruta, Moeen, and Agarwal (2016) discuss knowledge complementarities in research alliances.

Considered together, our results suggest that anticipated merger gains are largest for mergers between more productive firms, and it is these merged entities that perform well ex post. Exploring the microfoundations of how firm characteristics are combined to generate surplus post merger is a compelling topic for future work.

The rest of the paper proceeds as follows. Section 2 describes the empirical setting and the data used. Section 3 presents the matching model, the estimation framework, the identification strategy, and the computation approach. The more technical details are given in the Appendix. Section 4 presents the main results describing the estimated merger surplus function. Section 5 compares the estimated anticipated merger gains with the merger gains observed in a subsample of firms. Section 6 concludes.

2 Empirical Setting and Data

We study 2,819 two-firm mergers that occurred between 2008 and 2018, involving firms headquartered in one of six large Western European countries, Belgium, France, Germany, Italy, Spain, and the UK. The data come from two Moody's databases: Historical Orbis and Zephyr. One advantage of these data relative to commonly-used datasets such as Compustat in the US is that non-listed, private firms are included, which is made possible by European financial reporting requirements.

Historical Orbis provides annual balance sheet data for firms in the six European countries under study. We include both manufacturing and services firms in the sample.[10] From this dataset, we extract information on revenues, employment, total assets, and cost of materials.

The nominal variables are deflated using the EUKLEMS INTANProd (2021) price indices. Following De Ridder, Grassi, Morzenti et al. (2021), we deflate revenues using the gross output index, and material costs and costs of goods sold with the intermediate input index, for the relevant country and NACE Rev2 2-digit industry classification. Finally, we deflate total assets and the cost of employees with the total GDP index.[11]

[10] We exclude firms in financial service or the public sector, i.e. banks, insurance companies, public administration and defence. We also exclude firms in the extraction of crude petroleum and natural gas (NACE Rev. 2, two-digit code 06) because the industry-specific productivity coefficients were inconsistent with some of the structure imposed in later parts of the paper.

[11] The price indices run up to 2018 for Belgium and Germany. For the other countries, we extrapolate for 2018 (for 2017 and 2018 for the UK) based on past industry-level trends and the contemporaneous total GDP index.

Zephyr reports the details of M&A deals in Europe. From this database, we use the identity of the two participating firms, which we match with the firm-level data in Historical Orbis for the year prior to the merger. Zephyr also reports the stakes involved in the transaction; we limit the sample to deals that led to a change in the controlling owner of one of the merging parties: the initial stake was less than 50%, and the final stake was more than 50%. Zephyr identifies the acquirer and target firm in the transaction, labeling as the Acquirer the firm that has gained the controlling stake in the other. As we show in the next section, firms it labels as Targets are on average smaller than Acquirers.

2.1 Productivity and Input Scale

The matching model that we describe in more detail in the following section is estimated from variation in the relative frequency of observed mergers between firms with different observable characteristics. For computational reasons, we will characterize firms by two discrete dimensions, country and industry, and two continuous characteristics, input scale and productivity. In this subsection, we explain our methodology for estimating these continuous firm characteristics.

For each NACE 2-digit industry code, we use the Levinsohn and Petrin (2003) approach to estimate the following production function:

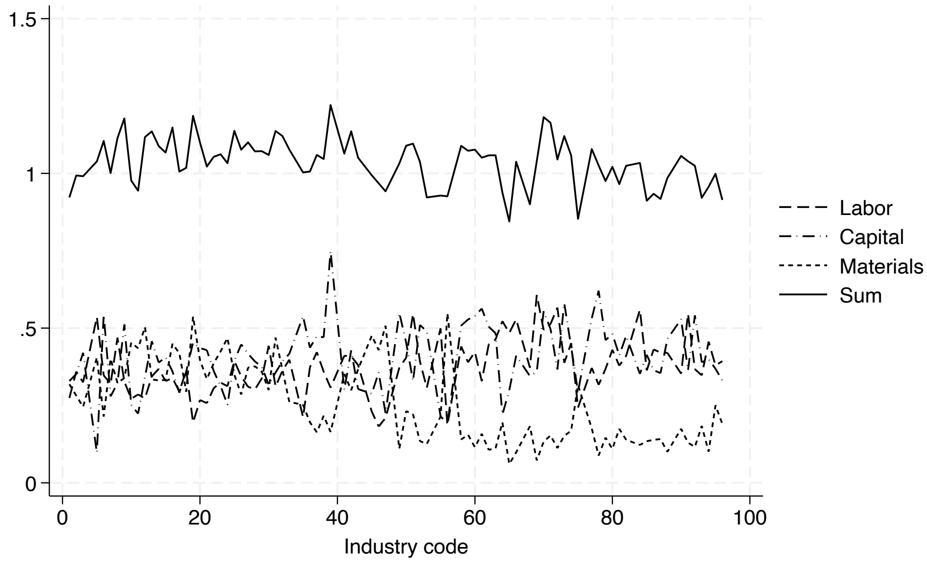
$$r_{jt} = z_{jt} + \alpha_L l_{jt} + \alpha_M m_{jt} + \alpha_K k_{jt}, \tag{1}$$

where r_{jt} is the log of revenues of firm j at time t, l_{jt} is the log of the number of employees, m_{jt} is the log of material costs, and k_{jt} is the log of total assets. The estimated coefficients α_L, α_K, and α_M are shown in Figure 1.[12] Using these estimates, we recover the firm (log) revenue-productivity z_{jt} and the (log) input scale, $\Omega_{jt} = \hat{\alpha}_L l_{jt} + \hat{\alpha}_M m_{jt} + \hat{\alpha}_K k_{jt}$.

Only a small share of all firms in the Historical Orbis database participate in bilateral M&A activity, as reported in Zephyr. Figure 2 shows the revenues, input scale, and revenue productivity (TFPr) distributions of firms that merge and those that do not merge. Target firms, and especially acquiring firms, exhibit significantly higher revenues compared to the average firm in their respective industry and country. David (2021) shows an analogous fact for firms' profitability. However, this

[12] We estimate the production function for UK firms separately, as they report COGS (cost of goods sold) rather than cost of materials.

Figure 1: Production Function Coefficients



Note: The coefficients shown here are the estimates for Belgium, France, Italy and Spain. The number of firm-year observations varies per industry classification (NACE Rev.2 two-digits). The mean number of observations per two-digit industry is $133,629$.

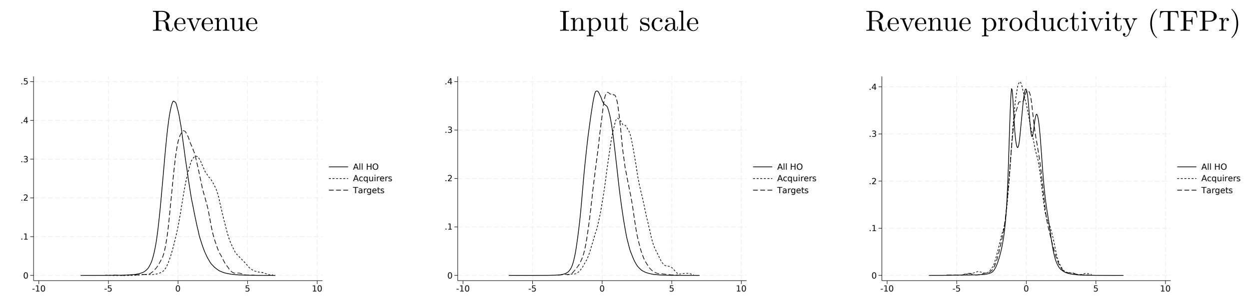
difference is mostly explained by input scale in our sample. Firms that participate in M&A are not more productive, as measured by TFPr, than the average firm in their corresponding market and acquiring firms do not have higher TFPr than targets, similar to the results in Braguinsky et al. (2015). However, we know that revenue productivity z is an imperfect measure of productivity. As Foster, Haltiwanger and Syverson (2008) show, it typically correlates positively with true efficiency (TFPq) but also with prices and markups.

To decompose revenue productivity, z_{jt}, into firm-level physical productivity and markup we follow De Loecker, Eeckhout, and Unger (2020) and estimate the firm's markup, μ_{jt}, using the cost minimization condition:

$$\mu_{jt} = \alpha_L \frac{Rev_{jt}}{W_t L_{jt}} \tag{2}$$

where Rev_{jt} and $W_t L_{jt}$ are revenues and labor cost, respectively, and α_L is the labor coefficient in the production function in equation (1). De Ridder et al. (2021) show that although the markups estimated from revenue data do not consistently recover the actual markup level, they successfully capture their ranking across firms. This serves our purpose: we are interested in capturing the *attractiveness* of matched firms relative to the pool of available candidates. Therefore, throughout

Figure 2: Merged and Non-Merged Firm Standardized Characteristics



Note: Variables in logs, standardized by country (BE, DE, ES, FR, IT, and UK) and broad industry. We consider five industries that aggregate various : basic manufacturing, sophisticated manufacturing, high-tech services, non high-tech services, and other. All HO pools all firm-year observations in Historical Orbis. Acquirer and Target describe the merging firms in the year prior to the deal.

this paper, firms' continuous characteristics are standardized relative to the firm's country and industry of operation.

Finally, we construct the following continuous variable to capture the firm's physical productivity:

$$\ln \text{TFPq}_{jt} = z_{jt} - \ln \mu_{jt} \tag{3}$$

where μ_{jt} is firm j's markup in year t.

Table 1 shows the correlation between the above computed variables for all firms in the six mentioned countries, pooled over the sample years (Panel A for all firms in the Historical Orbis dataset, Panel B for the firms in our sample). Not surprisingly, input scale and revenues are highly correlated. On the other hand, revenue productivity TFPr and total revenues are virtually uncorrelated in our data. This zero correlation is the result of two opposite forces: markups, which are higher for larger firms, are positively correlated with revenues, while quantity productivity TFPq is lower for larger firms and is negatively correlated with revenues.[13]

Figure 3 shows the distributions of markups and TFPq for merging firms (acquirers and targets) and for firms that did not merge. Acquiring firms have a higher markup than targets and than non-merging firms, but they have somewhat smaller TFPq. However, these difference are dwarfed by the differences in scale mentioned earlier.

[13] As in Foster, Haltiwanger, and Syverson (2008), TFPr and TFPq are positively but imperfectly correlated in our firm-level data.

Table 1: Firms' Characteristics - Covariance Matrix

Panel A: All firms (N = 10,096,516)

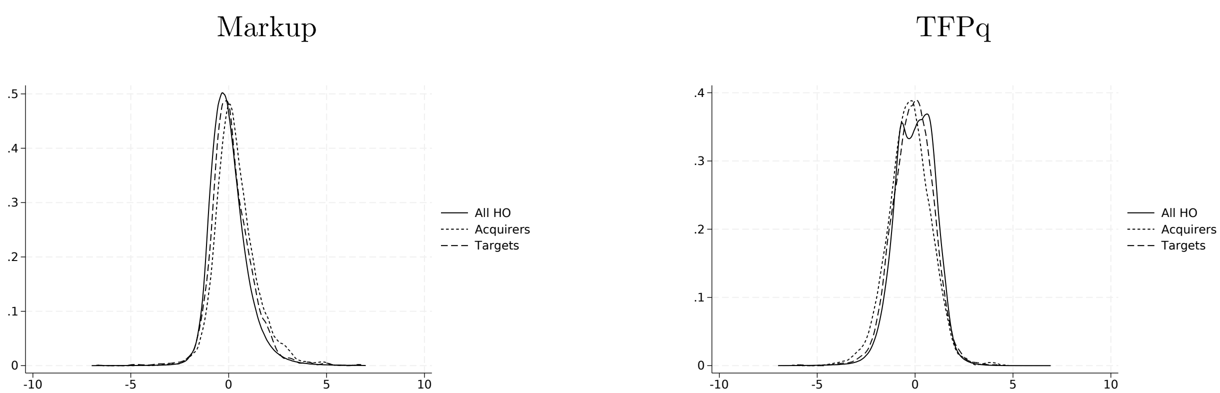
	Rev		Input-Scale		TFPr		Markup		TFPq	
	corr.	s.d.	corr.	s.d.	corr.	s.d.	corr.	s.d.	corr.	s.d.
Rev	1.00	-								
Input-Scale	0.77	(.0002)	1.00	-						
TFPr	0.06	(.0003)	-0.56	(.0003)	1.00	-				
Markup	0.41	(.0002)	0.20	(.0003)	0.20	(.0003)	1.00	-		
TFPq	-0.21	(.0003)	-0.65	(.0002)	0.75	(.0002)	-0.47	(.0003)	1.00	-

Panel B: Merged firms (N = 5,728)

	Rev		Input-Scale		TFPr		Markup		TFPq	
	corr.	s.d.	corr.	s.d.	corr.	s.d.	corr.	s.d.	corr.	s.d.
Rev	1.00	-								
Input-Scale	0.82	(.0075)	1.00	-						
TFPr	0.07	(.0132)	-0.47	(.0116)	1.00	-				
Markup	0.40	(.0121)	0.23	(.0129)	0.22	(.0129)	1.00	-		
TFPq	-0.22	(.0129)	-0.60	(.0106)	0.71	(.0093)	-0.50	(.0114)	1.00	-

Note: Variables in logs, standardized by country and industry. We consider five industries: basic manufacturing, sophisticated manufacturing, high-tech services, non high-tech services, and other. Panel A pools all firm-year observations in Historical Orbis. In Panel B, merged firms' characteristics correspond to the year prior to the deal.

Figure 3: Standardized Markup and Quantity Productivity



Note: Variables in logs, standardized by country and industry. We consider five industries: basic manufacturing, sophisticated manufacturing, high-tech services, non high-tech services, and other. Panel A pools all firm-year observations in Historical Orbis. In Panel B, merged firms' characteristics correspond to the year prior to the deal.

2.2 Merger Characteristics

Having described our continuous variables and the characteristics of firms that participate in M&A, we now move to the descriptive statistics of the observed matches. Table 2 summarizes the mergers in our data along the two discrete characteristics (country and industry); it gives the probabilities of selecting into domestic and within-industry mergers for firms by country and by industry. The total numbers are in the last row and last column of each panel. The country-level rows sum to 100% across country-partner columns in panel A, and the industry-level rows sum to 100% across industry-partner columns in panel B. In panel A, the large diagonal elements, presented in bold, reflect the large share of domestic mergers, that is, the within-country endogamy. Panel B shows the extent of within-industry endogamy. Domestic within-industry mergers are 68% of the total and domestic cross-industry mergers make up 21%. Cross-border within-industry mergers are 7% of total deals, and, finally, 3% are both cross-border and cross-industry .

Table 3 shows descriptive statistics for the firms participating in the four different types of mergers (cross-border or domestic, across- or within-industry). Cross-border mergers involve firms with larger input scales and markups. Domestic mergers, on the other hand, tend to involve more productive firms.

3 Empirical Framework

We start by describing the matching model and our approach to identification and estimation as if we observed firms that choose to remain standalone. Subsection 3.5 will explain how we adapt our methods to the data we use, which reports only realized mergers.

3.1 Firm value and merger surplus

Each firm i consists of a bundle of observable and unobservable characteristics that are summarized in a triple $\{a_i, x_i, \tilde{\varepsilon}_i\}$. The first element $a_i = (C_i, I_i)$ is an ordered pair, indicating the country where the firm is located and the industry of its primary activity. The second element x_i is a two-dimensional vector of continuous variables $x_i = (z_i, \Omega_i)$, where z_i is a measure of productivity of the firm and Ω_i is its input scale. Our main results proxy productivity by revenue productivity. All market participants observe a and x. The third element of the triple, $\tilde{\varepsilon}_i$, is a random variable

Table 2: Endogamy on Discrete Characteristics

Panel A: By Country

	Merging with firm in: Belgium	Germany	Spain	France	UK	Italy	N Mergers
Belgium	**89**	3	5	9	6	6	118
Germany	4	**77**	12	11	23	17	144
Spain	0	2	**1,203**	5	3	10	1,223
France	14	8	39	**213**	18	14	306
UK	4	4	8	17	**605**	16	654
Italy	2	3	13	8	8	**340**	374
N Mergers	113	97	1,280	263	663	403	2,819

Panel B: By Industry

	Merging with firm in: Basic Manuf.	Sophisticated Manuf.	Other	High-Tech Serv.	Non HT Serv.	N Mergers
Basic Manuf.	**222**	27	52	0	11	312
Sophisticated Manuf.	20	**479**	116	19	33	667
Other	32	69	**717**	14	63	895
High-Tech Serv.	1	7	21	**144**	45	218
Non HT Serv.	14	38	48	56	**571**	727
N Mergers	289	620	954	233	723	2,819

Table 3: Size and Productivity by Type of Merger

	N firms	Rev	Input-Scale	TFPr	TFPq	Markup
Cross-border, cross-industry group	162	21.8	16.7	5.1	5.6	0.5
		(1.9)	(2.5)	(1.6)	(1.8)	(0.9)
Domestic, cross-industry group	1,210	21.0	15.8	5.2	5.8	0.3
		(2.2)	(2.8)	(1.6)	(1.7)	(1.0)
Cross-border, within-industry group	422	22.0	17.0	5.0	5.5	0.3
		(1.9)	(2.6)	(1.7)	(1.8)	(0.8)
Domestic, within-industry group	3,844	21.0	15.6	5.4	6.0	0.3
		(2.0)	(2.7)	(1.6)	(1.8)	(1.0)

Note: Descriptive statistics of merged firms, computed the year prior to the merger. Variables in logs. Standard deviations in parenthesis.

that describes how firm i differs from other firms on dimensions that are observable to other market participants but not to us. The value of the firm depends on all three elements of the triple; we denote it $\Pi_{a_i}(x_i, \tilde{\varepsilon}_i)$.

We will often write "$i \in a$" to denote that $a_i = a$, and "$i \in O(a, x)$" if, moreover, $x_i = x$. We use the term "discrete characteristics" to refer to a_i and "continuous characteristics" for x_i. We denote A the number of values of the discrete characteristics a.

When firms i and j merge, the value of the new merged entity depends on the pre-merger characteristics of the two firms. If firm j has $a_j = b$ and $x_j = y$, then the value of the merged firm is $\Pi_{ab}(x, y, \tilde{\varepsilon}_i, \tilde{\varepsilon}_j)$. Therefore the merger gains, denoted $\tilde{\Phi}_{ab}(x, y, \tilde{\varepsilon}_i, \tilde{\varepsilon}_j)$, in percentage change, are given by

$$\tilde{\Phi}_{ab}(i, j) = \frac{\Pi_{ab}(x, y, \tilde{\varepsilon}_i, \tilde{\varepsilon}_j) - \Pi_a(x, \tilde{\varepsilon}_i) - \Pi_b(y, \tilde{\varepsilon}_j)}{\Pi_a(x, \tilde{\varepsilon}_i) + \Pi_b(y, \tilde{\varepsilon}_j)}. \tag{4}$$

That is, the merger generates positive gains if the new entity is more valuable than the sum of the pre-merger firm values. At this point, we impose some structure on how the unobservable (to us) characteristics $\tilde{\varepsilon_i}, \tilde{\varepsilon_j}$ of firms i and j enter the merger surplus function. We maintain the assumption made by Choo and Siow (2006) that the joint surplus function is independent of any interactions between the unobserved characteristics $\tilde{\varepsilon}_i$ and $\tilde{\varepsilon}_j$ conditional on the observed characteristics $\{a, x\}$ and $\{b, y\}$. This is known as the *separability* assumption. More precisely, we assume that there exists a matrix of functions $\mathbf{\Phi}$ and random vectors of functions $\boldsymbol{\varepsilon}^i$ such that in a match between firms $i \in O(a, x)$ and $j \in O(b, y)$, the joint surplus is

$$\tilde{\Phi}(i, j) = \Phi_{ab}(x, y) + \varepsilon^i_b(y) + \varepsilon^j_a(x). \tag{5}$$

By symmetry, $\Phi_{ab}(x, y) = \Phi_{ba}(x, y)$ and $\tilde{\Phi}(i, j) = \tilde{\Phi}(j, i)$. We denote $\tilde{\Phi}(i, 0) = \varepsilon^i_0$ for a firm that does not participate in a merger.

As Galichon and Salanié (2022) note, while restrictive, the separability assumption does allow for a limited form of "matching on unobservables" in that it allows surplus to arise from interactions between each firm's unobservable and the observables of its partner. It only rules out interactions between the unobservable characteristics of the two partners[14].

[14]By way of illustration, the education of CEOs is not recorded in our data. The separability assumption allows the surplus to be greater for a merger of a firm whose CEO studied at the LSE if the firm merges with any UK services firm. However, it rules out the possibility that the merger surplus is higher, controlling for all observable

3.2 Matching Equilibrium

The equilibrium consists of a feasible, stable matching, and a set of firm payoffs. A *matching* $\boldsymbol{\mu}$ is a specification of who merges with whom; we denote $\mu(i,j) = 1$ if i and j match, and $\mu(i,j) = 0$ if they don't. Note that by definition, $\mu(j,i) = \mu(i,j)$. *Feasibility* requires that each firm be matched to one partner at most (some firms may not merge): $\mu(i,0) + \sum_j \mu(i,j) = 1$. *Stability* requires that (i) no merging firm would prefer to opt out of the merger and (ii) no two firms would both prefer merging with each other to their observed outcome.

In equilibrium, any pair of merging firms must split the joint surplus. Each firm seeks the partner with whom its *net* surplus will be highest: the equilibrium payoff u_i of any firm i solves

$$u_i = \max_j (\tilde{\Phi}(i,j) - u_j),$$

where the maximization runs over all potential partners and the option not to merge (for which $\Phi(i,0) = 0$ and $u_0 = 0$). Firm i is matched in equilibrium to a firm j that achieves this maximum payoff in the equilibrium matching (or remains standalone if the optimum is achieved at $j = 0$).

We adapt the approach in Galichon and Salanié (2022)'s marriage market setting to fit the roommate problem.[15] We start by substituting equation (5) into the equilibrium payoff of firm $i \in O(a,x)$:

$$\begin{aligned} u_i &= \max_{b,y} \left(\Phi_{ab}(x,y) + \varepsilon_b^i(y) + \max_{j \in O(b,y)} \left(\varepsilon_a^j(x) - u_j \right) \right) \\ &= \max_{b,y} \left(\Phi_{ab}(x,y) + \varepsilon_b^i(y) - U_{ba}(y,x) \right), \end{aligned}$$

where we denote $U_{ba}(y,x) = \min_{j \in O(b,y)} (u_j - \varepsilon_a^j(x))$.

Suppose that firms $i \in O(a,x)$ and $j \in O(b,y)$ merge in a stable matching. Then the joint utility from the match must be split between the two firms: $u_i + u_j = \tilde{\Phi}(i,j)$. This yields

$$U_{ab}(x,y) + U_{ba}(y,x) = \Phi_{ab}(x,y). \tag{6}$$

---

characteristics, when the CEOs of both merging firms attended the LSE.

[15] Appendix A.1 demonstrates the potential bias in surplus estimates when modeling bilateral roommate selection as bipartite.

While $\Phi_{ab}(x,y)$ is symmetric, $U_{ba}(y,x)$ will normally differ from $U_{ab}(x,y)$: in any stable matching, the joint surplus generated by the observable characteristics is split between merging firms according to the value the market assigns to each of their observable characteristics. As always, this reflects both the relative scarcity and the relative desirability of these observables.

A simple extension of the argument in Shapley and Shubik (1971) shows that the equilibrium matching $\mu(i,j)$ maximizes the total joint surplus over the set of feasible matchings $\mathcal{M}$:

$$\mathcal{W} = \max_{\mu \in \mathcal{M}} \left(\sum_{i,j} \mu(i,j)\tilde{\Phi}(i,j) + \sum_{i} \mu(i,0)\tilde{\Phi}(i,0) \right) \tag{7}$$

where feasibility imposes that $\boldsymbol{\mu} \geq 0$; $\boldsymbol{\mu}$ is symmetric; and for each i, $\mu(i,0) + \sum_j \mu(i,j) = 1$. The symmetry constraint is specific to the roommate problem. While it may lead to the non-existence of a stable matching in general, Chiappori, Galichon, and Salanié (2019) show that this difficulty vanishes in large populations.

Our primary object of interest is the relationship between merger surplus and the observable pre-merger characteristics of the merging firms, as summarized in the function $\Phi_{ab}(x,y)$ in equation (5). We estimate features of this relationship by assuming that the observed matching $\hat{\boldsymbol{\mu}}$ is the equilibrium matching, and finding the functions $U_{ab}(x,y)$ such that the surplus function $\boldsymbol{\Phi}$ defined by equation (5) comes closest to generating $\hat{\boldsymbol{\mu}}$ in equilibrium.

We begin by partitioning firms based on observables. Let N_a be the number of firms with discrete characteristics a, and $f(x|a)$ be the density of the continuous characteristics of these firms. We denote μ_{ab} the number of mergers between firms with discrete characteristics a and b. Note that by construction, $\mu_{ab} = \mu_{ba}$.

Roommate matching introduces a complication: each (a,a) match has two firms with discrete characteristics a. To avoid distinguishing the two cases $a = b$ and $a \neq b$, it is convenient to define the pseudo-matching

$$\nu_{ab} = \mu_{ab} \times (1 + \mathbf{1}(a = b)).$$

With this notation,

$$N_a = \sum_{\substack{b=1 \\ b \neq a}}^{A} \mu_{ab} + 2\mu_{aa} = \sum_{b=1}^{M} \nu_{ab}.$$

On the other hand, we need not be concerned about $x = y$ since that happens with probability zero. Since all specifications we estimate in the paper include continuous characteristics, we can neglect the case $(a, x) = (b, y)$. Let $\mu(x, y|a, b)$ denote the joint density of the continuous characteristics in matches between a firm in a and a firm in b. By symmetry, $\mu(x, y|a, b) = \mu(y, x|b, a)$.

A pseudo-matching is feasible iff for all (a, x):

$$\sum_{b=1}^{A} \int \mu_{ab}(x, y)\ dy = N_a f(x|a). \tag{8}$$

We now choose distributions for the error terms in equation (5) in order to write down the choice probabilities for each firm i. Since x and y are continuous variables, we must depart from the standard apparatus of discrete choice models. We follow Dagsvik (1994) and Dupuy and Galichon (2014) and assume that

- each firm i selects a partner $(b, y, \varepsilon_b^i(y))$ from the points of a Poisson process with intensity $(1/M) \times dy \times \exp(-\varepsilon) d\varepsilon$;
- these Poisson processes are independently and identically distributed across firms i.

The Poisson specification can be seen as an analog of the multinomial logit model for continuous choice. Note that it is standardized, which implicitly sets the scale of surplus and utilities.

Under these assumptions, a firm $i \in O(a, x)$:

- remains a standalone firm with probability

$$\mu(\emptyset|a, x) = \frac{1}{1 + \sum_{d=1}^{A} \int \exp(U_{ad}(x, y)) dy};$$

- merges with a firm with discrete characteristics b with probability

$$\mu(b|a, x) = \frac{\int \exp(U_{ab}(x, y)) dy}{1 + \sum_{d=1}^{A} \int \exp(U_{ad}(x, y)) dy};$$

- and conditional on merging with a firm with discrete characteristics b, firm i merges with a firm that has continuous characteristics y with probability

$$\mu(y|a, x, b) = \frac{\exp(U_{ab}(x, y))}{\int \exp(U_{ab}(x, t)) dt},$$

so that

$$\mu(b,y|a,x) = \frac{\exp(U_{ab}(x,y))}{1+\sum_{d=1}^{A}\int \exp(U_{ad}(x,y))dy}.$$

The average expected equilibrium payoff of the firms in $O(a,x)$ is

$$u_a(x) = \log\left(1+\sum_{b=1}^{A}\int \exp(U_{ab}(x,y))dy\right).$$

3.3 Solving for the Equilibrium

In equilibrium, the pseudo-matching probabilities must be symmetric: $\mu_{ab}(x,y)$ must equal $\mu_{ba}(y,x)$ for all (a,b,x,y). This gives

$$N_a f(x|a)\frac{\exp(U_{ab}(x,y))}{1+\sum_{d=1}^{A}\int \exp(U_{ad}(x,t))dt} = N_b f(y|b)\frac{\exp(\Phi_{ab}(x,y)-U_{ab}(x,y))}{1+\sum_{c=1}^{A}\int \exp(\Phi_{cb}(z,y)-U_{cb}(z,y))dz}. \quad (9)$$

This system of equations in $\boldsymbol{U} = (U_{ab}(x,y))$ (or equivalently in $\boldsymbol{\mu} = (\mu_{ab}(x,y))$) has a solution, which is unique; the proof follows the logic in Dupuy and Galichon (2014). Note that taking logarithms, equation (9) can be rewritten as

$$\log \mu_{ab}(x,y) = \log\left(N_a f(x|a)\right) + U_{ab}(x,y) - u_a(x) = \log\left(N_b f(y|b)\right) + U_{ba}(y,x) - u_b(x),$$

so that

$$\log \frac{\mu_{ab}^2(x,y)}{N_a N_b f(x|a) f(y|b)} = \Phi_{ab}(x,y) - u_a(x) - u_b(y).$$

In this form, it is clear that we only need to solve for functions $e_a(x)$ that solve the system

$$\begin{aligned}\mu_{ab}(x,y) &= \exp(\Phi_{ab}(x,y)/2)\; e_a(x)e_b(y) \;\text{ for all }\; (a,b,x,y)\\ \sum_{b=1}^{A}\int \mu_{ab}(x,y)dy &= N_a f(x|a) \;\text{ for all }\; a,x.\end{aligned} \quad (10)$$

This can be done efficiently using a variant of the Iterative Proportional Fitting Procedure, which is described in Appendix A.3. Once we know the functions $e_a(x)$, we can compute $\mu_{ab}(x,y)$ and the utilities

$$u_a(x) = \log\left(N_a f(x|a)\right) - 2\log e_a(x)$$
$$U_{ab}(x,y) = \Phi_{ab}(x,y)/2 + \log e_a(x) - \log e_b(y).$$

The total surplus $\mathcal{W}$ in equation (7) is the sum of all individual utilities:

$$\mathcal{W}(\boldsymbol{U}) = \sum_{a=1}^{A} N_a \int f(x|a) u_a(x) dx = \sum_{a=1}^{A} N_a \int f(x|a) \log\left(1 + \sum_{b=1}^{A} \int \exp(U_{ab}(x,y)) dy\right) dx \quad (11)$$

which is a globally convex function of $\boldsymbol{U}$.

3.4 The Moment Matching Estimator

Our application uses the moment matching estimator proposed in Galichon and Salanié (2022). We specify $\Phi_{ab}(x,y)$ as a linear combination of known *basis functions* $(\phi^k_{ab}(x,y))_{k=1}^K$, with unknown weights λ_k:

$$\Phi^{\lambda}_{ab}(x,y) = \sum_{k=1}^{K} \lambda_k \phi^k_{ab}(x,y). \quad (12)$$

We will estimate the vector $\boldsymbol{\lambda}$ by fitting the observed first moments of the basis functions. To do this, we minimize the function

$$\mathcal{W}(\boldsymbol{U}^{\lambda}) - \frac{\boldsymbol{\lambda}}{2} \cdot \sum_{a,b=1}^{A} \int\int \phi_{ab}(x,y) \hat{\mu}_{ab}(x,y) dx dy \quad (13)$$

over the parameter vector $\boldsymbol{\lambda}$, where $\boldsymbol{U}^{\lambda}$ solves the system (9) for $\boldsymbol{\Phi} = \boldsymbol{\Phi}^{\lambda}$; $\boldsymbol{\phi}_{ab}(x,y)$ is the vector of basis functions $\phi^k_{ab}(x,y)$ for each $\{a,b,x,y\}$; and $\hat{\boldsymbol{\mu}}$ is the observed pseudo-matching.

The first-order conditions with respect to $\boldsymbol{\lambda}$ are

$$\frac{d\mathcal{W}}{d\boldsymbol{\lambda}}(\boldsymbol{U}^{\lambda}) = \frac{1}{2}\, E_{\hat{\boldsymbol{\mu}}}\ \phi_{ab}(x,y). \quad (14)$$

While computing the derivative of the social welfare with respect to $\boldsymbol{\lambda}$ may seem forbidding, in fact it is fairly straightforward. A simple adaptation of the arguments in Galichon and Salanié (2022) shows that the social welfare can be rewritten as

$$\mathcal{W}(\boldsymbol{U}^{\lambda}) = \max_{\boldsymbol{\mu}} \left(\frac{1}{2} \sum_{a,b=1}^{A} \int \int \mu_{ab}(x,y) \; (\boldsymbol{\lambda} \cdot \phi_{ab}(x,y))\, dxdy - \mathcal{E}(\boldsymbol{\mu}) \right), \tag{15}$$

where $\mathcal{E}(\boldsymbol{\mu})$ is the entropy of the pseudo-matching $\boldsymbol{\mu}$:

$$\mathcal{E}(\boldsymbol{\mu}) = \sum_{a,b=1}^{A} \int \int \mu_{ab}(x,y) \log \frac{\mu_{ab}(x,y)}{N_a f(x|a)} dxdy$$

and the maximization runs over the feasible pseudo-matchings.

Applying the envelope theorem to this program shows that the derivative in $\boldsymbol{\lambda}$ is simply the vector $\sum_{a,b=1}^{A} \mu_{ab}^{\lambda}(x,y)\boldsymbol{\phi}_{ab}(x,y)/2$, where $\boldsymbol{\mu}^{\lambda}$ is the equilibrium pseudo-matching that solves equation (9) when $\boldsymbol{\Phi} = \boldsymbol{\Phi}^{\lambda}$. Returning to (14) and canceling out the 1/2 factors, this gives

$$E_{\boldsymbol{\mu}^{\lambda}} \; \boldsymbol{\phi}_{ab}(x,y) = E_{\hat{\boldsymbol{\mu}}} \; \boldsymbol{\phi}_{ab}(x,y).$$

This shows that the moment matching estimator of $\boldsymbol{\lambda}$ fits the first moments of the basis functions. Now, by equation (15), the function $\boldsymbol{\lambda} \to W(\boldsymbol{U}^{\lambda})$ is convex, as the maximum of a set of affine functions. It follows that the objective function in expression (13) is globally convex and easy to minimize.

At the minimum of (13), the term

$$\frac{\hat{\boldsymbol{\lambda}}'}{2} \sum_{a,b=1}^{A} \int \int \left(\mu_{ab}^{\hat{\boldsymbol{\lambda}}}(x,y) - \hat{\mu}_{ab}(x,y) \right) \boldsymbol{\phi}_{ab}(x,y) dxdy$$

equals zero, so that the value of the objective function in (13) is simply the value of the entropy at $\boldsymbol{\mu}^{\hat{\boldsymbol{\lambda}}}$. This property, along with the close relationship between the entropy and the Kullback-Leibler divergence, makes it a natural measure of fit[16].

The observed pseudo-matching $\hat{\boldsymbol{\mu}}$ is, of course, not directly observed; we only observe $\hat{N}$ mergers with characteristics (a_d, x_d, b_d, y_d). Since x and y are both bi-dimensional, estimating $\hat{\mu}_{ab}(x,y)$ would be very challenging. Fortunately, $\hat{\boldsymbol{\mu}}$ only enters the moment matching objective function via

[16] See the Online Appendix E of Galichon and Salanié (2022) for more details.

the quadruple integral

$$\int\int\sum_{a,b}\phi_{ab}(x,y)\hat{\mu}_{ab}(x,y)dxdy,$$

which we simply estimate with

$$\sum_{d=1}^{\hat{N}}\phi_{a_d,b_d}(x_d,y_d).$$

3.5 Using EU mergers data

So far we have proceeded as if we observed all firms and all mergers in the universe of potential partners for firms in our six European countries (EU6). Our data, however, do not include deals that involve firms outside these six countries, and we only have data on the firms that do merge. Fortunately, the multinomial logit structure of our specification permits surplus estimation from the data we have. As is well-known, this specification incorporates a form of Independence of Irrelevant Alternatives; this is very useful for our purposes, as we can rely on the relative probabilities of the matches we do observe. In fact, we show in Appendix A.2 that we only need to rewrite the equilibrium conditions (9) as

$$\overline{N}_a\overline{f}(x|a)\frac{\exp(U_{ab}(x,y))}{\sum_{d=1}^{A}\int\exp(U_{ad}(x,t))dt}=\overline{N}_b\overline{f}(y|b)\frac{\exp(\Phi_{ab}(x,y)-U_{ab}(x,y))}{\sum_{c=1}^{A}\int\exp(\Phi_{cb}(z,y)-U_{cb}(z,y))dz},\tag{16}$$

where $\overline{N}_a$ and $\overline{N}_b$ refer to our sample of merging EU6 firms, and and $\overline{f}(x|a)$ and $\overline{f}(y|b)$ are the distributions of continuous characteristics within discrete group for our merging firms. After this redefinition, all of Section 3.4 remains valid, with straightforward changes in notation. We need to find functions $e_a(x)$ that solve the system

$$\begin{aligned}\mu_{ab}(x,y)&=\exp(\Phi_{ab}(x,y)/2)\ e_a(x)e_b(y)\ \text{ for all }\ (a,b,x,y)\\ \sum_{b=1}^{A}\int\mu_{ab}(x,y)dy&=\bar{N}_a\bar{f}(x|a)\ \text{ for all }\ a,x.\end{aligned}\tag{17}$$

Using only mergers data has consequences for the identification of $\boldsymbol{\lambda}$, however. Suppose that we add any term $h_b(y)$ to $\Phi_{ab}(x,y)$ in equation (16). Such a term will factor out of both the denominator and the numerator, as explained in more detail in Appendix A.2. Thus, and by symmetry, we can only identify $\Phi_{ab}(x,y)$ up to a sum of arbitrary terms $h_a(x)+h_b(y)$. To put it differently, we identify what Chiappori, Salanié, and Weiss (2017) call the *supermodular core*: all cross-differences

$$\Phi_{ab}(x,y) + \Phi_{a'b'}(x',y') - \Phi_{ab'}(x,y') - \Phi_{a'b}(x',y),$$

expressed in a utility scale that relies on our normalization of the standard error of the $\varepsilon_i(a)$ terms.

These cross-differences are the interactions between the observable characteristics of the potential partners that reveal the complementarities in merger surplus.

Returning to our linear expansion on basis functions, we can only identify the weights λ_k of the basis functions that contain interactions between the observable characteristics of both partners. In our application, we only use this subset of basis functions.

4 Results

In this section, we estimate versions of equation (12) with the data on observed mergers described in Section 2. The matching model allows us to convert the observed frequency of matches between firms with given characteristics into the surplus function that best rationalizes it. We first present the results and then provide an interpretation.

4.1 Baseline Results

As set out in Section 3, the surplus of the merger between two firms with observable characteristics $\{a, x\}$ and $\{b, y\}$ is $\Phi_{ab}(x, y)$, where a, b refer to the partners' discrete characteristics (country and industry) and x, y are the corresponding two-dimensional vectors of continuous variables, each of which has two elements: revenue productivity z and input scale Ω. Since we want the estimated coefficients to be comparable, we take quantile transforms of the z and Ω defined in Section 2.1, conditional on country and industry. As explained in Appendix A.3, we use the "vector ranks" procedure defined in Chernozhukov, Galichon, Hallin, and Henry (2017) in order to preserve the codependence between z and Ω. To minimize notation, we continue to use z and Ω for these vector ranks, which now take values in $[0, 1]$.

We specify the surplus as the following quadratic function of these attributes:

$$\Phi_{ab}(x,y) = \lambda^0_{ab} + \lambda^1_{ab} z_x z_y + \lambda^2_{ab} \frac{\Omega_x z_y + \Omega_y z_x}{2} + \lambda^3_{ab} \Omega_x \Omega_y \tag{18}$$

where the estimated coefficients λ_{ab}^k are allowed to vary depending on whether the two merging firms share discrete characteristics[17].

Table 4 shows the results of a basic specification where only λ_{ab}^0 is allowed to vary with the type of merger (by country and industry), but where the complementarities between the continuous characteristics are common across types. Irrespective of their productivity or scale, mergers in our dataset are more likely to occur between firms operating in the same industry ($I_a = I_b$) and/or in the same country ($C_a = C_b$); this shows up in the positive estimated constants in the surplus function.

Our main interest is in the complementarities between continuous characteristics. We begin by estimating the complementarities between revenue productivity (TFPr) and input scale in columns 1 (coefficient) and 2 (standard error). But, given that revenue productivity, derived from balance sheet data, combines actual productivity with the firm-specific price level, the synergies between revenue productivity could be driven by complementarities between partners' physical productivity or between markups. We further investigate this possibility using the measures of markup and actual quantity productivity derived in subsection 2.1. As explained by De Ridder et al. (2021), these measures do not provide a full decomposition of the revenue productivity. For our estimation procedure, it suffices that they preserve the ranking across firms.[18] The results in Table 4 re-estimate equation (18), replacing revenue productivity by quantity productivity (TFPq, in columns 3 and 4) or markup (columns 5 and 6).[19]

Table 4 shows that more productive firms tend to merge with more productive counterparts, which translates into a strong positive (revenue) productivity complementarity (on $z_x{*}z_y$). Columns 3 and 5 indicate that this reflects a stronger complementarity in efficiency than in markups, although both are large. In contrast, among potential matching firms of equally high productivity, the surplus decreases with the input scale of the merging partner ($z_x * Scale_y$). The negative coefficient is similar for revenue and efficiency productivity. That is, we find negative assortative matching

[17]Recall that since we only observe realized mergers, we can only identify the coefficients that involve interactions between merging firms' characteristics.

[18]See Demirer and Karaduman (2024) (for mergers between power plants) and Braguinsky et al. (2015) (for mergers in the cotton-spinning industry), who analyze effect of mergers using unusual detailed data on physical outcomes, inputs, and productivity.

[19]We replace TFPr with either TFPq or markups because estimating a 3×3 complementarity matrix is computationally burdensome in our data.

Table 4: Quadratic Model of Complementarities

	Revenue Prod		Quantity Prod		Markup	
	coeff	std	coeff	std	coeff	std
	(1)	(2)	(3)	(4)	(5)	(6)
Constant ($C_a = C_b, I_a \neq I_b$)	6.60	0.93	6.60	0.93	6.59	0.88
Constant ($C_a \neq C_b, I_a = I_b$)	3.62	0.85	3.62	0.93	3.62	0.88
Constant ($C_a = C_b, I_a = I_b$)	11.28	0.94	11.28	0.92	11.28	0.87
$z_x \times z_y$	8.50	1.54	10.87	2.30	8.34	2.29
$z_x \times Scale_y$	-9.54	1.27	-9.11	2.23	-1.65	2.99
$Scale_x \times Scale_y$	0.89	1.73	1.05	1.37	2.00	1.64

Note: Estimates of equation (18). In columns 1 and 2, z refers to revenue-productivity (TFPr); in 3 and 4, z refers to quantity-productivity (TFPq); in 5 and 6, z refers to the markup. Scale is input scale in all specifications. The constant terms are expressed relative to the unreported group ($C_a \neq C_b$, $I_a \neq I_b$).

between quantity productivity and input scale of the merging firms. The negative coefficient is much smaller and not significant in column 5, suggesting that the surplus when firms with equally high markups merge is not reduced by merger partner scale. Finally, synergies between the input scales of merging firms are much smaller and not significantly different from zero ($Scale_x * Scale_y$)[20].

4.2 Results by merger type: within- and across- countries and industries

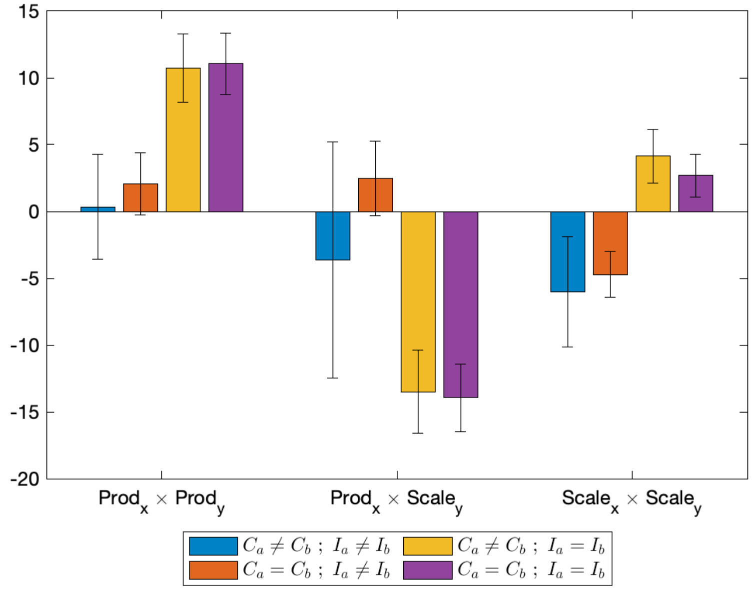
Merger motives, such as potential efficiency gains or increase in market power, are potentially different for within- versus cross-industry mergers, as well as for domestic (within-country) versus cross-border mergers. We now allow the complementarities to vary across those groups. We re-estimate all the coefficients for each of the four types of mergers. The results are shown in Figure 4 for revenue productivity and input scale, and in Figure 5, replacing revenue productivity with quantity productivity and markup.

The first noteworthy result is that within-industry mergers display very similar sources of complementarity regardless of whether they are domestic or cross-border, indicating there is no distinct motive for cross-border M&As in our data. However, implied complementarities are radically different in within- versus cross-industry mergers.

Within industry (yellow and purple bars), the strongest contributors to merger surplus are a

[20] This is in contrast with Akkus, Cookson, and Hortacsu (2016), who analyze bank mergers using two different dimensions of bank size: assets and number of branches. They find strong within-characteristic complementarity for both measures.

Figure 4: Revenue Productivity and Input Size



Note: Estimates (mean and 95% intervals) of complementarities between continuous characteristics in equation 18 by type of deals: cross-border/domestic x same/different industry. The specification also includes constants (λ^0), not shown here.

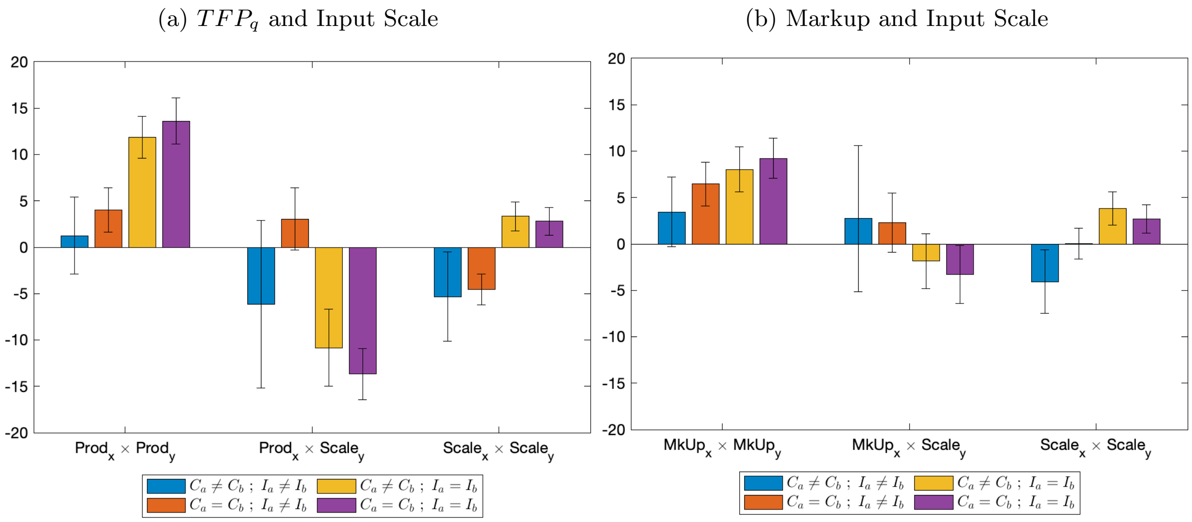
positive complementarity in (revenue) productivity and a negative one between input scale and productivity. Both are driven by synergies in efficiency: the surplus is higher when productive firms merge with other productive, but smaller firms. The complementarity in markups is also positive but much smaller (and very small in the cross characteristics results). Finally, we also observe positive input scale complementarity but of much smaller magnitude.

Firms that merge across industries (blue and orange bars) instead display negative complementarity in input scales. This is the only source of significant complementarity we estimate in unrelated mergers in Figure 4. When we explicitly allow for the role of markups, we find positive significant complementarity also for cross-border deals, especially within countries, but it is smaller than the within-industry markup complementarity highlighted earlier.

Finally, we see that there are no clear motives that are specific to cross-border deals versus within country deals: the dominant complementarities are similar.

We compute how the explanatory power of the model varies across the different characterizations of the firms and deals. Figure 6 shows, as a measure of fit, the value of the entropy in equation (15) for the different estimation models. The *homogeneous-coefficient* model, in blue, corresponds to

Figure 5: Quantity Productivity, Markup and Input Scale



Note: Estimates (mean and 95% intervals) of complementarities between continuous characteristics in equation (18) by type of deals: cross-border and different industry, domestic and same industry. Estimation also includes constants (λ^0) not shown.

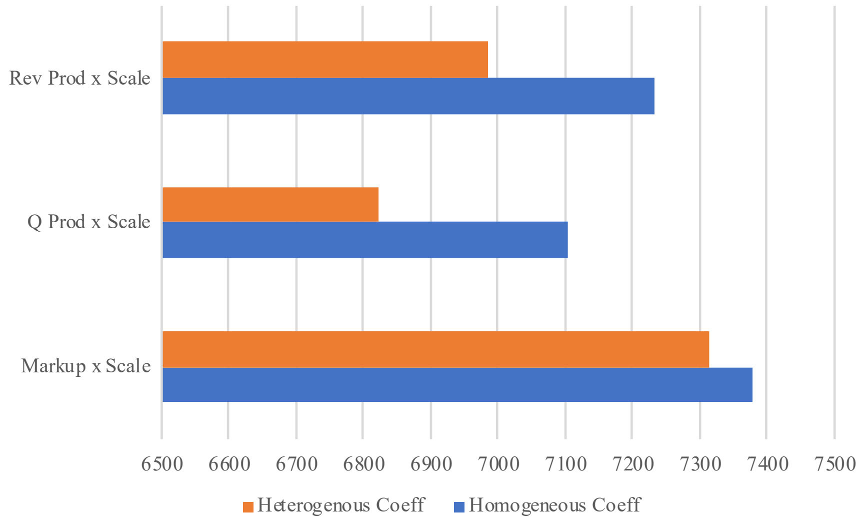
the estimations in Table 4. The *heterogeneous-coefficient* model, in orange, allows synergies to vary by type of merger as reported in Figures 4 and 5. We note that the model performs substantially better when firms are characterized by their technological productivity and the coefficients are allowed to accommodate the different patterns of within- and across-industry deals.

4.3 Economic Interpretation of the Matching Results

The complementarities we estimate rationalize observed mergers given the expected source of merger surplus as a function of the initial characteristics of the merging firms. Theoretical work across fields (industrial organization, international trade, finance, international business) has hypothesized (and tested) the sources and mechanisms driving mergers. Empirically, there is no reason to think that different motives are mutually exclusive; firms plausibly trade them off against each other. One can interpret our results as reflecting the dominant motives for synergies in our sample, given the scarcity of available merger partners in the market for corporate control.

Our finding of productivity complementarity is at odds with the common assumption that the technology of the target is replaced by the superior one from the acquirer after merging. This assumption is often explicitly or implicitly made in the international trade literature when

Figure 6: Goodness of Fit



Note: Entropy in (15) (measured using Kullback-Leibler divergence) based on model in equation (18). The *Homogeneous-coefficient* model imposes $\lambda^k_{ab} = \lambda^k$, for $k = 1, 2, 3$ for all ab. The *Heterogeneous-Coefficient* model allows λ^k_{ab} to vary for $a \neq b$ and $a = b$ (a, b being industry and/or country).

analyzing the welfare implications of cross-border investment flows;[21] as a result, the implied global productivity gains from multinational activity are very large in this literature. We find instead that, for the most part, productive foreign firms *cherry-pick* productive local firms with whom to merge. This is consistent with Javorcik and Arnold (2009) and Guadalupe, Kuzmina, and Thomas (2012): already high-performing local firms become subsidiaries of multinationals. It is also reflected in the findings of Bilir and Morales (2020) where new technologies in multinational headquarters increase the productivity of foreign affiliates if the affiliate is also innovative. This complementarity suggests that pre-merger firm productivity is indicative of technological absorptive capacity, and an important predictor of post-merger productivity growth, as proposed by Cohen and Levinthal (1990).

According to our results, productivity gains in a host country arise from cross-border mergers because adopting new technologies is easier for originally productive firms rather than because formerly unproductive local firms now attain the high productivity of a multinational parent. This suggests that the ability of countries to attract FDI is likely limited by the existing capabilities of

[21] See, for example, Ramondo and Rodriguez-Clare (2013) or Helpman, Melitz, and Yeaple (2004).

domestic firms.

We now turn to the interpretation of the large negative cross-characteristic complementarity for mergers between firms in the same industry (Scale $\times$ TFPq), whereby productive firms generate more value when merging with smaller firms. These results are at odds with the classic *q-theory* of mergers (Jovanovic and Rousseau, 2002), which typically predicts positive assortative matching in cross-characteristics—i.e., more productive firms are likely to merge with larger counterparts. The intuition behind this prediction is that firms can leverage their productivity by combining it with the relatively immobile assets of their merger partner (Nocke and Yeaple, 2007). Instead, our results suggest diseconomies of scale in the transmission of technology.

The smallest source of expected complementarities are, on average, those between the input scales of both firms. For firms in different industries, the scales of production are better described as substitutes; larger firms seek smaller partners, and vice-versa. For within-industry mergers, the effect is positive, which could reflect a market size expansion motive or an empire building motive. The market power motive is better captured by the complementarities in markups, which are the highest between firms operating in the same country and industry.

5 Comparing Predicted and Observed Merger Gains

The results in Section 4 show that the mergers in our data are consistent with bilateral selection of merger partners based on complementarities between each firm's pre-merger characteristics. We now explore whether firm outcomes post merger are related to our estimated expected merger surplus. If the parameters of the matching model are informative about firms' merger motives, and if firms are able to realize the anticipated gains from the merger, then we expect to see a positive correlation.[22] However, no such relationship will be present in the data if either (or both) of these conditions aren't met.

In this investigation, we are interested in the difference between the actual performance of the merged firm m in year t, denoted Π_{mt}, and the hypothetical counterfactual outcome had firms i and j not merged, denoted $\Pi_{m't} = \Pi_{it} + \Pi_{jt}$, where m' is the combination of the merging firms

[22] The management literature details various challenges for successful post-merger integration that could limit firms' ability to realize merger gains. Birkinshaw, Bresman, and Håkanson (2000) argue that human resource management and task integration are critical, and Reus, Lamont, and Ellis (2016) discuss how attempts to transfer knowledge within firm boundaries can destabilize power structures and hamper performance.

in the no-merger counterfactual. This difference is the empirical equivalent of the merger gains in equation (4), repeated here:

$$\tilde{\Phi}_{ab}(i,j) = \frac{\Pi_{ab}(x,y,\tilde{\varepsilon}_i,\tilde{\varepsilon}_j) - \Pi_a(x,\tilde{\varepsilon}_i) - \Pi_b(y,\tilde{\varepsilon}_j)}{\Pi_a(x,\tilde{\varepsilon}_i) + \Pi_b(y,\tilde{\varepsilon}_j)}.$$

The first term in the numerator is the performance of the merged firm m, for which we observe some empirical proxies, and the other terms are counterfactuals for each standalone firm, which we estimate from available data.

Following the production function estimations in subsection 2.1, we rationalize the merged firm and its non-merged counterfactual by a Cobb-Douglas production function, using the relevant industry's labor and capital shares. This gives firm profits as

$$\Psi_{\iota,t} = \frac{\mu_{\iota t} - 1}{\mu_{\iota t}}\ R_{\iota t}$$

where $R_{\iota t}$ is firm revenues and $\mu_{\iota t}$ is the firm's markup defined as in equation (2), as price over marginal cost.

We assume a correspondence between the value of the firm and its profits, of the form $\Pi_{\iota t} = \eta_t \Psi_{\iota,t}$. We can then approximate the merger gains $\tilde{\Phi}_{ab}(i,j)$ by the difference in logs between the actual and counterfactual profits

$$\begin{aligned} \tilde{\Phi}_{ab}(i,j) &\approx (\mu^{-1}_{m't} - \mu^{-1}_{mt}) + (r_{mt} - r_{m't}) \\ &\approx (\mu^{-1}_{m't} - \mu^{-1}_{mt}) + (z_{mt} - z_{m't}) + (\Omega_{mt} - \Omega_{m't}), \end{aligned} \tag{19}$$

where r_{mt} is the log of revenues of the merged firm in year t and $r_{m't}$ is the log of the sum of the counterfactual revenues of the merged firms in year t had they remained standalone.[23] We will further decompose the log of revenues in year t into input scale, $\Omega_{\iota t} = \alpha^I_L \log L_{\iota t} + \alpha^I_k \log K_{\iota t} + (1 - \alpha^I_L - \alpha^I_K)\log M_{\iota t}$, and revenue productivity ($z_{\iota t} = r_{\iota t} - \Omega_{\iota t}$), both of which are already defined in logs.

[23]This expression assumes that the production function coefficients α_L and α_K, and the factor of proportionality between profits and valuation η_t, are invariant to the merger having taken place.

5.1 Post-merger performance data

The firm identifiers in Zephyr allow us to find firm-level performance after the merger for firms that are included in Historical Orbis. The available data on merging firms i and j after $t > 0$ years post-merger reflect any merger gains that are realized.[24] The log of revenues, input scale, and revenue productivity for the actual merged firm can be computed using the data we observe on revenues, employment, and total assets, together with industry-specific production function coefficients.

We also use Historical Orbis data on revenues, employment, total assets, and materials, at the industry level to provide a control group for firms had they not merged, attaching a relevance weight to each control based on the positive selection in taking part in a merger shown in Figure 2. As described in Appendix A.4, the counterfactual growth rates for each of our merging firms are found by assuming that the revenues, employees, total assets, and materials of each firm, had they not merged, would have grown at the weighted growth rate of all firms in the same industry and country between years t and $t-1$.

We have post-merger data for $1,028$ acquirers and 925 targets in the first year after the merger. For 677 of the $2,819$ mergers used to estimate the matching model, we have data on both parties, and can construct the performance of the merged firm. We focus on comparing the outcomes of this subset of mergers to their own counterfactual outcome. The first rows of Table 5 summarize the distribution of the difference in the log of revenues between the merged firm and the counterfactual merged firm in $t = 1, 2, 3, 4$ years after merging. The middle set of rows summarizes the distribution of differences in input scale, and the last set of rows summarizes the distribution of differences in revenue productivity. We limit our attention to mergers where data is available for both firms and also where merging firms are in the same industry. This is because the equivalent of equation (19) for input scale and revenue productivity requires applying industry-specific production function coefficients and it unclear what the relevant coefficients would be for a cross-industry merger.[25]

[24]There is a growing literature investigating merger effects on firm performance that adopts various ways to address the endogeneity of merger decisions that is the main focus of this paper. For example, Blonigen and Pierce (2016) instrument for successful merger completion and find evidence of increases in markups and little productivity gain using differences in differences methods comparing to matched firms that do not merge, to firms in deals that are not completed, and to firms that will be acquired in the future.

[25]For a similar reason, we exclude deals that are within the same industry but where one firm is based in the UK. This is because the production function coefficients estimated for the UK in Section 2 differ from those from other countries.

Table 5: Distributions of post-merger performance relative to counterfactual

Years since merged:		1	2	3	4
Δr_t	Mean	-0.004	-0.040	-0.065	-0.129
	Std dev	0.208	0.309	0.384	0.495
	Min	-1.121	-1.846	-2.010	-2.017
	Max	0.606	0.715	0.972	0.987
	N	677	467	351	271
$\Delta\Omega_t$	Mean	-0.009	-0.036	-0.066	-0.112
	Std dev	0.163	0.248	0.303)	0.388
	Min	-0.586	-0.987	-1.100	-1.590
	Max	0.456	0.648	0.795	0.951
	N	647	436	330	251
Δz_t	Mean	0.008	0.004	0.018	0.010
	Std dev	0.148)	0.198)	0.233)	0.277)
	Min	-0.429	-0.669	-0.659	-0.820
	Max	0.566	0.568	0.786	0.570
	N	647	436	330	251

Notes: The table shows mean difference between the performance of the merged corporation in year t after the merger, for $t = 1, 2, 3, 4$, relative to unmerged counterfactual. The variables are the difference in log revenues, input scale, and revenue productivity. Values are winsorized at 1% and 99%. The sample is all mergers between firms in the same industry group, excluding any cross border mergers that involve a UK firm.

The values are winsorized at 1% in each tail.[26].

Table 5 shows that there is a large, and increasing, variance in post-merger performance across the mergers in the sample. For revenues and input scale, the mean performance relative to the counterfactual was slightly lower in each of four years after the merger, but these differences are very small. For revenue productivity, the mean performance was slightly higher than the counterfactual. However, the standard deviation of each distribution of performance differences is substantial. Figure 7 plots the distributions summarized in Table 5. For each of the three variables, the distribution becomes more dispersed as years since the merger pass[27,28].

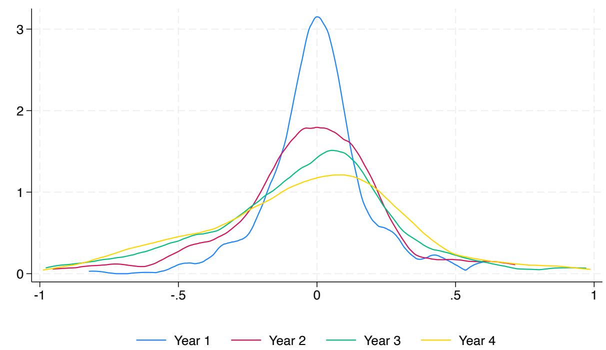
[26] The growth rates in Table 5 do not account for any organizational changes, some of the differences relative to the counterfactuals may reflect changes in the reporting basis for the firm-level accounts.

[27] Geurts and Van Biesebroeck (2019) study the long-term employment effects of mergers between Belgian firms between 2005 and 2012. They find that mergers likely to be motivated by market power show a permanent employment reduction of 14%, whereas those motivated by efficiency gains lead to employment expansions of 10%.

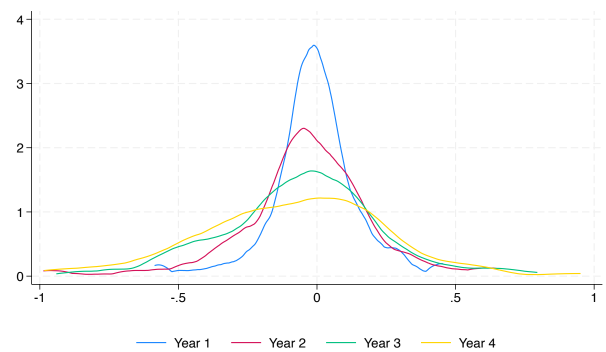
[28] Ashenfelter, Hosken, and Weinberg (2015) and Braguinsky et al. (2015) show industry-specific sources of pro-

Figure 7: Differences in post-merger performance by year

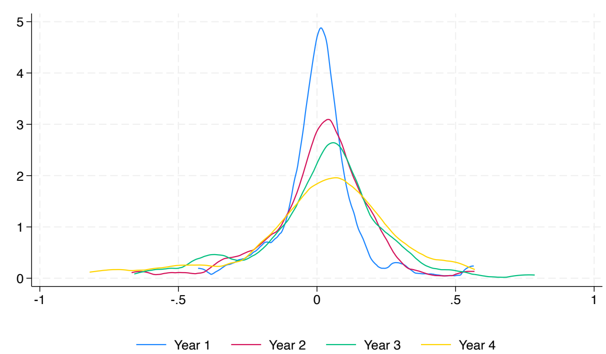
(a) Ln revenues



(b) Input scale



(c) Revenue productivity



Note: These figures present the distributions of differences between actual and counterfactual performance, as summarized in Table 5. The blue distribution is one year after the merger, the red is two years, green in three years, and green is four years.

5.2 Relationship between estimated surplus and post-merger performance

We now turn to the main post-merger analysis. Here, we evaluate whether the variation in post-merger performance, as summarized in Table 5 and Figure 7, is related to our expected merger surplus estimates. If it is, it would mean that the the surplus function and equilibrium model we develop is consistent with the firm's evaluation of their potential merger gains. It therefore also provides a test of the model.

We regress merged firm performance relative to the counterfactual on the predicted merger surplus $\widehat{\Phi}_{ab}(z_x, \Omega_x, z_y, \Omega_y)$ in equation (18), estimated in Section 4. The regression is

$$\Delta r_t = \ln R_{mt} - \ln(E_0[R_{m't}]) = \beta_t \ \hat{\Phi}_m + Y_m + \nu_{mt}, \tag{20}$$

where Y_m are year-of-merger fixed effects. We also estimate the analogous relationships with input scale, $\Delta\Omega_t$, and revenue productivity, Δz_t, replacing revenues on the left hand side.

Table 6 shows the estimated β_t coefficients from equation (20). Each coefficient is multiplied by 100 for ease of presentation. The coefficients in the columns with the heading 'M' (for merged) of Table 6 show that our estimated merger surplus and the post-merger performance of the consolidated entity are positively correlated.

To interpret the magnitudes of the estimated coefficients, we ask how much of the variation in the distribution of post-merger performance relative to counterfactual can be explained by the variation in predicted merger surplus. The coefficient of 0.14 for the difference in log revenues by the end of the first year after the merger (first row, column M) implies that a one standard deviation increase in $\hat{\Phi}_m$ (or 10.6), corresponds to a increase in predicted revenues of 1.4%, which is 7.5% of one standard deviation of the Δr_t variable shown in Table 5. By the end of year four after the merger, the coefficient of 0.88 gives predicted log revenues that are 10.2% higher, equivalent to 20.6% of the standard deviation in this variable.

Next, using expressions (2) and (3), we compute markup and physical productivity for the merged firms and their synthetic counterfactual had they not merged. We again investigate the connection between the predicted merger surplus $\widehat{\Phi}$ and the post-merger firm performance in terms of these two variables. The estimated coefficients, multiplied by 100, are given in Table 7. Here,

ductivity gains after merging relating to plant-level economies of scope and increased leverage of managerial talent, respectively.

Table 6: Correlations between post-merger performance and merger surplus

	Revenue (Δr_t)			Input Scale ($\Delta\Omega_t$)			Rev Prod (Δz_t)		
	M	A	T	M	A	T	M	A	T
Year 1	0.14*	0.22**	0.09	0.08	0.06	-0.15	0.09*	0.17**	0.19**
	(0.09)	(0.09)	(0.12)	(0.07)	(0.05)	(0.14)	(0.06)	(0.06)	(0.07)
N	677	1028	925	647	1008	897	647	1008	897
Year 2	0.49**	0.47**	0.37**	0.19	0.19	0.05	0.29**	0.23*	0.24
	(0.17)	(0.15)	(0.13)	(0.13)	(0.14)	(0.16)	(0.10)	(0.10)	(0.14)
N	467	732	651	436	710	624	436	710	624
Year 3	0.76**	0.79**	0.36	0.34	0.42*	0.10	0.46**	0.38*	0.15
	(0.25)	(0.27)	(0.20)	(0.20)	(0.22)	(0.25)	(0.16)	(0.18)	(0.12)
N	351	548	492	330	536	467	330	536	467
Year 4	0.88**	1.00**	0.49	0.50*	0.45	0.27	0.43**	0.54**	0.33
	(0.30)	(0.34)	(0.38)	(0.23)	(0.25)	(0.41)	(0.14)	(0.18)	(0.23)
N	271	423	383	251	410	362	251	410	362

Notes: The table shows regression coefficient β_t in equation (20): log of revenues, input scale, or TFPr between year 0 and year $t = 1, 2, 3, 4$ relative to counterfactual, regressed on the identified merger surplus terms ($\hat{\Phi}_m$). Each coefficient has been multiplied by 100 for ease of presentation. Columns with the heading M refer to consolidated merged entity and includes the sample of same industry mergers other than any cross-border mergers that include a firm from the UK. The columns labeled A and T refer to the unconsolidated acquiring and target firms, respectively. These observations are not limited to same-industry deals. Standard errors in parentheses are clustered by year. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$..

for the consolidated merged entity in column M, there are no significant correlations between post-merger performance relative to the counterfactual and estimated merger surplus. The coefficients in the M columns for physical productivity and markups are positive but not significant. But as we will see, this hides the different evolution of targets and acquirers.

The estimated surplus function Φ refers to the gains by the consolidated entity relative to the sum of two standalone matched firms, as analyzed in columns with the heading M. However, because our data also include unconsolidated information on the acquirer and target firms, we can also look at the post-merger performance of the firms labeled as the acquirer and the target separately. Recall that in the matching model, we did not segment merging firms into these two groups ex ante (acquirers and targets), and modeled mergers as the matching of roommates. However, after the partners chose each other, the deals are set and the merger is realized, the firm gaining control is labeled as the acquirer in Zephyr. We use this information here. The results are in columns titled A (for acquirer) and T (for target), and are, again, multiplied by 100 for ease of presentation.

Table 7: Post-merger quantity productivity and markups, regressed on merger surplus

	Physical Prod ($\Delta TFPq_t$)			Markup ($\Delta \mu_t$)		
	M	A	T	M	A	T
Year 1	0.06	1.60	6.15***	0.02	0.02	-3.09**
	(0.05)	(1.07)	(1.46)	(0.05)	(0.04)	(1.01)
N	636	1003	866	665	1024	895
Year 2	0.08	1.98*	4.57**	0.17	0.08	-2.31**
	(0.08)	(1.05)	(1.45)	(0.12)	(0.11)	(0.95)
N	425	705	599	456	730	628
Year 3	0.26	4.25**	6.47**	0.18	0.10	-3.04**
	(0.15)	(1.55)	(2.78)	(0.18)	(0.17)	(1.12)
N	324	533	446	346	547	474
Year 4	0.18	3.12*	2.17	0.16	0.24	-2.07
	(0.13)	(1.56)	(0.08)	(0.09)	(0.14)	(1.43)
N	244	404	341	262	419	361

Notes: The table shows regression coefficient β in equation (20): quantity productivity and markups after year $t = 1, 2, 3, 4$ relative to counterfactual, regressed on the identified merger surplus terms ($\hat{\Phi}$). Columns titled M refer to consolidated merged entity and include the sample of same industry mergers other than any cross-border mergers that include a firm from the UK. The columns labeled A and T refer to the unconsolidated acquiring and target firms, respectively. These observations are not limited to same-industry deals. Standard errors in parentheses, clustered by year. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

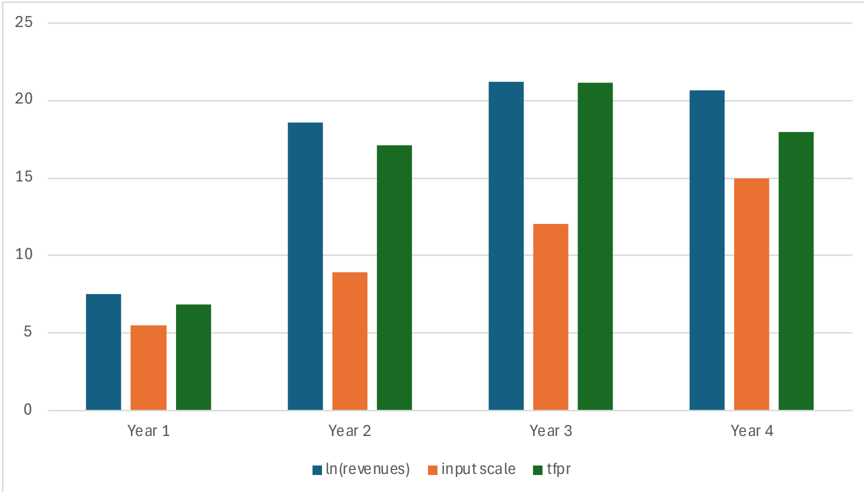
Assuming the acquirers and targets remain in the same industry as they were in before the merger, we are able to compute counterfactual performance measures for firms in all deals where firm-level data is available and, hence, do not limit observations to be from firms in same-industry deals. This gives us many more observations than in the M columns.[29] The correlations in Table 6 between the estimated surplus function and the growth rates of each of the two merged firms are positive. The coefficients for log revenues and revenue productivity are larger for the acquirer, rising over each of the four years after the merger. Input scale is also higher, but not significantly so.[30]

Physical productivity (TFPq in Table 7) is also positively associated with merger surplus for both firms over the four years studied when looking at the acquirer and target separately. We find

[29] The coefficient estimates for acquirer and target firms in Tables 6 and 7 are very similar when restricting the observations to those considered in column M.

[30] As shown in Figure 2, acquirer firms tend to be larger so their performance dominates the consolidated merged firm measures.

Figure 8: Share of performance variance explained by one std dev increase in $\hat{\Phi}$



Note: This figure shows the percentage share of one standard deviation in post-merger performance, relative to counterfactual, explained by a one standard deviation increase in $\hat{\Phi}$. The values correspond to $t = 1, 2, 3, 4$ years after the merger.

no significant correlation between the surplus function and markup in the case of acquirer firms, which, given its weight in the overall merged entity, explains the lack of correlation in column M. In the case of the target firms, the surplus function is correlated with a reduction in markups, relative to the no-merger counterfactual, in every year after the merger.

Note that in most specifications we also find increasing coefficients over time: as more years pass, the estimated predicted merger surplus explains more of the variation in outcomes relative to counterfactual, which suggests that the merger gains take time to materialize. The coefficients are significant for log revenues and revenue productivity, from the first year after the merger; for input scale, while they also grow over time, and are significant only in the fourth year. Figure 8 plots these variance share figures for each of the estimated coefficients in the M columns of Table 6.

Overall the results in this sections are quite striking. Much can happen in an industry that affects revenues, employment, capital growth and productivity. Furthermore firms are likely to innovate, expand their markets, and be subject to numerous firm dynamics. Yet, we are able to explain up to 20% of the variation in actual revenues and productivity by year four with the estimated merger surplus. This is evidence that the estimates from the matching model capture some part of the firms' merger motives and that merging firms were able to realize some of the anticipated merger gains from expected complementarities.

6 Conclusion

This paper develops a structural model of matching that takes into account key characteristics of merger markets in order to document and better understand firms' motives for merging. In particular, it allows for matching across multiple dimensions, and lets productivity and size have distinct effects. The model produces a rich set of results, both about expected complementarities between partners initial characteristics and also about the dynamics of performance post-merger given the predicted surplus.

While we develop the methodology to provide a rich characterization of expected complementarities, we had to make a number of choices that potentially limit the external validity of the empirical application. All our analysis refers to mergers in large European countries. While the estimation approach allows us to restrict the analysis to these mergers without inducing biases, it is possible that the results would be different in other environments. Future work could estimate the model in a wider range of economic settings. In addition, our definition of within-industry mergers is quite broad, and it would be interesting to provide a more detailed assessment of related mergers. Furthermore, since we only observe realized mergers, that by definition were allowed by antitrust authorities, our sample is selected from the set of all mergers that firms sought to achieve. An obvious question is whether disallowed mergers would have had partners matching on different characteristics.

Future work could also try to incorporate the dynamics of the merger market. Mergers are not independent events; they tend to occur in waves, as firms may react to a merger by seeking partners of their own. Similarly, defensive mergers may be triggered by the fear that competing firms may tie the knot. Some firms are involved in several mergers. These considerations suggest developing a richer model of equilibrium mergers that incorporates dynamics.

Finally, our results display complementarities that go against the testable predictions of several main theories of mergers: firms that consider mergers within their broad industry group prefer to match with partners of similar productivity levels; the input scale of the partner matters less, except in that more productive firms seem to prefer smaller partners. We do not know of existing theory that can fully rationalize our set of results. Further work is needed to uncover the microfoundations of the (signed) complementarities that we estimate, as well as on the dynamics of merger gains we document post-merger.

References

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

Akkus, Oktay, Anthony J. Cookson, and Ali Hortacsu. 2016. "The Determinants of Bank Mergers: A Revealed Preference Analysis." Management Science 62 (8):2241–2258.

Ashenfelter, Orley C, Daniel S Hosken, and Matthew C Weinberg. 2015. "Efficiencies brewed: pricing and consolidation in the US beer industry." The RAND Journal of Economics 46 (2):328–361.

Becker, Gary S. 1973. "A theory of marriage: Part I." Journal of Political Economy 81 (4):813–846.

Bilir, L. Kamran and Eduardo Morales. 2020. "Innovation in the Global Firm." Journal of Political Economy 128 (4):1566–1625.

Birkinshaw, Julian, Henrik Bresman, and Lars Håkanson. 2000. "Managing the post-acquisition integration process: How the human iintegration and task integration processes interact to foster value creation." Journal of management studies 37 (3):395–425.

Blonigen, Bruce A and Justin R Pierce. 2016. "Evidence for the effects of mergers on market power and efficiency." Tech. rep., National Bureau of Economic Research.

Braguinsky, Serguey, Atsushi Ohyama, Tetsuji Okazaki, and Chad Syverson. 2015. "Acquisitions, productivity, and profitability: evidence from the Japanese cotton spinning industry." American Economic Review 105 (7):2086–2119.

Chernozhukov, V., A. Galichon, M. Hallin, and M. Henry. 2017. "Monge-Kantorovitch Depths, Quantiles, Ranks and Signs." Annals of Statistics 45:223–56.

Chiappori, P.-A., B. Salanié, and Y. Weiss. 2017. "Partner Choice, Investment in Children, and the Marital College Premium." American Economic Review 107:2109–67.

Chiappori, Pierre-Andre, Alfred Galichon, and Bernard Salanié. 2019. "On Human Capital and Team Stability." Journal of Human Capital 13:236–259.

Choo, Eugene and Aloysius Siow. 2006. "Who marries whom and why." Journal of political Economy 114 (1):175–201.

Ciscato, Edoardo, Alfred Galichon, and Marion Goussé. 2020. "Like attract like? A structural comparison of homogamy across same-sex and different-sex households." Journal of Political Economy 128 (2):740–781.

Cohen, Wesley M and Daniel A Levinthal. 1990. "Absorptive capacity: A new perspective on learning and innovation." Administrative science quarterly :128–152.

Dagsvik, J.K. 1994. "Discrete and continuous choice, max-stable processes, and independence from irrelevant attributes." Econometrica 62 (5):1179–1205.

David, Joel. 2021. "The Aggregate Implications of Mergers and Acquisitions." Review of Economic Studies 88:1796–1830.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The rise of market power and the macroeconomic implications." The Quarterly Journal of Economics 135 (2):561–644.

De Ridder, Maarten, Basile Grassi, Giovanni Morzenti et al. 2021. "The Hitchhikers Guide to Markup Estimation." Tech. rep., LSE.

Demirer, Mert and Omer Karaduman. 2024. "Do Mergers and Acquisitions Improve Efficiency: Evidence from Power Plants." Tech. rep., Stanford GSB mimeo.

Dupuy, Arnaud and Alfred Galichon. 2014. "Personality traits and the marriage market." Journal of Political Economy 122 (6):1271–1319.

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

Fox, Jeremy T. 2010. "Identification in matching games." Quantitative Economics 1 (2):203–254.

———. 2018. "Estimating matching games with transfers." Quantitative Economics 9 (1):1–38.

Gale, David and Lloyd S Shapley. 1962. "College admissions and the stability of marriage." The American Mathematical Monthly 69 (1):9–15.

Galichon, A. 2016. Optimal Transport Methods in Economics. Princeton University Press.

Galichon, Alfred and Bernard Salanié. 2019. "Labeling Dependence in Separable Matching Markets." Tech. rep., mimeo, Columbia.

———. 2022. "Cupid's invisible hand: Social surplus and identification in matching models." Review of Economic Studies 89:2600–29.

Geurts, Karen and Johannes Van Biesebroeck. 2019. "Employment growth following takeovers." The RAND Journal of Economics 50 (4):916–950.

Guadalupe, Maria, Olga Kuzmina, and Catherine Thomas. 2012. "Innovation and foreign ownership." American Economic Review 102 (7):3594–3627.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2004. "Export Versus FDI with Heterogeneous Firms." American Economic Review 94 (1):300–316.

Hopenhayn, Hugo A. 1992. "Entry, exit, and firm dynamics in long run equilibrium." Econometrica :1127–1150.

Javorcik, Beata and Jens Arnold. 2009. "Gifted Kids Or Pushy Parents? Foreign Acquisitions And Plant Performance In Indonesia." Journal of International Economics 79 (1):42–53.

Jensen, Michael. 1986. "The Agency Cost of Free Cash Flow: Corporate Finance and Takeovers." American Economic Review 76:323–329.

Jovanovic, Boyan and Peter L Rousseau. 2002. "The Q-theory of mergers." American Economic Review 92 (2):198–204.

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

Mindruta, Denisa, Mahka Moeen, and Rajshree Agarwal. 2016. "A two-sided matching approach for partner selection and assessing complementarities in partners' attributes in inter-firm alliances." Strategic Management Journal 37 (1):206–231.

Nocke, Volker and Michael D. Whinston. 2022. "Concentration Thresholds for Horizontal Mergers." American Economic Review 112 (6):1915–48.

Nocke, Volker and Stephen Yeaple. 2007. "Cross-border mergers and acquisitions vs. greenfield foreign direct investment: The role of firm heterogeneity." Journal of International Economics 72 (2):336–365.

Ozcan, Yasin. 2015. "Innovation and acquisition: two-sided matching in M&A markets." Northwestern University .

Perry, Martin K. and Robert H. Porter. 1985. "Oligopoly and the Incentive for Horizontal Merger." The American Economic Review 75 (1):219–227.

Ramondo, Natalia and Andres Rodriguez-Clare. 2013. "Trade, Multinational Production, and the Gains from Openness." Journal of Political Economy 121 (2):273–322.

Reus, Taco H, Bruce T Lamont, and Kimberly M Ellis. 2016. "A darker side of knowledge transfer following international acquisitions." Strategic Management Journal 37 (5):932–944.

Rhodes-Kropf, Matthew and David T Robinson. 2008. "The market for mergers and the boundaries of the firm." The Journal of Finance 63 (3):1169–1211.

Roll, Richard. 1986. "The Hubris Hypothesis of Corporate Takeovers." Journal of Business 59:197–216.

Shapley, Lloyd S and Martin Shubik. 1971. "The assignment game I: The core." International Journal of game theory 1 (1):111–130.

Shleifer, Andrei and Robert W. Vishny. 1989. "Managerial Entrenchment: The Case of Manager-Specific Investments." Journal of Financial Economics 25 (1):123–139.

Trefethen, N. 2013. Approximation Theory and Approximation Practice. SIAM.

Williamson, Oliver E. 1968. "Economies as an Antitrust Defense: The Welfare Tradeoffs." The American Economic Review 58 (1):18–36.

Appendix

A.1 Bias when estimating a roommate model as bipartite, and vice versa

A.1.1 Bias when estimating a roommate model as bipartite

Imagine discrete characteristics play no role in the model. Suppose that the true DGP is a Choo and Siow roommate model with

$$\tilde{\Phi}(i,j) = x_i' B x_j + \varepsilon_{x_i,j} + \varepsilon_{x_j,i}.$$

By definition, B is symmetric; in fact, if $\mu(x,y)$ is the joint density of matches where one firm is x and the other is y, then we have

$$B = 2\frac{\partial^2 \log \mu}{\partial x \partial y}(x,y).$$

If we were to label some firms as acquirers and some as targets, and say each match must contain one of each, by drawing from each observed match (x,y):

- with probability $P_A(x,y)$, the acquirer will be x
- with probability $1 - P_A(x,y)$, the acquirer will be y.

This breaks the symmetry if $P_A(x,y) \neq P_A(y,x)$. For example, we could have (with x_1 and y_1 one-dimensional, representing say firm scale)

$$P_A(x,y) = \frac{1}{2} + g(x_1 - y_1)$$

with g odd, increasing, and $g(0) = 0$: the larger firm is more likely to be labeled as the acquirer.

Let $\mu_b(x_A, x_T)$ be the joint density of matches where the acquirer has characteristics x_A and the target has characteristics x_T. It is easy to see that

$$\mu_b(x_A, x_T) = \frac{\mu(x_A, x_T)P_A(x_A, x_T) + \mu(x_T, x_A)(1 - P_A(x_T, x_A)}{D}$$

where D makes μ_b a density. Since $\mu(y,x) = \mu(x,y)$, this gives

$$\begin{aligned}\frac{\partial^2 \log \mu_b}{\partial x_A \partial x_T}(x_A, x_T) &= \frac{\partial^2 \log \mu}{\partial x_A \partial x_T}(x_A, x_T) + \frac{\partial^2 (P_A(x_A, x_T) + 1 - P_A(x_T, x_A))}{\partial x_A \partial x_T} \\ &= \frac{B}{2} + \frac{\partial^2 (P_A(x_A, x_T) + 1 - P_A(x_T, x_A))}{\partial x_A \partial x_T}\end{aligned}$$

If we estimate a bipartite Choo and Siow model $x'_A C x_T$ on the (x_A, x_T) data where firms are labeled acquirers and targets, then we will get

$$\hat{C} = 2\frac{\partial^2 \log \hat{\mu}_b}{\partial x_A \partial x_T}(x_A, x_T) \simeq B + 2\frac{\partial^2 (P_A(x_A, x_T) + 1 - P_A(x_T, x_A))}{\partial x_A \partial x_T}$$

The second term is a bias. Take the $1/2 + g$ example; and specialize it further to $g(t) = -1/2 + L(t)$ with $L(t) = 1/(1 + \exp(-t/h))$ for some positive h. Note that g stays between $-1/2$ and $1/2$, it is 0 when $t = 0$, and it is odd, since $L(-t) = 1 - L(t)$. We need to compute

$$\begin{aligned} 2\frac{\partial^2 \log(-1/2 + L(x_A - x_T) + 1 + 1/2 - L(x_T - x_A))}{\partial x_A \partial x_T} &= 2\frac{\partial^2 \log L(x_A - x_T)/h}{\partial x_A \partial x_T} \\ &= (2/h)\frac{\partial L(x_T - x_A)}{\partial x_T} \\ &= (2/h^2) L(x_A - x_T) L(x_T - x_A), \end{aligned}$$

using, twice, the fact that $L'(t) = L(t)(1 - L(t))/h$.

At least in this reasonable example (given that the acquirers in our data are larger scale, see Figure 2), the bias is unambiguously positive: taking a bipartite model to roommate data when acquirers tend to be larger exaggerates the complementarity of scales. Depending on the extent of difference in scale in the data (a large difference would give us a small value of h in the L function in this example), the bias can be very small or very large.

A.1.2 Bias when estimating a bipartite model as roommate

Now suppose that the true DGP is Choo and Siow bipartite model with $x'_A C x_T$ (and C need not be symmetric). We forget the A, T labels and we estimate a roommate model (allowing any pair of firms to match in a merger). The density of characteristics in this population is just

$$f(x) = (f_A(x) + f_T(x))/2,$$

and the surplus in a pair (x, y), without the labels, should be taken to be $(x'Cy + y'Cx)/2$.

The joint density $\mu(x, y)$ must be symmetric and solve (along with an unknown function e)

$$2 \log \mu(x, y) = \frac{x'Cy + y'Cx}{2} + \log e(x) + \log e(y)$$

and $\int \mu(x, y) dy = f(x)$.

This gives the simple result

$$\hat{B} = 2\frac{\partial^2 \log \hat{\mu}(x, y)}{\partial x \partial y} \simeq \frac{C + C'}{2}$$

This tells us that the bias would be less severe in this direction: our estimator would be the symmetrized matrix C. In particular, the diagonal elements of B would equal those of C.

A.1.3 Conclusion

There is no systematic reason for the bias in Section A.1.1 to be smaller or larger than the bias in Section A.1.2. If P_A is symmetric (e.g. $h = +\infty$ in the example), there is no bias from the first approach. If C is symmetric, there is no bias from the second.

While in models of (heterosexual) marriage, the bipartite model is a reasonable description of the true DGP, there is no strong reason to believe that merging firms have well-defined roles as acquirer or target ex ante. Moreover, in our data, the distribution of input scale among firms labeled acquirers is shifted to the right of the same distribution of input scale among firms labeled as targets, as shown in Figure 2. We therefore suspect a bipartite structure would have led to biased estimated of the complementarities (at least in input scale), as shown in Section A.1.1, and have opted for a roommate structure in Section 3.

Finally, suppose that the A and T are really payoff-relevant, so that $\Phi(x_A, x_T) \neq \Phi(x_T, x_A)$. With perfectly transferable utility, when two firms decide to form a match then they should decide on the best assignment. That is, they solve

$$\Phi_R(x_A, x_T) = \max(\Phi(x_A, x_T), \Phi(x_T, x_A))$$

and the relevant joint surplus is Φ_R. This surplus is symmetric and takes us back to the roommate problem.

A.2 Equilibrium conditions using only observed matches

Our presentation of the model in Section 3 described a world market for mergers based on a representative sample of firms. Our data, however, only has observed mergers between firms in six EU countries (EU6); it excludes firms from other countries, as well as EU6 firms that remain standalone.

As a consequence, we observe neither N_a nor $f(x|a)$. Instead, we observe $\bar{N}_a$ and $\bar{f}(x|a)$ when $a \in EU6$: the number of firms in within-EU6 mergers and the distribution of their continuous characteristics. Since we observe all mergers between EU6 firms ($a, b \in EU6$), our data does give us $\mu_{ab}(x, y)$, the product of the number of mergers μ_{ab} between firms of discrete characteristics a and b, and the conditional joint pdf of their continuous characteristics $\mu(x, y|a, b)$.

For $a \in EU6$, the relationship between these latent and observed distributions of characteristics is given by subtracting the number of firms that do not participate in mergers and those that merge with a non-EU6 firm:

$$\overline{N}_a\overline{f}(x,a) = N_a f(x|a)(1-\mu(\emptyset|a,x)) - \sum_{b\notin EU6}\int \mu_{ab}(x,y)dy.$$

Now take a and b both in EU6. Denoting $I_{ad}(x) \equiv \int \exp\left(U_{ad}(x,y)\right)dy$, we can write

$$\mu(b,y|a,x) = \frac{\exp\left(U_{ab}(x,y)\right)}{1+\sum_{d\notin EU5} I_{ad}(x) + \sum_{d\in EU5} I_{ad}(x)}.$$

Therefore,

$$\begin{aligned}
\mu_{ab}(x,y) &= N_a f(x|a) \times \mu(b,y|a,x) \\
&= \frac{\exp\left(U_{ab}(x,y)\right)}{\sum_{d\in EU6} I_{ad}(x)} \times N_a f(x|a) \times \frac{\sum_{d\in EU6} I_{ad}(x)}{1+\sum_{d\notin EU6} I_{ad}(x) + \sum_{d\in EU6} I_{ad}(x)} \\
&= \frac{\exp\left(U_{ab}(x,y)\right)}{\sum_{d\in EU6} I_{ad}(x)} \times N_a f(x|a) \times \mu(EU6|a,x) \\
&= \frac{\exp\left(U_{ab}(x,y)\right)}{\sum_{d\in EU6} I_{ad}(x)} \times \bar{N}_a \bar{f}(x|a).
\end{aligned}$$

Taking logarithms gives an expression of the form $\log \mu_{ab}(x,y) = U_{ab}(x,y) + 2e_a(x)$. Since $\mu_{ab}(x,y) = \mu_{ba}(y,x)$ and $U_{ab}(x,y) + U_{ba}(y,x) = \Phi_{ab}(x,y)$, we obtain

$$\log \mu_{ab}(x,y) = \Phi_{ab}(x,y)/2 + e_a(x) + e_b(y). \tag{A.1}$$

While we cannot identify the functions e_a, equation (A.1) tells us that (given enough data) we can identify the functions $U_{ab}(x,y)$ nonparametrically up to additive terms. For instance,

$$\Phi_{ab}(x,y) - \Phi_{ab'}(x,y') - \Phi_{a'b}(x',y) + \Phi_{a'b'}(x',y') = 2\log\frac{\mu_{ab}(x,y)\mu_{a'b'}(x',y')}{\mu_{ab'}(x,y')\mu_{a'b}(x',y)}$$

is a simple function of the data. With our parametric specification $\Phi_{ab}(x,y) = \sum_{k=1}^{K}\lambda_k\phi_{ab}^k(x,y)$, we will only be able to identify a parameter λ_k if the double difference

$$\phi_{ab}^k(x,y) - \phi_{ab'}^k(x,y') - \phi_{a'b}^k(x',y) + \phi_{a'b'}^k(x',y')$$

is not constant. This only rules out basis functions of the form $\phi_a(x)$.

The set of equilibrium conditions given by equation (9) also needs to be modified: it becomes

$$\bar{N}_a\bar{f}(x|a)\frac{\exp(U_{ab}(x,y))}{\sum_{d\in EU6}\int\exp(U_{ad}(x,t))dt} = \bar{N}_b\bar{f}(y|b)\frac{\exp(\Phi_{ab}(x,y)-U_{ab}(x,y))}{\sum_{c\in EU6}\int\exp(\Phi_{cb}(z,y)-U_{cb}(z,y))dz}. \tag{A.2}$$

Finally, the moment matching estimator is still valid, with the only difference that the expected utilities become

$$u_a(x) = \log\left(\sum_{b \in EU6} \int \exp(U_{ab}(x,y))dy\right)$$

as the expected utility of each firm only includes its share of the surplus from mergers with EU6 firms.

A.3 Appendix: Numerical Methods

This appendix gives more detail on our implementation of the methods presented in the main text.

A.3.1 The IPFP Algorithm

To solve the system of equations (10), we apply the Iterative Proportional Fitting Procedure. Denoting $H_{ab}(x,y) = \exp(\Phi_{ab}(x,y)/2)$, we need to find functions $e_a(x)$ such that for all (a,x),

$$\bar{N}_a \bar{f}(x|a) = \sum_{b \in EU6} \int \mu_{ab}(x,y)dy = e_a(x) \sum_{b \in EU6} \int H_{ab}(x,y)e_b(y)dy. \tag{A.3}$$

To do so, we convert the equation into an iterative process. First, we choose an initial set of functions $e_a^{(0)}(x)$, which implies a total number of firms

$$\sum_{a,b \in EU6} \int \int H_{ab}(x,y)e_a^{(0)}(x)e_b^{(0)}(y)dxdy$$

equal to the observed number $2\hat{N}$. This can be done, for instance, by taking the constant functions

$$e_a^{(0)}(x) = \sqrt{\frac{\hat{N}_f}{\sum_{a,b \in EU6} \int \int H_{ab}(x,y)dxdy}} \quad \text{for all} \quad (a,x).$$

Then at step p, we compute the integral

$$J_a^{(p)}(x) \equiv \frac{\bar{N}_a \bar{f}(x|a)}{\sum_{b \in EU6} \int H_{ab}(x,y)e_b^{(p-1)}(y)dy}$$

and we renormalize it so as to keep the implied total number of firms equal to $\hat{N}_f$:

$$e_a^{(p)}(x) = J_a^{(p)}(x) \times \sqrt{\frac{\hat{N}_f}{\sum_{a,b \in EU6} \int \int H_{ab}(x,y)J_a^{(p)}(x)J_b^{(p)}(y)dxdy}}.$$

The convergence of IPFP in this case is a consequence of more general results in Galichon and

Salanié (2022). After convergence, we use the $e_a(x)$ functions to compute $\mu_{ab}(x, y)$, $u_a(x)$, and $U_{ab}(x, y)$ as explained in Section 3.3.

A.3.2 Vector Ranks

Each firm in a given country-industry cell a has two continuous characteristics: $x = (z_x, \Omega_x)$. In order to be able to compare our results across cells with different discrete characteristics, we transform x into its *vector ranks* $R_a(x)$.

In one dimension, the rank is simply the quantile transform of the variable. If we take productivity z_x by itself, its rank within a discrete cell a is simply $F_{z_x|z}(z_x|a)$. We could proceed in the same way with scale Ω_x and define the vector rank of x within a as the vector

$$\left(F_{z_x|z}(z_x|a), F_{\Omega_x|z}(\Omega_x|a)\right).$$

Proceeding dimension by dimension destroys the information about the correlation of scale and productivity, however. To preserve it, we use the definition of vector ranks proposed by Chernozhukov et al. (2017). Consider the following optimal transportation problem: we seek to determine the joint distribution G over $(u, x) \in [0, 1]^2 \times \mathbb{R}_+^2$ that minimizes the expected square distance

$$E_G\|x - u\|^2 = \int \|x - u\|^2 dG(u, x)$$

under the constraints that the marginal distributions

$$\int_x dG(u, x) \quad \text{and} \quad \int_u dG(u, x)$$

are respectively $U[0, 1]^2$ (the uniform distribution over the unit square) and $F_{x|a}$ (the distribution of the continuous characteristics in the discrete cell a).

This problem has a unique solution[31]; it is a distribution G whose support is a function $u = R_a(x)$ from $\mathbb{R}_+^2$ to $[0, 1]^2$. The function R_a maps the distribution of the bidimensional variable x into a uniform distribution on the unit square at the smallest quadratic cost. It is, in fact, the only function that both (i) maps the distribution $F_{x|a}$ into $U[0, 1]^2$ and (ii) is the gradient of a convex function. Its inverse is the vector quantile function $x = q_a(u)$.

Since gradients of convex functions are the natural multidimensional generalization of increasing functions, the function R_a has much to commend it as a definition of vector ranks[32].

In practice, we do not know the distribution $F_{x|a}$; we only have a sample $(x_i)_{i=1}^{N_a}$. We solve the optimal transport problem above using the empirical cdf. If we had only one dimension, the rank of

[31] See e.g. Galichon (2016, ch. 6).

[32] It is easy to see that in one dimension, it reduces to the standard quantile transform.

an observation x_i that is the k-th smallest in a could be defines as any number between $(k-1)/N_a$ and k/N_a. In two dimensions, it could be any pair of numbers in a certain convex polytope in $[0,1]^2$. We chose to take the barycenter of the polytope as our definition of the vector ranks.

Our paper uses vector ranks throughout: instead of using (a, x), we use the pair $(a, R_a(x))$. In the rest of this appendix, we write $R_a(x) = X = (X_z, X_\Omega)$.

A.3.3 Numerical Integration

A nice by-product of our use of vector ranks is that it greatly simplifies the computation of the integrals that appear in our formulæ. First consider the integrals that appear in our description of the IPFP algorithm in Appendix A.3.1. At each step, we need to approximate bidimensional integrals of the form

$$\int g(a, x)dx$$

and four-dimensional integrals

$$\int\int h(a, b, x, y)dxdy.$$

Since we use vector ranks X rather than x, the first integral becomes

$$\int_0^1 \int_0^1 g(a, X)dX.$$

To evaluate it, we use Chebyshev approximation and integration. Consider a function $h : [0,1]^2 \to \mathbf{R}$. First we define $\bar{h}(t) = h(2t-1)$, a function with arguments in $[-1,1]^2$. Then we apply Chebyshev approximation to get

$$\bar{h}(t_1, t_2) \simeq 4 \sum_{k,l=0}^{M} c_{kl} T_k(t_1) T_l(t_2)$$

where the T_k are the Chebyshev polynomials[33].

Using standard formulæ[34], we approximate

$$\int_0^1 \int_0^1 h(u_1, u_2)du_1 du_2 = \frac{1}{4} \int_{-1}^1 \int_{-1}^1 \bar{h}(t_1, t_2)dt_1 dt_2 \simeq \sum_{k=0, k \text{ even}}^{M} \sum_{l=0, l \text{ even}}^{M} \frac{c_{kl}}{(1-k^2)(1-l^2)}.$$

We extend this in the obvious way to approximate the four-dimensional integrals.

[33] We took $M = 8$.

[34] See e.g. Trefethen (2013).

A.3.4 The Moment Matching Estimator

To estimate the parameter vector λ, we need to minimize the function

$$\mathcal{M}(\boldsymbol{\lambda}) = \mathcal{W}(\boldsymbol{\lambda}) - \frac{\boldsymbol{\lambda}}{2} \cdot \sum_{a,b=1}^{A} \int\int \boldsymbol{\phi}_{ab}(X,Y)\hat{\mu}_{ab}(X,Y)dXdY$$

as explained in Section 3.4—using now the vector ranks. We start by estimating the moments by

$$\sum_{a,b=1}^{A} \int\int \boldsymbol{\phi}_{ab}(X,Y)\hat{\mu}_{ab}(X,Y)dXdY \simeq \sum_{d=1}^{\hat{N}} \boldsymbol{\phi}_{a_d,b_d}(x_d,y_d) \equiv \hat{\boldsymbol{C}}.$$

The value of $\mathcal{W}(\boldsymbol{\lambda})$ is

$$\frac{\lambda}{2}\int_{[0,1]^4} \sum_{a,b} \mu^{\lambda}_{ab}(X,Y)\phi_{ab}(X,Y)dXdY - \mathcal{E}(\boldsymbol{\mu}^{\lambda})$$

where

$$\mathcal{E}(\boldsymbol{\mu}^{\lambda}) = \int_{[0,1]^4} \sum_{a,b} \mu^{\lambda}_{ab}(X,Y) \log \frac{\mu^{\lambda}_{ab}(X,Y)}{\bar{N}_a} dXdY.$$

(Since the vector ranks X have a uniform distribution on the unit square, $\bar{f}(X|a)$ equals one everywhere.)

Putting everything together, we only need to choose $\boldsymbol{\lambda}$ to minimize

$$\mathcal{G}(\boldsymbol{\lambda}) \equiv \sum_{a,b\in EU6} \int\int \mu^{\lambda}_{ab}(X,Y)\left(\frac{\boldsymbol{\phi}_{ab}(x,y)\cdot\boldsymbol{\lambda}}{2} - \log\frac{\mu^{\lambda}_{ab}(X,Y)}{\hat{N}_a}\right)dXdY - \frac{\boldsymbol{\lambda}\cdot\hat{\boldsymbol{C}}}{2}$$

where $\hat{N}_a$ is the observed number of firms in cell a and

$$\mu^{\lambda}_{ab}(X,Y) = \exp(\boldsymbol{\phi}_{ab}(X,Y)\cdot\boldsymbol{\lambda}/2)e^{\lambda}_a(X)e^{\lambda}_b(Y)$$

is the solution obtained by the IPFP algorithm of Appendix A.3.1 for the parameter values $\boldsymbol{\lambda}$.

In practice, we apply the IPFP algorithm at the two-dimensional Chebyshev nodes $(n_k,n_l)_{k,l=1}^{M}$ to obtain the values of the e^{λ}_a at these nodes; then we approximate the four-dimensional integral using Chebyshev integration as explained in Appendix A.3.3.

We obtain the standard errors by the usual device: we write the first-order conditions

$$\begin{aligned} 0 &= \frac{1}{\sqrt{\hat{N}}} \hat{\mathcal{G}}'(\hat{\boldsymbol{\lambda}}) \\ &\simeq \frac{1}{\sqrt{\hat{N}}} \hat{\mathcal{G}}'(\boldsymbol{\lambda}_0) \\ &+ \frac{1}{\hat{N}} \hat{\mathcal{G}}''(\boldsymbol{\lambda}_0) \times \sqrt{\hat{N}}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0) \end{aligned}$$

which gives an asymptotic distribution

$$\sqrt{\hat{N}}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0) \simeq N\left(0, \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right)$$

with

$$\begin{aligned} \boldsymbol{I} &= \lim_{\hat{N}\to\infty} \frac{V\hat{\mathcal{G}}'(\boldsymbol{\lambda}_0)}{\hat{N}} \\ \boldsymbol{J} &= \text{plim}_{\hat{N}\to\infty} \frac{\hat{\mathcal{G}}''(\boldsymbol{\lambda}_0)}{\hat{N}}. \end{aligned}$$

The matrix $\hat{\mathcal{G}}'(\boldsymbol{\lambda}_0)/\hat{N}$ consists of two terms. Its variance comes from the second term,

$$\hat{E}\boldsymbol{\phi} = \frac{1}{\hat{N}} \sum_{d=1}^{\hat{N}} \boldsymbol{\phi}_{a_d,b_d}(x_d, y_d)$$

whose variance can be estimated as

$$\frac{1}{\hat{N}} \sum_{d=1}^{\hat{N}} (\boldsymbol{\phi}_{a_d,b_d}(x_d, y_d) - \hat{E}\boldsymbol{\phi})(\boldsymbol{\phi}_{a_d b_d}(x_d, y_d) - \hat{E}\boldsymbol{\phi})'.$$

The matrix $\boldsymbol{J}$ can simply be estimated by evaluating the Hessian of $\hat{\mathcal{G}}$ at $\hat{\boldsymbol{\lambda}}$.

A.4 Post-Merger Performance Analysis

A.4.1 Data about post-merger outcomes

This appendix details the data selection process for the analysis of post-merger performance in Section 5. We have 5,638 firm-level observations for the 2,819 bilateral mergers studied in this paper. The firm identifiers in Zephyr allow us to find firm-level performance after the merger for firms that are included in Orbis.[35] Historical Orbis (HO) is country specific and contains information about the firms that continued to file separate accounts post merger. HO contains the data on firm revenues, employment, total assets, and some measures of materials costs for $3,997$ of the $5,638$ firm listings.[36] To measure the performance of the merged firms in the years after the merger, we need values for revenues, employees, assets, and materials costs for both firms in the pair. We have revenues data for 677 deals, and data for all the variables for 647 deals, one year after the merger, and the number of deals covered falls in each subsequent year as one or both firms leave the data.[37]

Consider two firms i and j that merge. All available data on these firms after $t > 0$ years post-merger reflect any merger gains that are realized. The data from year t allow us to find the joint revenues of the actual merged firm, m, in year t, $R_{mt} = R_{it} + R_{jt}$. The actual revenue productivity and input scale of the merged firm m in years $t > 0$ can be computed by combining data on observed revenues, employment, total assets, and materials costs of firms i and j in year t. Defining $L_{mt} = L_{it} + L_{jt}$, $K_{mt} = K_{it} + K_{jt}$, and $M_{mt} = M_{it} + M_{jt}$, we estimate the revenue productivity for the merged firm in year $t \geq 0$ as:

$$\begin{aligned} z_{mt} &= \ln(R_{mt}) - \hat{\alpha}_l \ln(L_{mt}) - \hat{\alpha}_k \ln(K_{mt}) - (1 - \hat{\alpha}_l - \hat{\alpha}_k) \ln(M_{mt}) \\ &= r_{mt} - \hat{\alpha}_l l_{mt} - \hat{\alpha}_k k_{mt} - (1 - \hat{\alpha}_l - \hat{\alpha}_k) m_{mt}, \end{aligned} \tag{A.4}$$

and the input scale of the actual merged firm in year $t \geq 0$ as:

$$\Omega_{mt} = \hat{\alpha}_l l_{mt} + \hat{\alpha}_k k_{mt} + (1 - \hat{\alpha}_l - \hat{\alpha}_k) m_{mt}. \tag{A.5}$$

The $\hat{\alpha}$ estimates are computed at the two-digit NACE Rev. 2 industry level, following the procedure in Section 2.[38]

To relate the variation in the gains to merger surplus, we need to compare the observed gains to an estimate of the equivalent growth rates had the firms not merged. To do this, we return to HO to construct this counterfactual. We use data on firms' revenue, employment, total assets, and

[35] There are $5,301$ unique firms in the data. 258 firm identifiers are repeated because they take part in multiple deals in the sample.

[36] As in Section 4, for firms in UK, the materials measure is the cost of goods sold. For firms in all other countries, we use the materials cost HO variable.

[37] The data analyzed in Section 4 are from the year before the merger. Attrition from post-merger data does not necessarily imply firm death, but could reflect relabeling of the firm identifier after the merger, a selection bias whose sign is not obvious.

[38] In Section 2.1, we use the estimates produced using the Ackerberg, Caves, and Frazer (2015) approach at the two-digit industry level, using the acquirer firm's industry code, without imposing constant returns to scale. For the combined firms in the analysis, the mean estimated sum of the production function coefficients is 0.989, the median is 0.977, and the standard deviation is 0.097, suggesting that the constant returns assumption is reasonable.

materials, after the years of the mergers in our data to provide a control group for firms had they not merged. For each firm in HO, we calculate the percentage growth in these variables. Next, because the firms that provide good comparison firms for our merger sample are a non-representative subset of all firms in any country-industry, we attach a relevance weight to each control firm. To do this, and motivated by the positive selection on revenues into taking part in a merger shown in Figure 2, we perform a logistic regression of the indicator of being part of merger on firm revenues in 2008, which is the first year of HO data we collect. The estimated coefficient from this regression is used to predict the probability that each firm is part of a merger. Then, for each country-industry we sum up growth in revenues, input scale, and revenue productivity for all firms weighted by the probability they are part of a merger, excluding our sample firms that were actually part of a merger.

The counterfactual growth rates for each of our merging firms are found by assuming that the revenues, employees, total assets, and materials of each firm, had they not merged, would have grown at the weighted growth rate of all firms in the same industry and country between years t and $t-1$. For example, for revenues, let i' indicate firm i under the counterfactual that it did not merge with firm j, then the counterfactual revenues in each year $t > 0$ after the merger are:

$$R_{i't} = R_{i',t-1}(1 + g_{rt})$$

where g_{rt} is the weighted country-industry growth rate, and $R_{i'0} = R_{i0}$. Counterfactual employment, total assets, and materials costs are estimated in a similar way.

We construct a synthetic firm m' that consists of the two counterfactual standalone units i' and j'. The expected revenues for the synthetic firm m' in the years after the merger are assumed to simply be the sum of the counterfactual estimated revenues of firms i' and j':

$$E_0[R_{m't}] \quad = \quad R_{i't} + R_{j't}, \tag{A.6}$$

and the counterfactual expected number of employees, total assets, and materials costs are found analogously, $E_0[L_{m't}] = L_{i't} + L_{j't}$, $E_0[K_{m't}] = K_{i't} + K_{j't}$, and $E_0[M_{m't}] = M_{i't} + M_{j't}$.

Under the constant returns to scale assumption, the expected input scale of the synthetic counterfactual merged firm m' in years $t > 0$ is computed by combining data on estimated counterfactual employment, total assets, and materials of firms i' and j' in year t:

$$E_0[\Omega_{m't}] = \hat{\alpha}_l \ln(E_0[L_{m't}]) + \hat{\alpha}_k \ln(E_0[K_{m't}]) + (1 - \hat{\alpha}_l - \hat{\alpha}_k) \ln(E_0[K_{m't}]), \tag{A.7}$$

and the estimate of the expected revenue productivity of the synthetic counterfactual merged firm m' in years $t > 0$ is:

$$E_0[z_{m't}]) = \ln(E_0[R_{m't}]) - E_0[\Omega_{m't}], \tag{A.8}$$

using the same $\hat{\alpha}$ estimates as in equations (A.4) and (A.5).