HOW WIDE IS THE FIRM BORDER?*

Enghin Atalay
Ali Hortaçsu
Mary Jialin Li
Chad Syverson

We examine the within- and across-firm shipment decisions of tens of thousands of goods-producing and goods-distributing establishments. This allows us to quantify the normally unobservable forces that determine firm boundaries, that is, which transactions are mediated by ownership control, as opposed to contracts or markets. We find firm boundaries to be an economically significant barrier to trade: having an additional vertically integrated establishment in a given destination ZIP code has the same effect on shipment volumes as a 40% reduction in distance. These effects are larger for high value-to-weight products, faraway destinations, differentiated products, and IT-intensive industries. *JEL* Codes: F12, G34, L22, M11.

I. Introduction

A vast literature, beginning with Coase (1937), has sought to build an economic theory of the firm. A central question addressed in this literature is what forces determine which transactions occur within firm boundaries as opposed to across them. The literature has put forward many possible explanations for why some transactions are better moderated by the firm. The more prominent classes of explanations include the transaction costs theories first developed by Williamson (1971, 1973, 1979) and Klein, Crawford, and Alchian (1978); the property rights theory in Grossman and Hart (1986) and Hart and

*We thank Matt Backus, Francine Lafontaine, Nicola Pavanini, Sebastian Sotelo, Catherine Thomas, and Mike Whinston for their helpful and constructive comments. We also thank Erin Robertson for her excellent editorial assistance, and Frank Limehouse for his help at the Chicago Research Data Center. The research in this article was conducted while the authors were Special Sworn Status researchers of the U.S. Census Bureau at the Chicago Census Research Data Center and the UW-Madison branch of the Minnesota Census Research Data Center. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia, the Federal Reserve System, the Federal Reserve Board of Governors, or the U.S. Census Bureau. This article has been screened to ensure that no confidential data are revealed.

The Quarterly Journal of Economics (2020), 1845–1882. doi:10.1093/qje/qjz026.
Advance Access publication on August 19, 2019.

Moore (1990); the ownership-as-incentive instrument structure of Holmstrom and Milgrom (1991) and Holmstrom and Tirole (1991); the resource-based view of Wernerfelt (1984); the routines-based theory of Nelson and Winter (1982); and the knowledge-based explanation of Kogut and Zander (1992).[1]

The considerable empirical literature spurred by these theories has studied how such factors influence firm formation, size, and scope. The modal analysis in this literature identifies a likely (and hopefully exogenous) source of variation in the net gains of keeping a transaction inside the firm (e.g., greater R&D intensity) and then relates this variation to observed outcomes in firm structure. The estimated object of interest is the sign of the comparative static (e.g., whether increases in R&D intensity increase the extent of vertical integration, a question addressed by Acemoglu et al. 2010) and occasionally the magnitude of the relationship between the explanatory variable and firm structure outcomes.

What has not been attempted, however, is an estimate of actual magnitudes of the net benefits of internal transactions—the actual size of avoided transaction costs, or the benefit of retaining residual rights of control through ownership, or the advantage of internal incentives, and so on. This strikes us as an important missing piece. These benefits, after all, are the core empirical object in theories of the firm. Yet we do not know how big they actually are, or how they vary in magnitude across market environments. There are several reasons for this dearth of estimates of the magnitudes of "what makes a firm a firm." First, by their nature, the factors proposed by the theoretical literature tend to be shadow values. They are explicitly about nonmarket transactions and often about costs that are not paid, so they are inherently difficult to measure. More practically, even if one could imagine constructing a reasonable measure of these shadow values (using the payroll of a company's procurement department as a measure of transaction costs, for example), this would require highly detailed data. Furthermore, if such data exist, it would only be for specific firms in specific markets and perhaps only for specific transactions.[2] It would be difficult to extend any such measures

1. Gibbons (2005) discusses these various theories and distills the transaction cost, property rights, and incentive explanations into four formal theoretical structures.

2. We are aware of one case study, that of a naval shipbuilder, for which such detailed data exist (Masten, Meehen, and Snyder 1991). There, the authors estimate that the shipbuilder's costs would nearly double, relative to its observed cost-minimizing procurement choices, if all of its inputs were sourced externally.

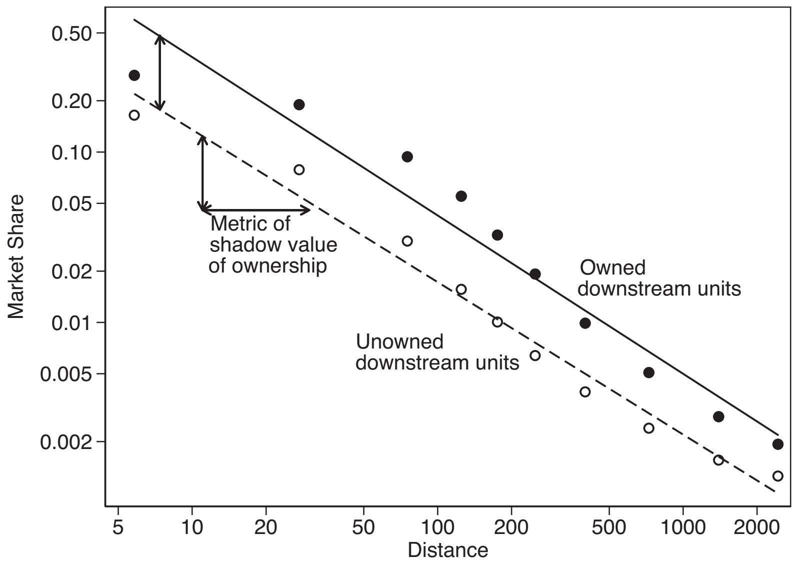

FIGURE I

Illustration of Our Approach

This figure portrays the relationship between trade flows and distance for transactions that take place across firm boundaries (dashed line) and within firm boundaries (solid line). The two vertical lines are of equal length. Thus, the horizontal line gives the distance-related reduction in trade flows equivalent to the reduction in trade flows associated with crossing firm boundaries.

to more general settings, at least without some model that empirically relates a transaction's observable characteristics to the net benefit of keeping that transaction within the firm.

This article proposes a method to measure the magnitude of the forces that shape firm boundaries. Our approach uses a firm-side analogue to the consumer concept of revealed preference to measure the shadow values of keeping transactions inside a firm. Specifically, we use firms' revealed choices about what, where, and to whom to ship to measure the implied shadow values of in-house transactions.

We detail our approach below, but the basic logic can be portrayed in a simple figure. Applying our data set of establishments' shipments, ownership, and location (which we describe in Section III), Figure I presents the relationship between transaction volumes and distance for two types of transactions: transactions internal to the firm (solid circles, with a solid fitted line) and transactions across firm boundaries (hollow circles, with a dashed fitted line). An extensive empirical literature has established that transaction volumes decline in distance because of various costs, ranging from physical transport costs, to monitoring, to coordination, and beyond. If we observe, all else equal, that firms systematically have a greater volume of internal than external transactions at any given distance (something that can be expressed in Figure I as the vertical distance between the two lines), it is because they perceive internal shipments as being less costly. Because we observe the overall relationship between shipment volumes and distance, which lets us characterize the magnitude of distance-based costs, we can obtain a cardinal measure of the "distance premium" of internal shipments—the perceived cost savings of keeping transactions within the firm. In other words, differences in the patterns of firms' within- and across-firm shipments reveal the hurdle they perceive for transacting outside their borders. We do not need to see these costs directly in the data. Firm behavior and the volume-distance relationship reveal to us what they are.

Aside from allowing us to measure what to this point has not been quantified, our approach has other advantages. For one, the literature has focused on comparing different governance structures based on how they mediate transactions. This is a comparison that our data on within- and across-firm shipments uniquely permit. In addition, we can apply our method to a wide swath of transactions, firms, and markets. We analyze millions of shipments from tens of thousands of establishments in the goods-producing and goods-distributing sectors in the United States. This allows us to characterize how our estimated shadow values vary with observable variables about the product being transacted, the production function of the firm, and even the attributes of specific transactions.[3]

3. It is important to note that our "revealed preference" approach allows us to remain agnostic about the specific source(s) of the shadow benefits of keeping transactions in-house, be they transaction cost savings, residual rights of control, advantages of incentive structures, elimination of the double marginalization problem, some other factor, or any combination thereof. A firm's decisions tell us how large it perceives these benefits to be, not the specific mechanism(s) through which they arise. This cost does come with a benefit, though; we do not need to rely on untestable assumptions about the source for measurement. Following a substantially different approach, Wallis and North (1986) gauge the aggregate importance of transaction costs by measuring the sizes of industries primarily engaged in conducting and intermediating transactions.

We find that the net benefits of keeping transactions in-house are substantial. They are equivalent in magnitude to the costs associated with decreasing the distance between separately owned counterparties by 40%. Moreover, the organizational and spatial structure of economic activity is significantly shaped by the forces that determine the boundaries of the firm. We characterize systematic patterns in the heterogeneity of firm boundary effects across different settings, finding that the net benefits of within-firm transactions are larger for more distant shipments, for high value-to-weight products, for more differentiated products, in industries that are more IT intensive, and for establishments that produce goods rather than just convey them. We also address the potential bias created by the endogeneity of establishment ownership and location.

In our earlier work (Atalay, Hortaçsu, and Syverson 2014), we documented that internal shipments were rare for many vertically integrated establishments. Using the same main data sources as in the current study, we computed the share of each establishment's shipments that were sent to other plants within the same firm. We found that for the median establishment at the upstream end of a production chain, less than 1% of its shipments are sent internally.[4] We interpreted this empirical finding as signifying that the primary rationale for common ownership for most production chains is to facilitate within-firm flows of intangible, rather than physical, inputs.

However, our earlier work does not imply that common ownership has no effect on firms' physical input-sourcing patterns. Our approach here isolates internal/external shipment differentials, holding all else constant. As such, it provides an estimate of the shadow value of ownership in physical shipments. However, this shadow value is just one of many factors, including the number and location of the same-firm and between-firm counterparties, that influence the prevalence of internal sourcing. It could well be (and our earlier work strongly suggests) that the balance of those factors usually makes external shipments the most profitable choice. That is, on net, those other factors end up outweighing the shadow value that we measure in this article. This can be true even if that shadow value is substantial in size, as we find here.

4. Because internal shipment shares are skewed across establishments, and because larger establishments tend to have larger internal shares, the weighted mean is 16%.

We offer the following analogy from cross-country trade. In the first half of 2019, more than two-thirds of Porsche's automobiles were sold to customers outside of Europe (where all Porsches are assembled).[5] That does not mean there are no foreign trade costs (explicit or implicit) associated with those sales, or that these costs are small in any absolute sense. Rather, other favorable factors make those foreign sales profitable on net despite the fact that Porsche must pay trade costs associated with those sales. Indeed, in this article we have formulated a method, following the voluminous international trade literature on trade costs, to measure a firm's shadow cost of shipping outside its ownership border. (This is the analog to Porsche's trade cost of shipping outside Europe.) This firm-boundary cost could be large, and here we find that is the case. Nevertheless, firms may still make most of their shipments outside their borders (Porsche may ship most of their automobiles outside Europe) if the myriad other influences on the value of a shipment (things that influence Porsche's profitability from a sale) typically outweigh the across-border cost. The fact that firms make most of their shipments to external customers is not contradictory to the costs of crossing firm boundaries being substantial.

Including the work already mentioned, this article relates to three literatures. First, our work contributes to the extensive literature that tries to test and quantify the importance of various theories of the firm. Lafontaine and Slade (2007, 2013) provide an excellent discussion of the empirical literature that investigates moral hazard, transaction cost, and property rights–based models of firm boundaries. Key contributions to this literature include Baker and Hubbard (2003, 2004), who use monitoring technology improvements to assess the role of moral hazard; Monteverde and Teece (1982) and Masten (1984), who, respectively, use differences in inputs' human and physical capital specificity to test transaction cost–based theories of the firm; and Acemoglu et al. (2010), who use supplier and customer R&D intensity to distinguish between transaction cost and property rights theories of the firm. In addition to exploring the determinants of firm boundaries, other work assesses the consequences of vertical integration. For example, Chipty (2001),

5. See https://newsroom.porsche.com/en/2019/company/porsche-deliveries-first-half-2019-18093.html (accessed July 20, 2019). As of July 2019, Porsche has three production facilities, in Bratislava, Leipzig, and Stuttgart.

Hortaçsu and Syverson (2007), and Forbes and Lederman (2010) assess vertical integration's impact on efficiency and competition in the cable TV, ready-mix concrete, and airline markets, respectively.

Second, although this article considers the interaction of ownership and domestic trade flows, it has clear connections to the literature on foreign direct investment and international trade; see Antràs and Yeaple (2014) for a useful review. Beyond considerations of factor abundance and proximity to consumers (Brainard 1997; Markusen and Venables 2000; Helpman, Melitz, and Yeaple 2004), firms' decisions on where to locate and whether to outsource certain inputs to foreign suppliers are shaped by the same "theory of the firm" explanations discussed in the previous paragraph. Related to transaction cost–based explanations, Fally and Hillberry (2018) construct a multi-industry, multicountry trade model of firm location and organization. The main trade-off in their model balances transaction costs against within-firm coordination costs. Tasks are integrated within the firm to save on the costs of transacting with suppliers or customers, but because of increasing marginal costs of coordinating tasks in the firm, not all tasks within a production chain are performed by the same firm. As transaction costs decline, product line fragmentation increases, and activity is spread out over a larger number of countries. Related to the property rights approach, Antràs and Chor (2013) model a multistage production process where the value of the final good is a function of investments made at each stage. Each stage may either be integrated with the final producer or outsourced to a supplier. A key prediction of the model is that integration at later (respectively, earlier) stages of production is more likely when investments along the chain are strategic complements (respectively, strategic substitutes). Antràs and Chor (2013) find empirical support for this prediction using aggregate data from the Census Related Party Database (this result is reaffirmed in firm-level data in Alfaro et al. 2019). In sum, the first two literatures examine how differences in proxies for transaction costs, property rights, and other factors shape firm boundaries domestically and internationally. Our complementary contribution is to measure the actual magnitude of the costs associated with transacting across firm boundaries.

Third, our work also has ties to the vast literature that uses gravity models to infer the costs associated with transacting with faraway counterparties; see Anderson and van Wincoop (2004), Costinot and Rodríguez-Clare (2014), and Head and Mayer (2014) for syntheses of this literature.[6] As emphasized in these literature reviews, the gravity equation of trade—according to which the flows of goods or services across two regions is directly proportional to the size of these regions and inversely proportional to the distance between them—emerges as the prediction of a broad class of trade models. Our contribution in this article is to leverage what is known from the gravity equation literature about distance-based trade impediments to obtain an estimate of the net benefit of internal transactions.[7]

II. The Gravity Equation

The framework we use to predict trade flows from establishments to destination ZIP codes borrows heavily from Eaton, Kortum, and Sotelo (2012). In particular, we adopt the model elements that yield a gravity equation that is both relatively simple to derive and allows for zero trade flows between pairs of regions. This latter element is important, as zero trade flows are common in our data. The model also aggregates up to the ZIP code level nicely. This is very useful, because although our data set is extremely detailed, it does have a limitation in that we observe a shipment's destination ZIP code rather than its recipient establishment within that ZIP code. We can use the model to directly derive an estimating equation that uses this more aggregate destination information. In this section, we sketch out the model

6. McCallum (1995) provides one of the first attempts to infer the "width" of national borders from trade flows. A complementary literature uses deviations from the law of one price as a way to measure the costs of trading across regions. We owe the title of our article to an exemplar of this literature, Engel and Rogers (1996).

7. Close to our work, Boehm (2017) applies a gravity equation–based methodology to recover the costs associated with imperfect contract enforcement. In countries with high legal costs for enforcing market transactions, firms will have a greater frequency of internal shipments and—to the extent that national accounts do not record internal shipments in input-output tables—lower expenditures in national input-output tables. Building on this work, Boehm and Oberfield (2018) document that Indian manufacturers rely more heavily on internal sourcing in states with slow enforcement of contracts. They quantify the aggregate importance of distortions caused by slow contract enforcement. Relative to Boehm (2017) and Boehm and Oberfield (2018), we provide an encompassing estimate of the net costs of transacting across firm boundaries.

assumptions, then jump to the estimating equation. Intermediate steps in our derivation are given in Online Appendices A and B.

We make two minor amendments to the Eaton, Kortum, and Sotelo (2012) model. First, we characterize the expected flows from specific sending establishments to destination regions (ZIP codes in the data, as discussed), as opposed to having both the origin and destination represent regions. Second, critically for our empirical question, we permit trade barriers to be lower when the sending and receiving establishment belong to the same firm.

Establishments operate in $1, \ldots, Z$ ZIP codes, with multiple establishments potentially located in each destination ZIP code z. We use i to refer to source ZIP codes. Establishments ("plants") can both produce/send and use/receive commodities. Each plant produces a single, horizontally differentiated traded commodity.[8] Denote the identity of a potential receiving establishment with its location z^e, and similarly refer to the sending establishment as i^e.[9]

8. In the empirical application in Sections III and IV, we construct market shares separately by commodity. We omit commodity-level superscripts throughout this section for notational simplicity. The analysis in this section can easily be extended to multiple traded commodities with constant expenditure on each commodity. This can be accommodated by a model in which a representative consumer in each ZIP code has Cobb-Douglas preferences over commodities; in Online Appendix E, we discuss a multi-industry model along these lines. In reality, some establishments sell multiple products. In our main sample, described below, 84% of the average establishments' sales come from its single largest commodity code. We abstract from multiproduct considerations and use establishments' industry and commodity interchangeably.

9. We do not attempt to directly model firms' decisions on where to locate their establishments, or which establishments to own, as in Antràs (2005), Keller and Yeaple (2013), or Ramondo and Rodríguez-Clare (2013). In an international setting, the aforementioned trade models emphasize that related-party and arm's-length trade are substitutes. A richer, more complete model would analyze location and ownership choices in combination with establishments' sourcing decisions. Due to the complexity of modeling both sets of choices in our context, in which there are thousands of possible locations, we do not pursue this richer model. We do, however, further discuss the endogeneity of firms' ownership and location decisions in Section IV.C. Also within the literature on foreign direct investment, Baier et al. (2008), Bruno et al. (2017), and Head and Mayer (forthcoming) apply gravity equations to jointly analyze aggregate FDI and international trade flows. Again, given the large number of potential locations in which firms can locate their different establishments and the granularity of our data, it would not be feasible to apply these papers' methods to our research question. Instead, our methodology for accounting for the endogeneity of ownership obviates the estimation of a gravity equation for firm location decisions.

Each sending establishment has access to a (random) number of linear production technologies, each of which allows it to transform l units of labor into $V \cdot l$ units of output. We assume that V is Pareto distributed with shape parameter θ and a lower cutoff $\underline{v}$ that can be set arbitrarily close to 0. We also assume that the (integer) number of establishment i^e's varieties with efficiency $V > v$ (for $v > \underline{v}$) is the realization of a Poisson random variable with mean $T_{i^e} v^{-\theta}$. In this expression, the parameter T_{i^e} reflects the overall productivity of establishment i^e.

Call x_i the cost of a unit of labor inputs for establishments in ZIP code i. There are also iceberg-style transportation costs that vary not only with distance but also based on ownership. Specifically, for establishment i^e to sell one unit of the commodity to plant z^e, it must produce $d_{zi} \geqslant 1$ units of output if plant z^e is owned by a different firm and $d_{zi}\delta_{zi} \geqslant 1$ units of output if the same firm owns it.[10] Furthermore, forming a relationship with establishment z^e requires a fixed number of workers F_{z^e} to be hired in ZIP code z.

So far, our assumptions have been on each supplier's technology and the trade barriers between each supplier and customer. These assumptions yield expressions for the probability that i^e will be among the lowest cost suppliers to z^e. From here, additional assumptions about how suppliers compete with one another are required to generate predictions of expected trade flows among customer-supplier pairs. In Online Appendix A, we delineate these assumptions, aggregate across all of the customers within each destination ZIP code, and impose a set of parametric restrictions between d_{zi}, δ_{zi}, and mileage.

In combination, as we demonstrate in Online Appendix A, our assumptions yield a relatively simple expression for i^e's expected market share as a function of (i) sending-establishment-specific terms, (ii) pair-specific observable variables, and (iii) a summation of destination-specific terms:

$$(1) \quad \mathbb{E}\left[\frac{X_{zi^e}}{X_z}|\Lambda\right] \approx \frac{\exp\left\{\alpha_{i^e} + \alpha_1 \cdot \log \text{mileage}_{z\leftarrow i} + \alpha_2 \cdot s_{zi^e} + \alpha_3 \cdot s_{zi^e} \cdot \log \text{mileage}_{z\leftarrow i}\right\}}{\sum_{i'=1}^{Z}\sum_{i'^e \in i'} \exp\left\{\alpha_{i'^e} + \alpha_1 \cdot \log \text{mileage}_{z\leftarrow i'} + \alpha_2 \cdot s_{zi'^e} + \alpha_3 \cdot s_{zi'^e} \cdot \log \text{mileage}_{z\leftarrow i'}\right\}}.$$

10. The additional costs associated with across-firm transactions, $\frac{1}{\delta_{zi}}$, reflect not only the costs of transacting with an already known business partner but also the costs related to searching for appropriate, trustworthy suppliers or customers. Providing evidence from an experiment in which small and medium-sized Chinese businesses were assembled in business associations, Cai and Szeidl (2018) indicate that the benefits of finding the right counterparties may be substantial.

Here, $\frac{X_{zi^e}}{X_z}$ equals the share of ZIP code z's expenditures sourced from supplier i^e. Conditioning on Λ indicates that there is some random component of trade barriers, namely, that the relationship between d_{zi} and mileage—and alternatively between δ_{zi} and mileage—contains some random component. Furthermore, s_{zi^e} equals the fraction of establishments in the destination ZIP code z that share ownership with the establishment i^e. Finally, $\alpha_{i^e} \equiv \alpha_0 + \log T_{i^e} - \theta \log x_i$ collects all of the relevant sending-establishment-specific unobservable terms.

There are two possible approaches to estimating the parameters involved in the expression for the expected market share. The first, advocated by Anderson and van Wincoop (2003), is to incorporate both destination and sending-establishment fixed effects:

$$
\begin{aligned}
\mathbb{E}\left[\frac{X_{zi^e}}{X_z}|\Lambda\right] &\approx \exp\{\alpha_1 \cdot \log \text{mileage}_{z\leftarrow i} + \alpha_2 \cdot s_{zi^e} \\
&\quad + \alpha_3 \cdot s_{zi^e} \cdot \log \text{mileage}_{z\leftarrow i} + \alpha_{i^e} + \alpha_z\}. \qquad (2)
\end{aligned}
$$

The destination fixed effects in equation (2) capture the terms in the denominator in equation (1). This theoretically motivated specification produces consistent estimates of the same-firm fraction, distance, and interaction terms.

One drawback of this approach is that, with tens of thousands of sending establishments and tens of thousands of destination ZIP codes, it is computationally taxing. As an alternative approach, in most of our specifications we follow the earlier literature on gravity equation estimation and regress $\frac{X_{zi^e}}{X_z}$ against sending-establishment fixed effects, distance terms, and destination-specific multilateral resistance terms (as discussed in Baier and Bergstrand 2009). These multilateral resistance terms involve subtracting off a first-order Taylor approximation of the terms in the denominator of the right-hand side of equation (1). Namely, for each pair-specific explanatory variable, g_{zi^e}, our regressions include $g_{zi^e} - \overline{g}_{z\cdot} - \overline{g}_{\cdot i^e} + \overline{g}$ as the covariate; $\overline{g}_{z\cdot}$, $\overline{g}_{\cdot i^e}$, and $\overline{g}$, respectively, denote the average value of the covariate g_{zi^e} for a given destination ZIP code z, for a given establishment i^e, or across all sending establishment-destination ZIP code pairs. In essence, the multilateral resistance terms apply the mechanics of linear models with two-way fixed effects to the gravity relationship.

An appropriate estimator for either specification is the multinomial pseudo maximum likelihood estimator, which can be implemented via a Poisson regression; see Santos Silva and Tenreyro (2006), Head and Mayer (2014, section 5.2), or Sotelo (2019).

III. Data Sources and Definitions

Our analysis employs two large-scale data sets maintained by the U.S. Census: the Longitudinal Business Database (LBD) and the Commodity Flow Survey (CFS). We supplement these data with two sets of industry-level definitions from past work: our definitions of vertically related industry pairs (from Atalay, Hortaçsu, and Syverson 2014) and Rauch (1999)'s product differentiation classification.

Our benchmark sample is drawn from the establishments surveyed in the 2007 CFS. Like its predecessors, the 2007 CFS contains a sample of establishments operating in the economy's goods-producing and goods-distributing sectors: mining; manufacturing; wholesale; electronic shopping and mail-order houses; and newspaper, book, and music publishers. Once a quarter, each surveyed establishment is asked to report up to 40 randomly selected shipments that it made on a given week in that quarter.[11] Relevant for our purposes, the data include each shipment's origin and destination ZIP code, weight, and dollar value.[12] The sample contains approximately 4.3 million shipments made by roughly

11. For each surveyed establishment, the set of shipments that we observe is only a small fraction of the shipments actually sent. In Online Appendix Table 10, we corroborate that our benchmark results are not sensitive to the sparsity of our shipment data.

12. Transfer pricing—whereby firms shift reported sales from high corporate tax to low-tax jurisdictions—may potentially lead us to mismeasure shipment values for intrafirm shipments. Bernard, Jensen, and Schott (2006) and Davies et al. (2018) document that this behavior is common in cross-border transactions. For two reasons, transfer pricing is likely to play a much smaller role in our data set of domestic shipments. First, while corporate tax rate differences do exist across states, they are small relative to differences that exist across countries. Furthermore, existing multijurisdictional apportionment agreements limit the ability of multiestablishment firms to engage in transfer pricing in their domestic shipments. Second, the CFS responses are kept confidential and by law may not be used for legal proceedings, including those related to taxation. Thus, CFS respondents have no economic incentive to shift revenues across establishments in their survey responses.

58,000 establishments.[13] Because we are interested in characterizing the shipment patterns of establishments that could make same-firm shipments, we only keep establishments from multiunit firms. This reduces the sample size to approximately 35,000 establishments.[14] Our main analysis focuses on data from 2007. In supplemental analyses, we control for past shipping behavior using the 2002 CFS. In these analyses, our sample consists of the 9,000 establishments from multiunit firms that are surveyed in both the 2002 and 2007 versions of the CFS. Throughout the article, we limit our analysis to domestic shipments. Although the CFS includes shipments for export, the data only report the ZIP code of the shipment's port of departure from the United States and its destination country; we do not see the specific destination within the foreign country or anything about ownership of the receiving establishment. Thus, we cannot construct either of the key variables for our analysis for exported shipments.

While the CFS is a shipment-level data set, we sum up across shipments within a surveyed establishment-destination ZIP code pair to obtain each observation in our analysis data set.[15] We create the sample as follows. We first segment the 2007 CFS by the six-digit North American Industry Classification System (NAICS) industry of the shipping plant. For each industry, we collect all destination ZIP codes that receive at least one shipment from any establishment. We then create the Cartesian product of all shipping plants and all destination ZIP codes for that industry. Our sample consists of the aggregation of these Cartesian products across all six-digit industries. Our benchmark sample has 190 million sending establishment–destination ZIP code observations.

13. Census disclosure rules prohibit us from providing exact sample size counts.

14. It would, of course, be feasible to include single-unit firm establishments in our estimation of the relationship between trade flows, common ownership, and distance. Doing so would only increase the precision of our estimate of the effect of distance on trade flows with no impact on our internal-shipment coefficient estimates. Online Appendix Table 9 in Online Appendix C confirms this.

15. Note that the CFS allows us to observe the destination ZIP code of the shipment, not the identity of the particular receiving establishment. As a result, our level of observation is demarcated by a (shipping) establishment on one side but a ZIP code on the other. It means we must infer internal shipments as a function of the prevalence of downstream establishments owned by the shipping establishment's firm (our model helpfully provides the form of this function under its assumptions) rather than being able to observe these internal shipments directly.

The main variables of interest in the next section's empirical specification are the market share and distance measures. The market share for a shipping plant i^e in destination z is the total value of shipments from i^e to z divided by the total shipments sent to z by all plants in i^e's six-digit NAICS industry. Our main analysis relates this market share to measures of the distance, be they literal or figurative, between i^e and the establishments located in ZIP code z. The physical, great-circle distance between two ZIP codes is straightforward to compute using the ZIP codes' longitudes and latitudes. A key figurative distance measure s_{zi^e} is the fraction of downstream establishments in ZIP code z owned by the same firm that owns establishment i^e; below, we call this variable the "same-firm ownership fraction."[16] To compute this fraction, we restrict attention to the establishments in ZIP code z that could conceivably use the product that establishment i^e is shipping. For example, if i^e is a cement manufacturer, we would not want to include dairy producers, auto wholesalers, or gas stations when computing s_{zi^e}. To discern which establishments are downstream of i^e and could conceivably use i^e's output, we apply the algorithm introduced in our earlier work (Atalay, Hortaçsu, and Syverson 2014). Namely, we find industry pairs I, J for which at least 1% of the output of industry I is purchased by establishments in industry J. (In Online Appendix Table 10, we reassess our main empirical findings for other choices of this cutoff.) Then, when computing s_{zi^e} for each establishment $i^e \in I$ we sum only over the plants in ZIP code z that belong to a downstream industry J.

Table I presents summary statistics for our sample of establishment-destination ZIP code pairs. Panel A indicates that the total value shipped (summing across all potential sending establishments i^e) is highly skewed. Although the median six-digit product-destination ZIP code shipment total is around \$1.6 million, the mean is around \$14.5 million. Second, the average market share, $\frac{X_{zi^e}}{X_z}$, equals 0.004. Only 0.7% of sending establishments

16. Throughout the article, we refer to i^e and z^e as commonly owned if the two establishments have the same census firm identifier. We draw on the LBD—a U.S. Census–compiled registry of all establishments with at least one employee—to identify the firm identifiers for each establishment in each ZIP code. The Census Bureau draws on multiple data sources and performs multiple checks to produce census firm identifiers that closely reflect the true ownership patterns that exist across establishments. We outline these data sources and checks in Online Appendix C.1 of Atalay, Hortaçsu, and Syverson (2014).

TABLE I
SUMMARY STATISTICS

	Percentile						
	10	25	50	75	90	Mean	Std. dev.
Panel A: Entire sample							
Total shipment value to z ($ millions)	0.1	0.3	1.6	7.6	27.5	14.5	94.1
Market share, $\frac{X_{zi^e}}{X_z}$	0	0	0	0	0	0.004	0.061
Panel B: If there is a shipment from i^e to z							
Number of total downstream ests. at z	0	2.0	7.5	18.5	42.5	17.26	30.49
Number of same-firm downstream ests. at z	0	0	0	0	0	0.041	0.250
Number of same-firm establishments at z	0	0	0	0	0	0.113	0.622
Same-firm ownership fraction	0	0	0	0	0	0.0051	0.0455
Panel C: If there is no shipment from i^e to z							
Number of total downstream ests. at z	0	1.0	5.0	13.5	31.0	12.90	24.86
Number of same-firm downstream ests. at z	0	0	0	0	0	0.009	0.110
Number of same-firm establishments at z	0	0	0	0	0	0.026	0.240
Same-firm ownership fraction	0	0	0	0	0	0.0009	0.0166
Panel D: Log mileage. . .							
if the same-firm ownership fraction = 0	5.58	6.22	6.77	7.29	7.65	6.66	0.87
if the same-firm ownership fraction > 0	5.22	6.02	6.69	7.26	7.64	6.54	1.04
if there is no shipment from i^e to z	5.60	6.22	6.77	7.29	7.65	6.67	0.85
if there is a shipment from i^e to z	2.78	4.10	5.54	6.53	7.16	5.23	1.65

Notes. The sample consists of pairs of sending establishments and destination ZIP codes, i^e–z, for which at least one shipment by an establishment in the same industry as i^e was sent to ZIP code z. The market share equals the ratio of the shipments sent by i^e to ZIP code z, relative to the total amount sent by all establishments in the same industry as i^e to ZIP code z. The total number of i^e–z pairs in the sample is 189.6 million. Of these, for 1.4 million pairs there is at least one shipment from i^e to z (with 188.2 million pairs with no shipments). Of the 189.6 million i^e–z pairs, the same-firm ownership fraction is greater than 0 for 1.4 million pairs (and equals 0 for the remaining 188.1 million pairs).

have any shipments to z. In short, zero trade flows are exceedingly common in our sample of i^e–z pairs.

Panels B and C split i^e–z pairs by the presence or absence of shipments from i^e to z. The two takeaways from these panels are that (i) establishments tend to ship to ZIP codes that contain some potential counterparties with which they share ownership, but (ii) same-firm shares are still low, even in ZIP codes that receive at least one shipment. For the mean i^e–z pair, 12.9 establishments in z belong to industries downstream of sender i^e. But of these 12.9, only 0.01 establishments, on average, share ownership with the sender. Shipments are more likely to be sent to ZIP codes in which at least one of the potential recipients belongs to the same firm as the sender. For destination ZIP codes that receive at least one shipment from i^e, 0.51% of the potential recipients share ownership with the sender, compared to 0.09% when no shipment is sent.

Panel D offers a summary of ownership and shipment distances. Not surprisingly (and consistent with gravity models of the type we leverage here), shipments become less likely as the distance to a potential recipient increases. The median distance between sending establishments and destination ZIP codes that receive at least one shipment is 254 miles, while it is 870 miles for pairs with no shipments. The relationship between ownership and distance is a priori less clear cut. On the one hand, by choosing to locate establishments far apart, firms can economize on shipping costs to their customers. On the other hand, the costs of managing establishments may be increasing in distance.[17] As it turns out, establishments under common ownership tend to be closer to one another. For i^e–z pairs with a potential recipient in z owned by the firm that also owns i^e, the 10th percentile distance is 184 miles, and the 25th and 50th percentile distances are 411 and 804 miles, respectively. In contrast, for pairs in which no such common ownership link exists, the 10th, 25th, and 50th percentile distances are uniformly larger: 264, 501, and 866 miles, respectively.

To sum up, we can draw the following three conclusions from Table I. First, for any particular destination ZIP code, it is rare for there to be an establishment sharing ownership with the sender.

17. For instance, Giroud (2013) and Kalnins and Lafontaine (2004, 2013) demonstrate that proximity allows a firm's headquarters to monitor and acquire information from the firm's other establishments, thereby increasing those establishments' productivity and, in turn, profitability.

TABLE II
RELATIONSHIP BETWEEN DISTANCE, COMMON OWNERSHIP, AND MARKET SHARES

Dependent variable: $\frac{X_{zi^e}}{X_z}$	(1)	(2)	(3)	(4)	(5)	(6)
Same-firm ownership fraction	2.596	2.828	2.941	2.633	2.854	2.911
	(0.047)	(0.049)	(0.047)	(0.040)	(0.040)	(0.041)
Log mileage	− 0.923	− 0.962	− 0.944			
	(0.004)	(0.003)	(0.005)			
Distance $\leqslant$ 10 miles				4.215	4.355	4.460
				(0.017)	(0.017)	(0.018)
Distance $\in$ (10, 50] miles				3.611	3.777	3.874
				(0.015)	(0.016)	(0.017)
Distance $\in$ (50, 100] miles				2.647	2.817	2.876
				(0.008)	(0.015)	(0.016)
Distance $\in$ (100, 200] miles				1.750	1.897	1.922
				(0.013)	(0.013)	(0.014)
Distance $\in$ (200, 500] miles				0.709	0.802	0.788
				(0.008)	(0.009)	(0.010)
Distance > 1,000 miles				− 0.487	− 0.584	− 0.340
				(0.010)	(0.013)	(0.020)
Multilateral resistance	None	Unweighted	Weighted	None	Unweighted	Weighted

Notes. All regressions include sending-establishment fixed effects. The sample includes 190 million i^e–z pairs drawing on the shipments made by 35,000 establishments. In columns (4)–(6), the omitted distance category contains ZIP code pairs that are between 500 and 1,000 miles apart. Standard errors are clustered at the level of the sending establishment. With the exception of Online Appendix Table 12, we apply this clustering in all subsequent tables.

Second, pairs of establishments that are owned by the same firm and belong to vertically related industries tend to be located closer to one another than the typical upstream-downstream pair. Finally, a potential destination ZIP code that contains an establishment sharing ownership with the sending firm tends to receive more shipments. So our data on domestic shipments indicate that firms choose to locate their establishments close to one another and that distance and common ownership shape shipment frequencies.

IV. RESULTS

IV.A. Benchmark Specification

Table II reports our baseline regression results relating distance and ownership to the share of a ZIP code's purchases of a given product purchased from a sending establishment i^e. Our benchmark specification is given by equation (2), where we first (momentarily) fix α_3—the coefficient on the distance-ownership interaction term—to be equal to 0, and then use the Baier and Bergstrand (2009) multilateral resistance terms to proxy for the destination ZIP code fixed effect. The columns differ according to how we model the relationship between distance and the market share (either logarithmically or, more flexibly, with a sequence of categorical variables) and which multilateral resistance term we include (whether the averages that are being subtracted off of the distance and ownership measures are weighted by the trade flows or are unweighted).[18] Through the trade-offs between distance and ownership, firms reveal in their shipment patterns the costs they perceive in transacting outside their borders. Given that transaction costs generally increase with distance, if establishments are systematically more likely to ship a greater distance to same-firm establishments than other-firm establishments (or, equivalently, ship a greater volume internally than externally at any given distance), this indicates that they see a differential cost in transacting within rather than between firms.

Consistent with a large body of evidence drawing on international trade flows (Disdier and Head 2008), we find that the elasticity of bilateral trade flows with respect to distance is close to 1. Newer to the literature and the focus of our study is the estimate embodied in the same-firm ownership share coefficient. We find values of approximately 2.5 to 3. Interpreting the magnitude of these coefficients requires a short calculation. Our same-firm ownership metric is the fraction of establishments in downstream ZIP code z that are owned by i^e's firm. For the average i^e–z pair, there are 12.9 potential recipients (establishments in industries that are downstream of i^e) in the destination ZIP code. Using $r_{i^e z}$ to refer to the number of potential recipients in ZIP code z, the average (across i^e–z pairs) of $\frac{1}{1+r_{i^e z}}$ equals 0.315. Thus, the addition of a same-firm establishment in the destination ZIP code is associated with the same change in i^e's market share in z as a reduction in the distance from i^e to z by a factor of $\exp\left(\frac{0.315 \cdot 2.828}{-0.962}\right) \approx 0.40$, a 60% reduction. This implied "distance premium" of ownership in-

18. When computing $g_{zi^e} - \overline{g}_{z\cdot} - \overline{g}_{\cdot i^e} + \overline{g}$ in columns (2) and (5) of Table II, $\overline{g}_{z\cdot}$, $\overline{g}_{\cdot i^e}$, and $\overline{g}$ are simple, unweighted averages. In columns (3) and (6), we also compute averages but instead weight observations by the observed flows from the sending establishment multiplied by the observed flows to the destination ZIP code. Throughout this section, we exclude i^e–z pairs for which i^e resides in z, because the log(mileage) variable is undefined for these pairs. The results from our regressions would be unchanged in an alternative specification in which we included these i^e–z pairs in our regression sample while also including, as a covariate, an indicator variable describing whether i^e is located in ZIP code z. See Online Appendix C, Table 14.

TABLE III

RELATIONSHIP BETWEEN DISTANCE, COMMON OWNERSHIP, AND MARKET SHARES: INTERACTIONS AND SENSITIVITY ANALYSIS

Dependent variable:	$\frac{X_{zi^e}}{X_z}$ (1)	$\frac{X_{zi^e}}{X_z}$ (2)	$\frac{X_{zi^e}}{X_z}$ (3)	$\frac{X_{zi^e}}{X_z}$ (4)	$1\ \frac{X_{zi^e}}{X_z} > 0$ (5)
Same-firm ownership fraction	1.605	2.641	3.090	0.000	2.948
	(0.132)	(0.026)	(0.026)	(0.018)	(0.040)
Log mileage	−0.964	−0.961	−0.962	−0.023	−0.964
	(0.003)	(0.001)	(0.001)	(0.007)	(0.003)
Interaction between log mileage	0.279		0.218		
and same-firm ownership fraction	(0.023)		(0.015)		
Sample		Benchmark		$\frac{X_{zi^e}}{X_z} > 0$	Benchmark
Destination ZIP code fixed effects	No	Yes	Yes	No	No
Multilateral resistance	Unweighted	None	None	Unweighted	Unweighted

Notes. All regressions include sending-establishment fixed effects. With the exception of the second column, the sample includes 190 million i^e–z pairs, drawing on the shipments made by 35,000 establishments. In the fourth column, the sample includes the 1.4 million i^e–z pairs with positive trade flows.

creases somewhat as we first include (column (2)) and then use a weighted version of (column (3)) a multilateral resistance control. The final three columns replace log mileage with a flexible set of indicators for various distance categories to capture any nonlinearities in distance effects. The same-firm ownership coefficients change little.

With an additional assumption on θ—which, in our Section II model, parameterizes the heterogeneity of productivity draws—we can express the cost savings of common ownership explicitly and directly, not indirectly as a function of distance. Using α_2 to refer to the coefficient on the same-firm ownership fraction and our maintained parameterization on trade costs, the cost reduction associated with common ownership equals $(\alpha_2 + 1)^{-\frac{1}{\theta}}$; see equation (7) in Online Appendix A. With $\alpha_2 = 2.83$ and two values of θ that span the range adopted by the literature (see section 5.3 of Costinot and Rodríguez-Clare 2014), the costs of trade under common ownership are multiplied by a factor of 0.71 (with $\theta = 4$) or 0.85 (with $\theta = 8$). In the remainder of the section, we apply the "distance premium" as our metric of the benefit of common ownership, because it does not depend on θ. However, with this extra parameter choice, all of our ensuing regression results can be restated as a direct cost reduction.

In Table III, we explore how the relative importance of common ownership varies by distance, the intensive versus extensive margins of trade, and the impact of destination fixed effects on our estimates. The first column includes an interaction of the same-firm ownership fraction with logged distance, allowing the relationship between ownership and the probability of shipping to a location to vary with distance. To help with interpretation, we demean the distance variable when including an interaction term in our specification. The interaction term has a positive coefficient, implying that the link between same-firm presence and the market shares is stronger for more distant destinations. An additional same-firm downstream establishment in the destination (again, equivalent to an increase in the same-firm ownership fraction by 0.315) in destinations at the 10th, 50th, and 90th percentile distances has the same impact on trade flows as a reduction in shipping distance by 57%, 69%, and 80%, respectively. (The main effect of distance is somewhat larger in magnitude in this specification.) In columns (2) and (3), we apply destination ZIP code fixed effects, obviating the use of the multilateral resistance terms used in our specifications above. The coefficient estimates are reassuringly similar to that in the benchmark specification.

Columns (4) and (5) explore the intensive versus extensive margins of trade. In column (4), we restrict our sample to pairs of sending establishments and destination ZIP codes with positive trade. In column (5), we revert to the benchmark sample but modify the dependent variable so that it equals 1 if the sending establishment ships to the destination ZIP code. We find that conditional on positive sales, there is basically no relationship between trade flows, distance, and the same-firm ownership fraction variable. In contrast, the likelihood that an establishment ships to a given destination ZIP code is strongly increasing in our ownership variable and strongly decreasing in distance. These findings follow from the Eaton, Kortum, and Sotelo (2012) framework we apply: the lack of an intensive margin reflects the balance of two opposing forces. First, holding fixed the set of supplying establishments which supply ZIP code z, lower trade barriers imply higher sales (higher $\frac{X_{zie}}{X_z}$). On the other hand, lower trade barriers expand the set of establishments that can profitably enter each destination ZIP code. Because these "marginal suppliers" are relatively low productivity and have relatively low sales, their inclusion into the set of suppliers lowers the average of $\frac{X_{zie}}{X_z}$. In the Eaton, Kortum, and Sotelo (2012) model, these two forces exactly offset. Moreover, our estimates in columns (4) and (5) accord with the empirical findings in Hillberry and Hummels (2008).

TABLE IV
RELATIONSHIP BETWEEN DISTANCE, COMMON OWNERSHIP, AND MARKET SHARES: PANEL REGRESSIONS

Dependent variable:	$\frac{X_{zi^e,2007}}{X_{z,2007}}$ (1)	$\frac{X_{zi^e,2007}}{X_{z,2007}}$ (2)	$\frac{X_{zi^e,2007}}{X_{z,2007}}$ (3)	$\frac{X_{zi^e t}}{X_{zt}}$ (4)
Same-firm ownership fraction from 2007	2.970 (0.088)	2.415 (0.085)	1.779 (0.106)	2.770 (0.078)
Log mileage	−0.911 (0.006)	−0.792 (0.006)	−0.792 (0.006)	
$X_{zi^e} \cdot (X_z)^{-1}$ from 2002		2.153 (0.017)	2.150 (0.017)	
Same-firm ownership fraction from 2002			1.049 (0.123)	
Fixed effects	Sending establishment			Sending estab.× destination ZIP

Notes. The sample includes 43 million i^e–z pairs, drawing on the shipments made by 9,000 establishments included in the 2002 and 2007 versions of the CFS. In all specifications, we calculate the unweighted multilateral resistance terms.

Using the 1997 CFS, they also find that trade flows decrease with distance essentially entirely through the extensive margin.[19]

Up to now, we have excluded past shipment information from our list of explanatory variables. We did so primarily because the set of establishments that are surveyed by the Census changes from one edition to the next, meaning that including past shipment information as an explanatory variable reduces the sample size considerably. But using data from an earlier version of the CFS, we can examine how changes in ownership reshape establishments' shipment patterns, accounting for past shipment decisions. In the first column of Table IV, we replicate our benchmark specification, using as a sample the set of establishments that were surveyed in both the 2002 and 2007 versions of the CFS. The coefficient on common ownership is similar to that in our benchmark sample, while the coefficient on distance is slightly smaller in magnitude. In the second column, we include $\frac{X_{zi^e}}{X_z}$ from

19. Within the empirical international trade literature, the extensive margin plays a primary—though not total—role in shaping trade flows; see Head and Mayer (2014, 186). Beyond the obvious international versus domestic distinction, there are a number of potential explanations for the difference in the estimated role of the extensive margin. In the CFS, each supplier is an individual establishment. In contrast, within the international trade literature, senders comprise multiple establishments, implying a greater range of products and thus a larger scope for the intensive margin to operate.

the 2002 CFS as an additional regressor, then include past ownership as an explanatory variable in column (3). Controlling for past market shares, the distance premium of an additional same-firm establishment is 62%, similar to that in our previous benchmark specification. When we include past ownership as an additional covariate, both past and contemporaneous ownership are positively associated with trade flows.[20] In the final column, we apply the most comprehensive set of fixed effects possible, those at the sending establishment × destination ZIP code pair level. Our regression exploits only variation in ownership between 2002 and 2007 within these pairs. (We omit distance as an explanatory variable, because it does not vary within i^e–z pairs.) Our coefficient estimate on the common ownership term is 2.77, slightly smaller than the coefficient from our benchmark specification.

Overall, across a wide variety of specifications, we report a substantial, economically meaningful distance premium of common ownership. In reconciling our large distance premium with low overall internal shares (reported in our earlier work, Atalay, Hortaçsu, and Syverson 2014), note that, for most sending establishments i^e, only a small fraction of the potential recipients of i^e's shipments belong to the same firm as i^e. Even if common ownership confers a substantially higher probability an establishment will send to a particular recipient, average internal shares will remain small because there are so few commonly owned potential recipients.

IV.B. Interactions with Industry Characteristics

We build on our benchmark analysis by exploring whether there are systematic variations in the associations among distance, ownership, and transactions. We begin in Figure II, with plots of the coefficient estimates and confidence intervals of the relationships between distance and our market share variable (left) and the relationships between the same-firm ownership fraction and the sending establishment's market share (right) for the 19 broadly defined industries that comprise our sample.[21]

20. The positive coefficient on past ownership is consistent with previous work documenting that postmerger restructuring often takes several years (e.g., Focarelli and Panetta 2003).

21. For the most part, these industries are defined at the three-digit level. However, to maintain sufficiently large sample sizes to conform with census disclosure avoidance rules, we combine some three-digit industries: Food is the

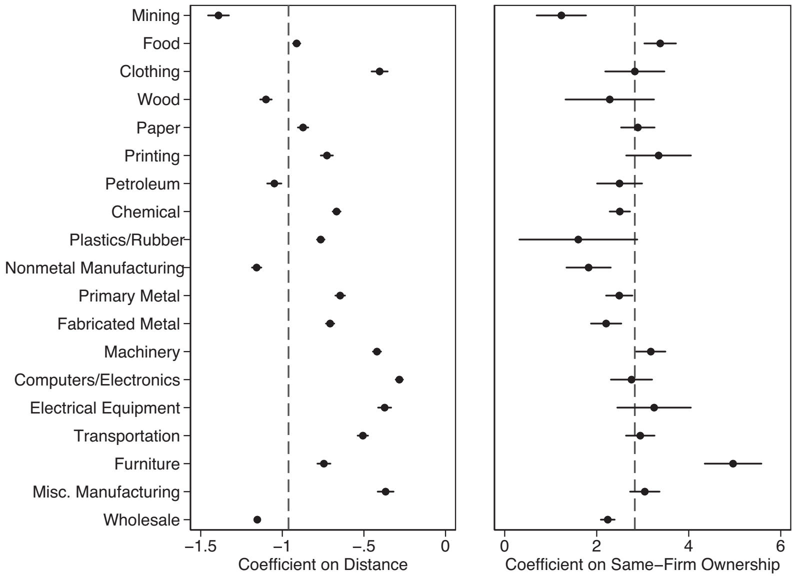

FIGURE II

Coefficient Estimates and Confidence Intervals, by Two/Three-Digit Industry of the Sending Establishment

The left side gives the coefficient estimate (and corresponding ±1.96 standard error confidence interval) of the log of mileage on the sending establishment's market share. The right side gives the coefficient estimate and corresponding confidence interval of the same-firm ownership share variable. These coefficients and confidence intervals result from a specification analogous to Table II, column (2), run separately for each two- or three-digit NAICS industry. The dashed lines in each panel present the coefficient estimates from the pooled sample.

Unsurprisingly, industries with the strongest relationship between trade flows and distance produce bulky (and thus costly to ship) products: mining, nonmetal manufacturing, and wood. In

combination of NAICS codes 311 and 312; Clothing is the combination of NAICS codes 313, 314, 315, and 316. Wholesale is the combination of NAICS codes 421 through 429. Complementing this section's analysis, in our earlier article (Atalay, Hortaçsu, and Syverson 2014, Online Appendix Table A.4) we also explored differences across industries. There, we computed the fraction of establishments which are vertically integrated (for which there is a same-firm plant in an industry downstream of the sender) and the share of vertically integrated establishments with any within-firm shipments. To highlight some of the results from that table, less than 40% of the sampled furniture manufacturers were at the upstream end of a within-firm production chain. In contrast, more than 90% of petroleum refiners were.

addition, trade flows are more responsive to distance in the wholesale sector than in manufacturing. Industries with the largest estimates of α_2 (the coefficient on the same-firm ownership fraction) include furniture, printing, and electrical equipment. Conversely, for the mining, nonmetal manufacturing, wood, and wholesale industries, the coefficient estimates of α_2 are relatively small. In combination, these estimates suggest that trade flows respond more heavily to distance for certain perhaps heavy-to-ship products and to common ownership in other industries.

Returning to the benchmark sample of 190 million observations, we interact the key explanatory variables in the specifications with several measures of industry attributes. The results are shown in Table V. In the first column, we group industries by the average value-to-weight ratio of shipments made by industry establishments in our CFS sample. Low value-to-weight (i.e., bulky) shipments exhibit a stronger relationship with distance, consistent with our results above. Moreover, the relationship between trade flows and firm ownership is stronger for these high value-to-weight commodities. Both patterns imply that our distance premium of common ownership is greater for high value-to-weight commodities. Specifically, the distance premium for above-median value-to-weight commodities is 77% ($=1-\exp[\frac{(2.460+1.038)\cdot 0.315}{(-1.075+0.330)}]$).[22] It is 51% for below-median value-to-weight commodities.

The second column of Table V probes the determinants of trade flows separately for goods distributors (mainly wholesalers, but also some mail-order retail catalogs) and goods producers (manufacturers and mining establishments). Bernard et al. (2010) and Ahn, Khandelwal, and Wei (2011), among others, demonstrate that wholesalers have different exporting patterns compared to manufacturers and play a special role in facilitating international trade. Complementary to this work, we find that the domestic shipments of wholesalers/mail-order retailers and manufacturers/mining establishments differ as well. First, the shipments

22. To compute this distance premium, the three relevant numbers are (i) the increase in the same-firm ownership fraction from an additional commonly owned downstream establishment in the destination ZIP code, 0.315; (ii) the slope of the relationship between the sending establishment's market share and the same-firm ownership fraction for above-median value-to-weight commodities, 2.460 + 1.038; and (iii) the slope of the relationship between the sender's market share and the log mileage variable for the same set of commodities, −1.075 + 0.330. The numbers within (ii) and (iii) come from Table V, column (1), adding the main and interaction effects.

TABLE V

RELATIONSHIP BETWEEN DISTANCE, COMMON OWNERSHIP, AND MARKET SHARES: INTERACTIONS WITH INDUSTRY CHARACTERISTICS

Dependent variable: $\frac{X_{zi^e}}{X_z}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Same-firm ownership fraction	2.460	3.135	2.584	2.552	2.731	2.576	3.103
	(0.066)	(0.060)	(0.101)	(0.105)	(0.093)	(0.097)	(0.103)
Log mileage	− 1.075	− 0.811	− 0.974	− 0.939	− 0.869	− 0.864	− 0.707
	(0.004)	(0.004)	(0.009)	(0.009)	(0.005)	(0.008)	(0.007)
Same-firm fraction × value-to-weight indicator	1.038						
	(0.097)						
Same-firm fraction × indicator for distributors		− 0.851					
		(0.097)					
Same-firm fraction × differentiated goods indicator			0.304	0.381			
			(0.126)	(0.129)			
Same-firm fraction × traded-on-exchange indicator			0.102	0.134			
			(0.294)	(0.263)			
Same-firm fraction × IT-intensity indicator					0.314		
					(0.125)		
Same-firm fraction × e-commerce indicator						0.441	
						(0.123)	
Same-firm fraction × capital intensity indicator							− 0.381
							(0.132)
Log mileage × value-to-weight indicator	0.330						
	(0.007)						
Log mileage × indicator for distributors		− 0.351					
		(0.006)					
Log mileage × differentiated goods indicator			0.262	0.224			
			(0.011)	(0.011)			
Log mileage × traded-on-exchange indicator			0.012	0.012			
			(0.026)	(0.021)			
Log mileage × IT-intensity indicator					0.246		
					(0.009)		
Log mileage × e-commerce indicator						0.161	
						(0.009)	
Log mileage × capital intensity indicator							− 0.106
							(0.009)
Rauch's classification	—	—	Conserv.	Liberal	—	—	—
In-sample mean: $\frac{1}{(1+r_{i^e z})}$	0.315	0.315	0.343	0.343	0.339	0.339	0.339

Notes. All regressions include sending-establishment fixed effects. In column (3), "Conserv." refers to Rauch's conservative classification, which assigns more commodities to be classified as reference-priced or differentiated. Rauch's liberal classification assigns a larger fraction of commodities as sold on an organized exchange. In columns (3) and (4), the omitted category includes reference-priced goods. The sample in columns (1) and (2) includes 190 million observations, representing 35,000 establishments. The sample in columns (3) and (4) includes 49 million observations, representing 16,000 establishments. The sample in columns (5), (6), and (7) includes 56 million observations, representing 18,000 establishments. There are 100 million observations corresponding to distributors (column (1)); 57 million observations corresponding to high value-to-weight industries (column (2)); 700,000 observations corresponding to exchange-traded commodities and 38 million observations corresponding to differentiated products using the conservative Rauch classification (column (3)); 1.3 million observations corresponding to exchange-traded commodities and 37 million observations corresponding to differentiated products using the liberal Rauch classification (column (4)); 20 million observations corresponding to high IT intensity industries (column (5)); 35 million observations corresponding to high e-commerce intensity industries (column (6)); and 38 million observations corresponding to high capital intensity industries (column (7)).

of distributors are more sensitive to distance, consistent with Hillberry and Hummels's (2003) characterization of manufacturers and wholesalers belonging to a hub-and-spoke arrangement.[23] Moreover, the relationship between shipment intensity and common ownership is weaker for distributors (see the "Same-firm fraction × indicator for distributors" term). When comparing the two effects, we see that the distance premium is 46% for distributors and 70% for establishments in other industries. In the remaining columns of Table V, the industry-level variables are measured only for the manufacturing sector, meaning we examine the interactions of observable characteristics within the subset of establishments with the 70% distance premium.

In columns (3) and (4), we apply Rauch's (1999) classification to check whether common ownership plays a larger role in facilitating physical input flows for goods more likely to involve relationship-specific investments. Rauch classifies manufactured products into three categories, in ascending order of relationship specificity: products that are traded on an organized exchange; those that are not traded in an organized market but are reference priced in trade publications; and those which are neither exchange traded nor reference priced. We find that for the most differentiated products—those in the last of the three categories—the slope of the relationship between market shares and the same-firm ownership fraction is significantly larger than it is for reference-priced commodities or exchange-traded commodities. The distance premium for these differentiated products is 75%, and it is 60% for reference-priced products, and 62% for exchange-traded products.[24] The larger value for differentiated products is consistent with Monteverde and Teece (1982), Masten (1984), and Masten, Meehan, and Snyder (1989, 1991), all of whom posit that the potential for costly hold up between an input supplier and input customer will tend to be larger for products that are complex or specific to the customer-supplier relationship.

23. According to Hillberry and Hummels, in this hub-and-spoke configuration, "[g]oods are manufactured in the hub and dispersed, sometimes at great distances, to a number of wholesaling spokes spread throughout the country. The wholesaling spokes then distribute, over very short distances, to retailers" (1090).

24. In computing these premia, note that within the subsample in columns (3) and (4) an additional same-firm establishment in the destination ZIP code increases the same-firm ownership fraction by 0.343, as opposed to 0.315 in the benchmark sample in columns (1) and (2).

In columns (5) and (6), we consider industries' use of new technologies. In column (5), we group industries based on the ratio of their investment in information technology to their total value of shipments. The results in Table V, column (5) indicate a distance premium for industries with above-median IT intensities of 81%, compared to 66% for below-median industries. In column (6), we group industries based on the fraction of their sales conducted through the internet. Industries with above-median e-commerce shares have a distance premium of 77%, as opposed to a 64% distance premium for low e-commerce industries. These results complement Acemoglu et al. (2007), along with more recent work by Fort (2017) and Forman and McElheran (2017), which tie the arrival of new information technologies to an increase in production fragmentation. In our setup, this would correspond to a decline in the average same firm ownership fraction, with larger declines occurring in more IT-intensive industries. Here, we find that the relationship between the volume of shipments and common ownership is stronger for IT-intensive industries for a given configuration of establishments across firms and locations.

Finally, in the international setting, Antràs (2003, 2005) demonstrates that intrafirm shipments are more prevalent in industries with a higher capital intensity and in countries with higher capital-labor ratios. Motivated by these results, in the final column of Table V, we compare the relationships between shipment intensity, common ownership, and distance by the capital intensity (dollar value of capital stock divided by total value of shipments) of an industry. The distance premia for above-median and below-median capital intensity industries are, respectively, 68% and 77%. It is unclear if capital intensity has much bearing on the relative importance between distance and firm ownership on domestic trade flows.

Table VI summarizes the results from this section. Overall, we find that the distance premium of ownership is significantly greater for high value-to-weight commodities, for producers (as opposed to distributors), for differentiated commodities, for commodities with IT-intensive production technologies, and for commodities with a high fraction of e-commerce sales.

IV.C. Quasi-Exogenous Changes in Common Ownership

Up to this point, we have refrained from lending a causal interpretation to our regression estimates. Location and

TABLE VI

DISTANCE PREMIUM OF COMMON OWNERSHIP

Industry characteristics	Characteristic 1: distance premium	Characteristic 2: distance premium	Characteristic 3: distance premium
Value-to-weight	Above median : 0.77	Below median: 0.51	
Producers or distributors	Producers: 0.70	Distributors: 0.46	
Differentiation (conserv.)	Exchange traded: 0.62	Reference priced: 0.60	Differentiated: 0.75
Differentiation (liberal)	Exchange traded: 0.63	Reference priced: 0.61	Differentiated: 0.76
IT intensity	Above median: 0.81	Below median: 0.66	
E-commerce intensity	Above median: 0.77	Below median: 0.64	
Capital intensity	Above median: 0.68	Below median: 0.77	

Notes. This table presents the distance premium of common ownership for different sets of commodities. The distance premium gives the reduction in distance equivalent to—in its relationship with trade flows—an additional same-firm downstream establishment in the destination ZIP code. The seven rows of this table correspond to the seven columns of Table V.

ownership choices could well be endogenous to expected shipment destinations. There could be unobserved factors specific to i^e–z pairs which make both common ownership and trade flows more prevalent. Previous work has detected many factors, including: common social identities (Combes, Lafourcade, and Mayer 2005), transportation infrastructure (Giroud 2013; Donaldson 2018), and communication links (Portes and Rey 2005). Moreover, establishment pairs i^e–z^e for which the idiosyncratic returns to trading are exceptionally high may find it optimal to merge with one another. Either these omitted variables or the endogeneity of s_{zi^e} would lead our previous regressions to overstate the causal impact of common ownership on trade flows.

Recognizing these issues, we seek to identify the causal effect of ownership on shipment patterns by using instances where firms acquire establishments for reasons other than the favorability (or lack thereof) of those establishments' locations vis-à-vis their expected shipments. Namely, we look at cases where new within-firm vertical links are created when a subset of establishments experiences an ownership change that is incidental to a large multiestablishment acquisition by its new parent firm. The logic of this approach is that when two multi-industry firms merge—or when a multi-industry firm purchases multiple establishments from another firm—it is unlikely that those establishments in the merging firms' secondary and tertiary lines of business triggered the acquisition. The identifying assumption is that the acquiring firm's motivation for the merger was to acquire the establishments in the acquired firm's primary lines of business, not so that it could own a peripheral establishment.[25]

25. Hastings and Gilbert (2005) and Hortaçsu and Syverson (2007) use a related strategy of exploiting within-firm, cross-market variation following a

We implement our approach as follows. We first use the LBD to identify mergers that occurred between 2002 and 2007.[26] From the set of establishments that were part of a merger or acquisition, we define our subset of incidental merger establishments by identifying establishments that satisfy the following criteria: (i) both the acquired firm and the acquiring firm contain at least three segments, where a segment is defined by four-digit NAICS codes; and (ii) the sending establishment's sector is in neither of the premerger firms' top S segments.[27] Among the 35,000 establishments in our benchmark sample, 2,400 satisfy criteria (i) and (ii) when S equals 1 (i.e., 2,400 establishments were acquired between 2002 and 2007 and did not belong to either the acquiring or the acquired firm's top segment), and 1,100 satisfy criteria (i) and (ii) for S equal to 3.

Figure III illustrates these criteria for a hypothetical merger between two firms. In this figure, there are two firms, where each firm has multiple establishments across multiple business segments. Each symbol represents a separate establishment in one of seven possible segments: automotive transportation, airplane manufacturing, bicycle manufacturing, computer manufacturing, electric lighting manufacturing, ship manufacturing, and tire manufacturing. Before the merger, the top three segments for Firm 1 are automotive manufacturing, airplane manufacturing, and bicycle manufacturing. For Firm 2, the top segments are automotive manufacturing, tire manufacturing, and airplane manufacturing. Because both firms have multiple establishments in more than three segments, a merger of the two firms would satisfy the first two criteria of the previous paragraph.

---

multiple-market merger to identify the effect of firm boundaries. In these earlier papers, the dependent variable of interest was the downstream market price rather than the propensity to ship to a given location, as is the focus here.

26. We define establishment i^e as being purchased in a merger or acquisition in year t if three conditions are met. First, i^e's firm identifier switches between year t and year $t+1$. Second, i^e's new firm identifier, as of year $t+1$, was already present as of year t (i.e., there was already existing a firm which could potentially have acquired i^e). This second criterion is necessary because it rules out several common scenarios—like changes in legal form of organization—which are unrelated to a change of ownership but are associated with changes in firm identifiers. Third, we require that i^e's firm identifier does not revert back to its original identifier in year $t+2$ or later.

27. For the purpose of ranking each firm's top segments, we include establishments in all sectors, not only those in the CFS sample frame. We rank segments according to the payroll of the establishments within each segment.

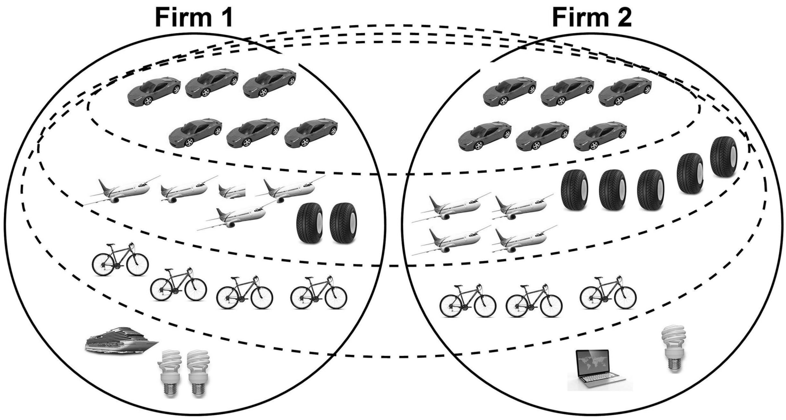

FIGURE III

Incidental Merger Example

Both firms have multiple segments, with each segment potentially containing multiple establishments. Each establishment is represented by an individual symbol (e.g., with a car representing an automotive manufacturing plant; a plane representing an airplane manufacturer). The three dashed ellipses, for $S \in \{1, 2, 3\}$, enclose the establishments that are excluded from the set of incidental merger establishments.

Depending on the chosen value of S, the number of plants classified as "incidental" to the merger would vary. With $S = 1$, all establishments outside of automotive manufacturing would be classified as incidental merger plants. For $S = 3$, ship, electric lighting, and computer manufacturing plants would be classified as incidental to the merger.

After identifying the incidental mergers in the sample, we construct an instrumental variable for our same-firm ownership fraction. For each i^e–z pair, we count the number of establishments in z (belonging to an industry that is downstream of i^e) that belong to the same firm as i^e as a result of an incidental merger but were part of a different firm from i^e before the merger. Our instrument takes this count and then divides by the number of total plants in z that are downstream of i^e.[28] For establishments i^e that were not part of an incidental merger, our instrument is equal to 0.

28. With S equal to 1, there are 14,400 sending establishment-destination ZIP code pairs for which our instrumental variable is greater than 0. With S equal to 2, the number of observations for which our instrument is greater than 0 decreases to 8,900. With S equal to 3, this same figure falls to 5,300.

TABLE VII

RELATIONSHIP BETWEEN DISTANCE, COMMON OWNERSHIP, AND MARKET SHARES: CONTROL FUNCTION ESTIMATES

Dependent variable: $\frac{X_{zi^e}}{X_z}$	Control function estimates			Baseline
	(1)	(2)	(3)	(4)
Same-firm ownership fraction	1.785	1.815	1.607	2.828
	(0.322)	(0.371)	(0.582)	(0.049)
Log mileage	– 0.963	– 0.963	– 0.963	– 0.962
	(0.003)	(0.003)	(0.003)	(0.003)
Residual from the first stage	1.050	1.016	1.223	–
	(0.325)	(0.374)	(0.584)	–
First Stage:				
Fraction of establishments in z in an incidental merger	1.015	1.027	1.028	–
	(0.001)	(0.001)	(0.001)	–
Number of segments	1	2	3	–

Notes. All regressions include sending-establishment fixed effects. The first-stage regressions also include log mileage as a covariate. The sample includes 190 million i^e–z pairs, drawing on the shipments made by 35,000 establishments. In the final row, "Number of segments" refers to the S we used when identifying which establishments were part of an incidental merger. In all specifications, we calculate the unweighted multilateral resistance terms. The last column reports our baseline results (Table II, column (2)) without attempting to address potential endogeneity in the same-firm ownership fraction variable.

Because of our large sample size and nonlinear gravity specification, we implement the estimation using a two-stage control-function-based estimator. In the first stage, we use a linear regression to regress our endogenous same-firm ownership fraction on the instrumental variable along with log mileage and sending-establishment fixed effects. The residual from this regression is then included as an additional covariate in a second-stage regression, which, as before, is a fixed effect Poisson model. In Online Appendix D, we discuss the underlying assumptions needed for consistent estimates and report the results from a Monte Carlo exercise on this approach. In the Monte Carlo simulations, we find that our control-function estimator provides precise and unbiased estimates for samples with a few hundred ZIP codes and sending establishments, samples that are smaller than the ones used in this section.

The first three columns of Table VII present our control function estimates. Here, the coefficient estimate of the same-firm ownership fraction is approximately one-third smaller than the estimates in Table II. (On the other hand, the estimates related to the importance of distance are as before.) Increasing the same-firm ownership fraction in the destination ZIP code by 0.315

TABLE VIII

RELATIONSHIP BETWEEN DISTANCE, COMMON OWNERSHIP, AND MARKET SHARES: SENSITIVITY ANALYSIS TO CONTROL FUNCTION ESTIMATES

Dependent variable: $\frac{X_{zi^e}}{X_z}$	(1)	(2)	(3)	(4)	(5)	(6)
Same-firm ownership fraction in 2007	1.293	1.575	1.246	1.346	1.258	1.359
	(0.549)	(0.686)	(0.452)	(0.558)	(0.442)	(0.540)
Log mileage	− 0.912	− 0.912	− 0.792	− 0.793	− 0.792	− 0.792
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
$X_{zi^e} \cdot (X_z)^{-1}$ from 2002			2.159	2.159	2.151	2.151
			(0.017)	(0.017)	(0.017)	(0.017)
Same-firm ownership fraction from 2002					1.415	1.345
					(0.326)	(0.393)
Residual from the first stage	1.689	1.401	1.176	1.107	0.529	0.424
	(0.555)	(0.691)	(0.459)	(0.546)	(0.453)	(0.551)
First Stage:						
Fraction of establishments in z in an incidental merger	1.028	1.035	1.028	1.035	1.038	1.050
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Number of segments	1	2	1	2	1	2

Notes. All regressions include sending-establishment fixed effects. The first-stage regressions also include log mileage as a covariate. In addition, if included in the second stage, the first-stage regressions include the 2002 values of same-firm ownership fraction and market shares as explanatory variables. The sample includes 43 million i^e–z pairs, drawing on the shipments made by 9,000 establishments included in the 2002 and 2007 versions of the CFS. In the final row, "Number of segments" refers to the S we used when identifying which establishments were part of an incidental merger. In all specifications, we calculate the unweighted multilateral resistance terms.

(corresponding to adding a single commonly owned establishment in that ZIP code) has the same impact on trade flows as decreasing the distance between the origin and destination by 40%.[29]

Our incidental merger instrument exploits changes in ownership, yet our Table VII regression uses variation from a single cross-section of the CFS. To get at a panel-like design, in Table VIII we extend our analysis to include data on past ownership and trade flows. We first replicate the first two columns of Table VII using the subset of establishments that are surveyed in the 2002 and 2007 vintages of the CFS. Our estimates of the effect of the distance premia of common ownership are somewhat lower, by approximately a quarter when $S = 1$ and a tenth when $S = 2$. In columns (3) and (4), we include previous shipment behavior as an explanatory variable. Based on the coefficient estimates

29. Head and Mayer (2014, Table 4) report that, in the context of trade across countries, the effect on trade flows of a common language is equivalent to a 30% reduction in distance. The effect of a colonial link is equivalent to a 50% distance reduction. Our 40% figure lies in between these two distance premia.

from these two columns, the distance premia of common ownership equals 39% and 42%, respectively. These premia are identical to those from Table VII. In the final two columns of Table VIII, we introduce past ownership and find that this variable is positive and statistically significant. Its inclusion, however, does not alter our estimates of the distance premium of common ownership.

IV.D. Sensitivity Analysis

In Online Appendix C, we perform eight sets of exercises to explore the sensitivity of the results in this section. (This is in addition to the robustness checks previously described in notes 11, 14, and 18.) First, our definition of the set of establishments with which a supplier can potentially enter into a trading relationship relies on choosing a cutoff value (of the share of the upstream industry's sales that are purchased by the downstream industry) in order to determine which pairs of industries are vertically linked with one another. Choosing a higher cutoff leads us to define fewer industries as vertically linked, in turn leading to fewer establishments in each destination ZIP code that are potential receivers of i^e's shipment. We verify that our main results are robust to our choice of cutoff value. In our second exercise, we argue that the distance premium of common ownership is the same for establishments belonging to small versus large firms. Third, we assess whether the distance premium varies with the level of geographic aggregation. We reestimate our regressions with counties as opposed to ZIP codes as the geographic region. Then we reestimate our regressions on the subsample of ZIP codes with the number of establishments in the destination exceeding progressively larger thresholds. Fourth, we evaluate the impact of different assumptions on the spatial correlation of the standard errors. Fifth, we verify that our main results are robust to different weighting methods—whether we use payroll to weight establishments when computing the same-firm ownership fraction or whether we use CFS sampling weights. Sixth, our sample of sending establishments and domestic ZIP codes excludes exports and imports. We demonstrate that our estimate on the same-firm ownership fraction is nearly identical for the subsample of industries for which the export intensity is low (less than 10%) or high. Seventh, we assess whether our estimated interaction of distance and ownership on trade flows (Table III, columns (1) and (3)) remains the same after accounting for the endogeneity of firm ownership. Finally, as an alternative to the control function approach, we apply a GMM procedure—due to Wooldridge (1997) and Windmeijer (2000)—to estimate the relationship between trade flows, common ownership, and distance. Here, both the coefficient estimates and the standard errors are somewhat larger than those in Table VII.

V. Conclusion

Establishments are substantially more likely to ship to destinations that are (i) close by and (ii) contain downstream establishments that share ownership with the sender. In this article, we used data on shipments made by tens of thousands of establishments throughout the manufacturing, mining, and wholesale sectors of the United States to characterize the relationships between transaction volume, distance, and common ownership. We find that all else equal, establishments send internal shipments further (or, equivalently, have a greater propensity to make internal shipments at any given distance). The magnitude of this differential willingness to ship implies that the shadow benefit of internal transactions is substantial: an extra same-firm downstream establishment in the destination ZIP code has roughly the same effect on transaction volumes as a 40% reduction in distance. In Online Appendix E, we apply these estimates to a simple multisector general equilibrium trade model. This exercise suggests that there could be a notable aggregate reduction in both trade flows and welfare from current levels without the trade-enhancing effects of common ownership.[30]

Quantifying the magnitude and aggregate effects of other benefits associated with common ownership—beyond facilitating physical input flows—is an exciting topic for future research. In an earlier paper (Atalay, Hortaçsu, and Syverson 2014), we argued that the primary motivation for common ownership of production chains is to share intangible inputs across establishments, with the mitigation of transaction costs as a secondary concern. However, due to data limitations, we could only provide circumstantial

30. The Online Appendix E exercise aims to gauge the aggregate importance of the benefits of vertical integration that specifically relate to sourcing physical inputs more easily. This exercise does not attempt to quantify other benefits of common ownership via the sharing of intangible inputs. Nor does it attempt to measure the private (e.g., due to managers' limited span of control) or societal (e.g., due to decreased competition) costs of vertical integration.

evidence in favor of the intangible input hypothesis.[31] Now, thanks to new survey data being collected and linked to census micro data (Buffington et al. 2017; Bloom et al. 2019), it is possible to directly quantify the extent to which profitability-increasing management practices respond to changes in firm boundaries (Bai, Jin, and Serfling 2018), and thus it should also be possible to evaluate aggregate productivity in counterfactual environments in which firms' sharing of intangible managerial inputs is muted.

Federal Reserve Bank of Philadelphia
University of Chicago
Compass Lexecon
University of Chicago

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online. Code replicating tables and figures in this article can be found in Atalay et al. (2019), in the Harvard Dataverse, doi:10.7910/DVN/MCZLLB.

References

Acemoglu, Daron, Philippe Aghion, Rachel Griffith, and Fabrizio Zilibotti, "Vertical Integration and Technology: Theory and Evidence," *Journal of the European Economic Association*, 8 (2010), 989–1033.

Acemoglu, Daron, Philippe Aghion, Claire Lelarge, John Van Reenen, and Fabrizio Zilibotti, "Technology, Information, and the Decentralization of the Firm," *Quarterly Journal of Economics*, 122 (2007), 1759–1799.

Ahn, JaeBin, Amit K. Khandelwal, and Shang-Jin Wei, "The Role of Intermediaries in Facilitating Trade," *Journal of International Economics*, 84 (2011), 73–85.

Alfaro, Laura, Pol Antràs, Davin Chor, and Paola Conconi, "Internalizing Global Value Chains: A Firm-Level Analysis," *Journal of Political Economy*, 127 (2019), 508–559.

Anderson, James E., and Eric van Wincoop, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93 (2003), 170–192.

———, "Trade Costs," *Journal of Economic Literature*, 42 (2004), 691–751.

Antràs, Pol, "Firms, Contracts, and Trade Structure," *Quarterly Journal of Economics*, 118 (2003), 1375–1418.

———, "Property Rights and the International Organization of Production," *American Economic Review*, 95 (2005), 25–32.

31. We wrote: "It is difficult to directly test our 'intangible input' explanation for vertical ownership structures because such inputs are by definition hard to measure. Ideally, we would have information on the application of managerial or other intangible inputs (like managers' time-use patterns across the different business units of the firm) across firm structures. Such data do not exist for the breadth of industries which we are looking at here, however" (1141).

Antràs, Pol, and Davin Chor, “Organizing the Global Value Chain,” *Econometrica*, 81 (2013), 2127–2204.

Antràs, Pol, and Stephen R. Yeaple, “Multinational Firms and the Structure of International Trade,” *Handbook of International Economics*, 4 (2014), 55–130.

Atalay, Enghin, Ali Hortaçsu, Mary Jialin Li, and Chad Syverson, “Replication Data for: ‘How Wide Is the Firm Border?’,” (2019), Harvard Dataverse, doi:10.7910/DVN/MCZLLB.

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson, “Vertical Integration and Input Flows,” *American Economic Review*, 104 (2014), 1120–1148.

Bai, John, Wang Jin, and Matthew Serfling, “Management Practices and Mergers and Acquisitions,” Mimeo, 2018.

Baier, Scott L., and Jeffrey H. Bergstrand, “Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Effects Using the Gravity Equation,” *Journal of International Economics*, 77 (2009), 77–85.

Baier, Scott L., Jeffrey H. Bergstrand, Peter Egger, and Patrick A. McLaughlin, “Do Economic Integration Agreements Actually Work? Issues in Understanding the Causes and Consequences of the Growth of Regionalism,” *World Economy*, 31 (2008), 461–497.

Baker, George P., and Thomas N. Hubbard, “Make versus Buy in Trucking: Asset Ownership, Job Design, and Information,” *American Economic Review*, 93 (2003), 551–572.

———, “Contractibility and Asset Ownership: On-Board Computers and Governance in U.S. Trucking,” *Quarterly Journal of Economics*, 119 (2004), 1443–1479.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott, “Wholesalers and Retailers in US Trade,” *American Economic Review*, 95 (2010), 408–413.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott, “Transfer Pricing by U.S.-Based Multinational Firms,” Mimeo, 2006.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen, “What Drives Differences in Management Practices?,” *American Economic Review*, 109 (2019), 1648–1683.

Boehm, Johannes, “The Impact of Contract Enforcement Costs on Outsourcing and Aggregate Productivity,” Mimeo, 2017.

Boehm, Johannes, and Ezra Oberfield, “Misallocation in the Market for Inputs: Enforcement and the Organization of Production,” Mimeo, 2018.

Brainard, S. Lael, “An Empirical Assessment of the Proximity-Concentration Trade-Off between Multinational Sales and Trade,” *American Economic Review*, 87 (1997), 520–544.

Bruno, Randolph, Nauro Campos, Saul Estrin, and Meng Tian, “Economic Integration, Foreign Investment and International Trade: The Effects of Membership of the European Union,” Mimeo, 2017.

Buffington, Catherine, Lucia Foster, Ron Jarmin, and Scott Ohlmacher, “The Management and Organizational Practices Survey (MOPS): An Overview,” *Journal of Economic and Social Measurement*, 42 (2017), 1–26.

Cai, Jing, and Adam Szeidl, “Interfirm Relationships and Business Performance,” *Quarterly Journal of Economics*, 133 (2018), 1229–1282.

Chipty, Tasneem, “Vertical Integration, Market Foreclosure, and Consumer Welfare in the Cable Television Industry,” *American Economic Review*, 91 (2001), 428–453.

Coase, Ronald H., “The Nature of the Firm,” *Economica*, 16 (1937), 386–405.

Combes, Pierre-Philippe, Miren Lafourcade, and Thierry Mayer, “The Trade-Creating Effects of Business and Social Networks: Evidence from France,” *Journal of International Economics*, 66 (2005), 1–29.

Costinot, Arnaud, and Andrés Rodríguez-Clare, “Trade Theory with Numbers: Quantifying the Consequences of Globalization,” *Handbook of International Economics*, 4 (2014), 1880–2002.

Davies, Ronald B., Julien Martin, Mathieu Parenti, and Farid Toubal, “Knocking on Tax Haven’s Door: Multinational Firms and Transfer Pricing,” *Review of Economics and Statistics*, 100 (2018), 120–134.

Disdier, Anne-Célia, and Keith Head, “The Puzzling Persistence of the Distance Effect on Bilateral Trade,” *Review of Economics and Statistics*, 90 (2008), 37–48.

Donaldson, Dave, “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,” *American Economic Review*, 108 (2018), 899–934.

Eaton, Jonathan, Samuel Kortum, and Sebastian Sotelo, “International Trade: Linking Micro and Macro,” in *Advances in Economics and Econometrics: Theory and Applications*, Tenth World Congress, Daron Acemoglu, Manuel Arellano and Eddie Dekel, eds. (Cambridge: Cambridge University Press, 2012), 329–370.

Engel, Charles, and John Rogers, “How Wide Is the Border?,” *American Economic Review*, 86 (1996), 1112–1125.

Fally, Thibault, and Russell Hillberry, “A Coasian Model of International Production Chains,” *Journal of International Economics*, 114 (2018), 299–315.

Focarelli, Dario, and Fabio Panetta, “Are Mergers Beneficial to Consumers? Evidence from the Market for Bank Deposits,” *American Economic Review*, 93 (2003), 1152–1172.

Forbes, Silke J., and Mara Lederman, “Does Vertical Integration Affect Firm Performance? Evidence from the Airline Industry,” *RAND Journal of Economics*, 41 (2010), 765–790.

Forman, Chris, and Kristina McElheran, “The Digital Reorganization of Firm Boundaries: IT Use and Vertical Integration in U.S. Manufacturing,” Mimeo, 2017

Fort, Teresa C., “Technology and Production Fragmentation: Domestic versus Foreign Sourcing,” *Review of Economic Studies*, 84 (2017), 650–687.

Gibbons, Robert, “Four Formal(izable) Theories of the Firm?,” *Journal of Economic Behavior and Organization*, 58 (2005), 200–245.

Giroud, Xavier, “Proximity and Investment: Evidence from Plant-Level Data,” *Quarterly Journal of Economics*, 128 (2013), 861–915.

Grossman, Sanford J., and Oliver Hart, “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration,” *Journal of Political Economy*, 94 (1986), 691–719.

Hart, Oliver, and John Moore, “Property Rights and the Nature of the Firm,” *Journal of Political Economy*, 98 (1990), 1119–1158.

Hastings, Justine S., and Richard J. Gilbert, “Market Power, Vertical Integration and the Wholesale Price of Gasoline,” *Journal of Industrial Economics*, 53 (2005), 469–492.

Head, Keith, and Thierry Mayer, “Gravity Equations: Workhorse, Toolkit, and Cookbook,” *Handbook of International Economics*, 4 (2014), 131–195.

———, “Brands in Motion: How Frictions Shape Multinational Production,” *American Economic Review*, forthcoming.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple, “Export versus FDI with Heterogeneous Firms,” *American Economic Review*, 94 (2004), 300–316.

Hillberry, Russell, and David Hummels, “Intranational Home Bias: Some Explanations,” *Review of Economics and Statistics*, 85 (2003), 1089–1092.

———, “Trade Responses to Geographic Frictions: A Decomposition Using Micro-Data,” *European Economic Review*, 52 (2008), 527–550.

Holmstrom, Bengt, and Paul Milgrom, “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design,” *Journal of Law, Economics, and Organization*, 7 (1991), 24–52.

Holmstrom, Bengt, and Jean Tirole, “Transfer Pricing and Organizational Form,” *Journal of Law, Economics, and Organization*, 7 (1991), 201–228.

Hortaçsu, Ali, and Chad Syverson, “Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices,” *Journal of Political Economy*, 115 (2007), 250–301.

Kalnins, Arturs, and Francine Lafontaine, “Multi-Unit Ownership in Franchising: Evidence from the Fast-Food Industry in Texas,” *RAND Journal of Economics*, 35 (2004), 747–761.

———, "Too Far Away? The Effect of Distance to Headquarters on Business Establishment Performance," *American Economic Journal: Microeconomics*, 5 (2013), 157–179.
Keller, Wolfgang, and Stephen R. Yeaple, "The Gravity of Knowledge," *American Economic Review*, 103 (2013), 1414–1444.
Klein, Benjamin, Robert G. Crawford, and Armen Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, 21 (1978), 297–326.
Kogut, Bruce, and Udo Zander, "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology," *Organization Science*, 3 (1992), 383–397.
Lafontaine, Francine, and Margaret Slade, "Vertical Integration and Firm Boundaries," *Journal of Economic Literature*, 45 (2007), 629–685.
———, "Inter-Firm Contracts," in *Handbook of Organizational Economics*, Robert Gibbons and John Roberts, eds. (Princeton: Princeton University Press, 2013), 958–1013.
Markusen, James R., and Anthony J. Venables, "The Theory of Endowment, Intra-industry and Multinational Trade," *Journal of International Economics*, 52 (2000), 209–234.
Masten, Scott E., "The Organization of Production: Evidence from the Aerospace Industry," *Journal of Law and Economics*, 27 (1984), 403–417.
Masten, Scott E., James W. Meehan, Jr., and Edward A. Snyder, "Vertical Integration in the U.S. Auto Industry: A Note on the Influence of Transaction Specific Assets," *Journal of Economic Behavior and Organization*, 12 (1989), 265–273.
———, "The Costs of Organization," *Journal of Law, Economics, and Organization*, 7 (1991), 1–25.
McCallum, John, "National Borders Matter: Canada-U.S. Regional Trade Patterns," *American Economic Review*, 85 (1995), 615–623.
Monteverde, Kirk, and David J. Teece, "Supplier Switching Costs and Vertical Integration in the Automobile Industry," *Bell Journal of Economics*, 13 (1982), 206–212.
Nelson, Richard R., and Sidney G. Winter, "*An Evolutionary Theory of Economic Change*," (Cambridge, MA: Harvard University Press, 1982).
Portes, Richard, and Helene Rey, "The Determinants of Cross-Border Equity Flows," *Journal of International Economics*, 65 (2005), 269–296.
Ramondo, Natalia, and Andrés Rodríguez-Clare, "Trade, Multinational Production, and the Gains from Openness," *Journal of Political Economy*, 121 (2013), 273–322.
Rauch, James E., "Networks versus Markets in International Trade," *Journal of International Economics*, 48 (1999), 7–35.
Santos Silva, J. M. C., and Silvana Tenreyro, "The Log of Gravity," *Review of Economics and Statistics*, 88 (2006), 641–658.
Sotelo, Sebastian, "Practical Aspects of Implementing the Multinomial PML Estimator," Mimeo, 2019.
Wallis, John J., and Douglass North, "Measuring the Transaction Sector in the American Economy, 1870–1970," in *Long-Term Factors in American Economic Growth*, Stanley L. Engerman and Robert E. Gallman, eds. (Chicago: University of Chicago Press, 1986).
Wernerfelt, Birger, "A Resource-Based View of the Firm," *Strategic Management Journal*, 5 (1984), 171–180.
Williamson, Oliver, "The Vertical Integration of Production: Market Failure Considerations," *American Economic Review*, 61 (1971), 112–123.
———, "Markets and Hierarchies: Some Elementary Considerations," *American Economic Review*, 63 (1973), 316–325.
———, "Transaction-Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics*, 22 (1979), 233–261.
Windmeijer, Frank, "Moment Conditions for Fixed Effects Count Data Models with Endogenous Regressors," *Economics Letters*, 68 (2000), 21–24.
Wooldridge, Jeffrey M., "Multiplicative Panel Data Models without the Strict Exogeneity Assumption," *Econometric Theory*, 13 (1997), 667–678.

Online Appendix to *How Wide Is the Firm Border?*

Enghin Atalay, Ali Hortaçsu, Mary Jialin Li, Chad Syverson

A Calculations Related to Section II

In this appendix, we derive Equation 1. For the reader's convenience, portions of the text (particularly the following two paragraphs) draw on the exposition of Section II.

Each sending establishment has access to a (random) number of linear production technologies, each of which allows the plant to transform l units of labor into $V \cdot l$ units of output. We assume that V is Pareto distributed with shape parameter θ and a lower cutoff $\underline{v}$ that can be set arbitrarily close to 0. We also assume that the (integer) number of establishment i^e's varieties with efficiency $V > v$ (for $v > \underline{v}$) is the realization of a Poisson random variable with mean $T_{i^e} v^{-\theta}$. In this expression, the parameter T_{i^e} reflects the overall productivity of establishment i^e.

Call x_i the cost of a unit of labor inputs for establishments in ZIP code i. There are also iceberg-style transportation costs that vary not only in distance, but also based on ownership. Specifically, for establishment i^e to sell one unit of the commodity to plant z^e, it must produce $d_{zi} \geq 1$ units of output if plant z^e is owned by a different firm and $d_{zi}\delta_{zi} \geq 1$ units of output if the same firm owns it. Furthermore, forming a relationship with establishment z^e requires a fixed number of workers F_{z^e} to be hired in ZIP code z. Given these assumptions, the unit cost of a variety with an idiosyncratic productivity draw V selling to plant z^e is

$$\psi_{z^e i^e}(v) = \frac{x_i}{V} d_{zi} \left(\delta_{zi}\right)^{\mathbf{1}^{SF}(z^e, i^e)},$$

where $\mathbf{1}^{\mathrm{SF}}$ is an indicator for a within-firm relationship between establishments i^e and z^e.

Using properties of the Poisson distribution, the number of varieties that can be sold to establishment z^e at a cost less than or equal to ψ is the realization of a Poisson random variable with parameter $\Phi_{z^e}\psi^{\theta}$, with

$$\Phi_{z^e} \equiv \sum_{i=1}^{Z} \sum_{i^e \in i} T_{i^e} \left(x_i d_{zi}\right)^{-\theta} \cdot \left(\left(\delta_{zi}\right)^{\mathbf{1}^{SF}(z^e, i^e)}\right)^{-\theta},$$

where $i^e \in i$ indicates that we are summing over the set of plants that reside in ZIP code i.

Our last set of assumptions, again following the setup in Eaton, Kortum, and Sotelo (2012), relates to establishments' entry and pricing decisions. We assume that (a) upstream establishments compete monopolistically when serving each downstream establishment, (b) the downstream establishment z^e combines inputs from its suppliers according to a CES aggregator, (c) each upstream establishment takes as given both the downstream establishment's total expenditures X_{z^e} on intermediate inputs and its unit labor cost x_z, and (d) upstream establishments decide to sell to establishment z^e so long as the profits net of the fixed cost F_{z^e} are non-negative, with low-cost sending establishments making their decisions first. This setup provides three results concerning the margins of trade. First, conditional on selling a non-zero amount to recipient z^e, sales by different sending establishments are independent of the cost parameters x_i, d_{zi}, and δ_{zi}. These parameters affect only the extensive margin of trade, not the intensive margin. Second, the probability that a given variety produced by establishment i^e is among the lowest-cost varieties that are able to profitably enter is given by:

$$\begin{aligned}\pi_{z^e i^e} &= \frac{\Phi_{z^e i^e}}{\Phi_{z^e}}, \text{ with} \\ \Phi_{z^e i^e} &\equiv T_{i^e}\left(x_i d_{zi}\left(\delta_{zi}\right)^{\mathbf{1}^{SF}(z^e,i^e)}\right)^{-\theta}.\end{aligned} \tag{3}$$

Third, and related to the first two results, the fraction of z^e's expenditures purchased from upstream establishment i^e, in expectation, equals

$$\mathbb{E}\left[\frac{X_{z^e i^e}}{X_{z^e}}\right] = \frac{\Phi_{z^e i^e}}{\Phi_{z^e}}. \tag{4}$$

In Online Appendix B, we aggregate Equation 4 up to the sending establishment by destination ZIP code pair in order to match the level of aggregation in our data, as discussed above:

$$\begin{aligned}\pi_{zi^e} &\equiv \frac{\Phi_{zi^e}}{\Phi_z} \approx \mathbb{E}\left[\frac{X_{zi^e}}{X_z}\right], \text{ where} \\ \Phi_{zi^e} &\equiv T_{i^e}\left(x_i d_{zi}\right)^{-\theta}\left(1 - s_{zi^e} + s_{zi^e}\left(\delta_{zi}\right)^{-\theta}\right), \\ \Phi_z &\equiv \sum_{i'=1}^{Z} \sum\nolimits_{i'^e \in i'} \Phi_{zi'^e}, \text{ and} \\ s_{zi^e} &\equiv \sum_{z^e \in z} \frac{X_{z^e}}{X_z} \mathbf{1}^{SF}\left(z^e, i^e\right)\end{aligned} \tag{5}$$

is the expenditure-weighted share of downstream establishments in the destination ZIP code owned by the same firm of the sending establishment i^e. The $\left(1 - s_{zi^e} + s_{zi^e}(\delta_{zi})^{-\theta}\right)$ term reflects a weighted average of the trade-facilitating effects of common ownership: A fraction s_{zi^e} of the establishments in the destination shares ownership with the sender and has lower trade costs by a factor of δ_{zi}. For the remaining $1 - s_{zi^e}$ establishments in the destination, there is no analogous reduction in trade costs. Finally, throughout the paper, we use X_{zi^e}/X_z to refer to the market share of establishment i^e in ZIP code z. In the empirical analysis, in the body of the paper, this market share is specific to the industry of establishment i^e.

Consider a first-order Taylor approximation around the point at which sending establishment i^e has no same-firm establishments in the downstream ZIP code:[34]

$$1 + s_{zi^e}\left(\left(\delta_{zi}\right)^{-\theta} - 1\right) \approx \exp\left\{s_{zi^e}\left(\left(\delta_{zi}\right)^{-\theta} - 1\right)\right\}.$$

Using this approximation, we can rewrite Equation 5 as

$$\mathbb{E}\left[\frac{X_{zi^e}}{X_z}\right] \approx \frac{\exp\left\{\log T_{i^e} - \theta \log x_i - \theta \log d_{zi} + s_{zi^e}\left(\exp\left[-\theta \log \delta_{zi}\right] - 1\right)\right\}}{\sum_{i'=1}^{Z} \sum_{i'^e \in i'} \exp\left\{\log T_{i'^e} - \theta \log x_{i'} - \theta \log d_{zi'} + s_{zi'^e}\left(\exp\left[-\theta \log \delta_{zi'}\right] - 1\right)\right\}}. \tag{6}$$

We parameterize the relationship between distance and same-firm ownership on trade flows to be

$$\begin{aligned} -\theta \log d_{zi} + s_{zi^e}\left(\exp\left[-\theta \log \delta_{zi}\right] - 1\right) = \alpha_0 + \alpha_1 \cdot \log \text{mileage}_{z \leftarrow i} \\ + \alpha_2 \cdot s_{zi^e} + \alpha_3 \cdot s_{zi^e} \cdot \log \text{mileage}_{z \leftarrow i} + \log \varepsilon_{z,i^e}. \end{aligned} \tag{7}$$

In this equation, the ε_{z,i^e} reflect the random unobservable component of trade costs from establishment i^e to destination z, costs which are unrelated to mileage and common ownership. The ε_{z,i^e} are constructed as in Eaton, Kortum, and Sotelo (2012), as the ratio of Gamma-distributed random variables (see their footnote 21), and are independent across i^e–z pairs.[35]

Plugging Equation 7 into Equation 6 yields the following equation relating i^e's market share in destination ZIP code z:

$$\mathbb{E}\left[\frac{X_{zi^e}}{X_z}|\Lambda\right] \approx \frac{\exp\left\{\alpha_{i^e} + \alpha_1 \cdot \log \text{mileage}_{z \leftarrow i} + \alpha_2 \cdot s_{zi^e} + \alpha_3 \cdot s_{zi^e} \cdot \log \text{mileage}_{z \leftarrow i}\right\}}{\sum_{i'=1}^{Z} \sum_{i'^e \in i'} \exp\left\{\alpha_{i'^e} + \alpha_1 \cdot \log \text{mileage}_{z \leftarrow i'} + \alpha_2 \cdot s_{zi'^e} + \alpha_3 \cdot s_{zi'^e} \cdot \log \text{mileage}_{z \leftarrow i'}\right\}}.$$

This equation is equivalent to Equation 1 in Section II. As we write in that section of the paper, "conditioning on Λ indicates that there is some random component of trade barriers, namely that

[34] With this approximation, the relationship between the same-firm ratio s_{zi^e} and the expected market share is log-linear. Since in our sample the average value for s_{zi^e} equals 0.0009, the approximation error is inconsequential.

[35] First, define

$$\Lambda_{zi^e} \equiv \frac{\exp\left\{\alpha_{i^e} + \alpha_1 \cdot \log \text{mileage}_{z \leftarrow i} + \alpha_2 \cdot s_{zi^e} + \alpha_3 \cdot s_{zi^e} \cdot \log \text{mileage}_{z \leftarrow i}\right\}}{\sum_{i'=1}^{Z} \sum_{i'^e \in i'} \exp\left\{\alpha_{i'^e} + \alpha_1 \cdot \log \text{mileage}_{z \leftarrow i'} + \alpha_2 \cdot s_{zi'^e} + \alpha_3 \cdot s_{zi'^e} \cdot \log \text{mileage}_{z \leftarrow i'}\right\}}$$

as the observable component of trade costs. To compute ε_{z,i^e}, consider a set of random variables ϑ_{zi^e} drawn (independently across i^e–z pairs) from a Gamma distribution with scale parameter $\frac{\Lambda_{zi^e}}{\eta^2}$ and shape parameter $\frac{\eta^2}{\Lambda_{zi^e}}$, for some $\eta > 0$. The idiosyncratic components of trade costs are defined as $\varepsilon_{z,i^e} \equiv \frac{\vartheta_{zi^e}}{\vartheta_{ii^e}}$. Based on the properties of the Gamma distribution, with this parameterization the expression for the expected trade flows (conditional on the observable trade cost variables) retains a convenient multinomial logit form.

the relationship between d_{zi} and mileage — and, alternatively, between δ_{zi} and mileage —contains some random component. Furthermore, s_{zi^e} equals the fraction of establishments in the destination ZIP code z that share ownership with the establishment i^e. And, finally, $\alpha_{i^e} \equiv \alpha_0 + \log T_{i^e} - \theta \log x_i$ collects all of the relevant sending-establishment-specific unobservable terms."

B Derivation of Equation 5 from Equation 4

The goal of this appendix is to relate Equations 4 and 5. Begin with π_{zi^e}, the fraction of shipments to ZIP code z that come from establishment i^e. As a reminder, these calculations refer to the share of sales of a given product in ZIP code z that come from different sending establishments. As in Section II and Online Appendix A, we omit commodity or industry superscripts:

$$\begin{aligned}
\pi_{zi^e} &= \frac{\Phi_{zi^e}}{\Phi_z} \\
&= \frac{T_{i^e}(x_i d_{zi})^{-\theta}\left(1 - s_{zi^e} + s_{zi^e}(\delta_{zi})^{-\theta}\right)}{\sum_{i'=1}^{Z}\sum_{i'^e \in i} T_{i'^e}(x_{i'} d_{zi'})^{-\theta}(1 - s_{zi'^e} + s_{zi'^e}(\delta_{zi'})^{-\theta}))} \\
&= \frac{\sum_{z^e \in z} \frac{X_{z^e}}{X_z} T_{i^e}(x_i d_{zi}(\delta_{zi})^{\mathbf{1}^{SF}(z^e, i^e)})^{-\theta}}{\sum_{i'=1}^{Z}\sum_{i'^e \in i'}\sum_{z^e \in z} \frac{X_{z^e}}{X_z} T_{i'^e}(x_{i'} d_{zi'}(\delta_{zi'})^{\mathbf{1}^{SF}(z^e, i'^e)})^{-\theta}}.
\end{aligned}$$

In this expression, Φ_{zi^e} is the parameter associated with the Poisson distribution that characterizes the number of varieties that i^e can supply the average customer in z at a price less than ψ. Similarly, Φ_z parameterizes the distribution of the total number of varieties that can be supplied to z at a price less than ψ. In the equations above, the second line follows from the definitions of Φ_z and Φ_{zi^e}, while the third line follows from the definition of s_{zi^e} (which, again, is the expenditure-weighted fraction of establishments in the destination ZIP code that share ownership with the sender). Next, we apply the definition of Φ_{z^e}:

$$\begin{aligned}
\pi_{zi^e} &= \frac{\sum_{z^e \in z} \frac{X_{z^e}}{X_z} \Phi_{z^e i^e}}{\sum_{z^e \in z} \frac{X_{z^e}}{X_z} \Phi_{z^e}} \\
&= \frac{\sum_{z^e \in z} \frac{X_{z^e}}{X_z} \frac{\Phi_{z^e i^e}}{\Phi_{z^e}} \Phi_{z^e}}{\sum_{z^{e'} \in z} \frac{X_{z^{e'}}}{X_z} \Phi_{z^{e'}}} \\
&= \sum_{z^e \in z} \frac{X_{z^e}}{X_z} \frac{\Phi_{z^e i^e}}{\Phi_{z^e}} \cdot \frac{\Phi_{z^e}}{\sum_{z^{e'} \in z} \frac{X_{z^{e'}}}{X_z} \Phi_{z^{e'}}} && (8) \\
&\approx \sum_{z^e \in z} \frac{X_{z^e}}{X_z} \frac{\Phi_{z^e i^e}}{\Phi_{z^e}} . && (9)
\end{aligned}$$

Above, the approximation results from the fact that the fraction $\Phi_{z^e} / \left(\sum_{z^{e'} \in z} \frac{X_{z^{e'}}}{X_z} \Phi_{z^{e'}} \right)$ is, on average (averaging over the establishments z^e in the destination z), close to but not equal to 1. To see this, note that

$$\frac{\Phi_{z^e}}{\sum_{z^{e'} \in z} \frac{X_{z^{e'}}}{X_z} \Phi_{z^{e'}}} = \left[\sum_{z^{e'} \in z} \frac{X_{z^{e'}}}{X_z} \cdot \frac{\sum_{i=1}^{Z} \sum_{i^e \in i} T_{i^e} (x_i d_{zi})^{-\theta} \cdot \left((\delta_{zi})^{\mathbf{1}^{SF}(z^{e'}, i^e)} \right)^{-\theta}}{\sum_{i=1}^{Z} \sum_{i^e \in i} T_{i^e} (x_i d_{zi})^{-\theta} \cdot \left((\delta_{zi})^{\mathbf{1}^{SF}(z^e, i^e)} \right)^{-\theta}} \right]^{-1}. \tag{10}$$

Thus, $\Phi_{z^e} / \left(\sum_{z^{e'} \in z} \frac{X_{z^{e'}}}{X_z} \Phi_{z^{e'}} \right)$ is substantially greater than 1 to the extent that z^e has more nearby same-firm establishments than the other establishments located in destination z. (Note that z^e only appears in the $\mathbf{1}^{SF}(z^e, i^e)$ term within the right-hand side of Equation 10.) Since Equation 8 sums over establishments in the destination, and since $\Phi_{z^e} \cdot \left(\sum_{z^{e'} \in z} \frac{X_{z^{e'}}}{X_z} \Phi_{z^{e'}} \right)^{-1}$ will tend to be above 1 for some destination establishments, tend to be below 1 for others, and near 1 on average, the right-hand side of Equation 8 will be close to the right-hand side of Equation 9. In the original Eaton, Kortum, and Sotelo formulation, there was no cost advantage of internal shipments: $\delta_{zi} = 1$. So, the only variables that shape i-to-z expected trade flows are the same for all destination ZIP code establishments. As a result, in Eaton, Kortum, and Sotelo (2012) there is no need for an approximation. In our context, the approximation error should be small.

Moving forward, we apply the definition of $\pi_{z^e i^e}$, and then use Equations 3 and 4 to substitute out the $\pi_{z^e i^e}$ terms:

$$\begin{aligned}
\pi_{zi^e} &\approx \sum_{z^e \in z} \frac{X_{z^e}}{X_z} \pi_{z^e i^e} \\
&= \sum_{z^e \in z} \frac{X_{z^e}}{X_z} \mathbb{E} \left[\frac{X_{z^e i^e}}{X_{z^e}} \right] \\
&= \sum_{z^e \in z} \mathbb{E} \left[\frac{X_{z^e i^e}}{X_z} \right] \\
&= \mathbb{E} \left[\frac{X_{zi^e}}{X_z} \right].
\end{aligned}$$

The final expression is equivalent to Equation 5.

C Additional Robustness Checks

In this section, we discuss 11 sets of robustness checks, aimed at examining the sensitivity of the Section IV results to alternative sample construction and estimation methods.

Table 9: Relationship between Distance, Common Ownership, and Market Shares: Sensitivity to Firm Size

Dependent Variable: $\frac{X_{zi^e}}{X_z}$	(1)	(2)	(3)	(4)	(5)
Same-firm ownership fraction	2.828	2.811	2.813	2.832	2.824
	(0.049)	(0.049)	(0.052)	(0.055)	(0.047)
Log mileage	-0.962	-0.987	-1.003	-1.019	-0.936
	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)
Firm Size to be in Sample	Multi-Unit	$\geqslant$5 Ests.	$\geqslant$10 Ests.	$\geqslant$20 Ests.	All

Notes: The first column reproduces column (2) of Table 2. In columns (2) through (5), we vary the sample according to the size of the firm of the sending establishment. In columns (1), (2), (3), (4), and (5), the sample sizes are 190 million, 149 million, 125 million, 103 million, and 302 million, respectively, representing the shipments made by 35,000, 27,000, 23,000, 18,000, and 57,000 establishments. In all specifications, we calculate the unweighted multilateral resistance terms.

In our benchmark regression, we restrict our sample to establishments belonging to multi-unit firms. We apply this restriction because establishments belonging to single-unit firms mechanically cannot possibly sell to another establishment in their firm (as no such establishment exists). However, even in our restricted sample, an establishment belonging to a two-establishment firm will only have a positive same-firm ownership fraction in one destination ZIP code, with zeros elsewhere. To see whether most of our observations are drawn from relatively small firms like these or whether the relationship between trade flows and our same-firm ownership fraction varies with firm size (the number of establishments belonging to i^e's firm), we re-estimate the regression from column (2) of Table 2 only using observations from large firms. In columns (2) through (4) of Table 9, we progressively restrict the sample to sending establishments belonging to 5-establishment, 10-establishment, or 20-establishment firms. The estimated coefficients across the first four columns are similar to one another. In column (5), we expand our sample to include establishments in single-unit firms. While these establishments cannot possibly have any within-firm shipments, their inclusion may affect our estimate of the sensitivity of trade flows to distance. Column (5) indicates that our results are unchanged by the inclusion of establishments belonging to single-unit firms.

Second, in constructing the samples in any of our regression specifications, a key step is to define pairs of industries that are upstream/downstream of one another. This step is necessary in order to construct the same-firm ownership fraction s_{zi^e}. Under a definition in which many pairs of industries are classified as vertically linked, the number of downstream establishments for a sending establishment i^e will be relatively large. As a result, the same-firm ownership fraction (which, as a reminder, computes the fraction of downstream establishments in the destination ZIP

Table 10: Relationship between Distance, Common Ownership, and Market Shares: Sensitivity to IO Link Definition and to the Number of Sampled Shipments

Dependent Variable: $\frac{X_{zi^e}}{X_z}$						
	(1)	(2)	(3)	(4)	(5)	(6)
Same-firm ownership fraction	2.828	2.038	1.909	2.586	2.853	3.021
	(0.049)	(0.039)	(0.033)	(0.067)	(0.066)	(0.059)
Log mileage	-0.962	-0.963	-0.963	-0.899	-0.939	-0.899
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.007)
Same-firm ownership fraction×				0.021	-0.054	-0.016
Indicator:≥100 shipments				(0.093)	(0.093)	(0.090)
Log mileage×				-0.048	-0.047	-0.091
Indicator:≥100 shipments				(0.007)	(0.006)	(0.001)
Multilateral Resistance		Unweighted		None	Unweighted	Weighted
Cutoff for IO links (Percent)	1	2	3	1	1	1

Notes: The first column reproduces column (2) of Table 2. Relative to the first column, in columns (2) and (3) we vary the cutoff share of (6-digit NAICS) industry I's revenues that must go to industry J for the I,J industry pair to be defined as vertically linked. The sample contains 190 million i^e–z pairs, representing the shipments of 35,000 establishments.

code that belong to the same firm as i^e) will tend to be relatively large with higher cutoff values.[36] In our benchmark definitions, we choose a 1 percent cutoff — that is, we define 6-digit NAICS industry J to be downstream of 6-digit NAICS industry I if at least 1 percent of industry I's sales are sent to industry J. In the second and third columns of Table 10, we consider increasingly restrictive definitions, with 2 and 3 percent cutoffs. In these two columns, the estimated coefficient on the log mileage term is similar to the estimate of the benchmark specification. The coefficient estimates for the same-firm ownership fraction term are smaller by approximately one-third. However, since the number of downstream establishments (with the more restrictive definition of vertical linkages) is lower, the resulting distance premium in the specifications in columns (2) and (3) are 69 percent and 73 percent, somewhat larger than the 60 percent of the benchmark specification. Based on these columns, we conclude that our benchmark results are robust to increasingly restrictive definitions on the extent to which industries are vertically linked.

As discussed earlier in this paper, the CFS only contains a subset of the shipments that each surveyed establishment made during the survey year. Surveyed establishments only report on a subset of weeks — one week per quarter — within the year and report on only a maximum

[36]In this fraction, both the numerator and the denominator will be smaller with higher cutoffs. However, applying definitions in which few pairs of industries are classified as vertically integrated, the denominator decreases more than the numerator does.

of 40 shipments per quarter.[37] Our third set of exercises examines the robustness of our main results to this sparsity of our dataset. We first split the sample in two: establishments that reported on at least 100 shipments — the median number in our sample — and those establishments that reported on fewer than 100 shipments. We then regress our market share variables against the same-firm ownership fraction and distance variables, with both explanatory variables interacted with an indicator variable equal to 1 if the establishment reported at least 100 shipments. The idea behind this exercise is that establishments with fewer than 100 shipments are more likely to have reported on all of the shipments that they made in the weeks they were surveyed. Columns (4) to (6) of Table 10 contain the results of our exercise. These columns indicate that the relationship between trade flows and ownership does not significantly differ according to the number of shipments per surveyed establishment. The coefficient on distance is larger, by about 5 percent, for establishments that report fewer than 100 shipments. This could reflect some shipment costs that increase with distance, but not at a one-for-one rate with the scale of the shipment. Plants that make costly-to-ship faraway transactions may economize by batching larger values within the same shipment.

Our fourth set of robustness checks explores the sensitivity of our results to the type of destination region. In arriving at our main results, a key ingredient was the number of potential recipients (i.e., number of establishments in industries that are downstream of the sender) in the destination ZIP code. Moreover, within our sample, there is substantial variation in the number of potential recipients. Motivated by this variation, in Table 11 we explore the robustness of our results to restricting the sample based on the number of potential recipients in the destination ZIP code. In the first column, we report our benchmark results. Restricting the sample to ZIP codes with an increasingly greater number of recipient plants has no impact on the estimated coefficient of distance on trade flows, but increases the coefficient estimate of common ownership. By itself, these larger coefficients would imply larger distance premia in subsamples of $i^e - z$ pairs for which there are many downstream establishments in ZIP code z. However, since the $(1 + \text{plants} \in z)^{-1}$ term decreases with our sample restriction, the net effect is to have smaller distance premia of ownership for ZIP codes with a larger number of recipients: Our distance premium of ownership is 45 percent when restricting to ZIP codes with at least five potential recipients and 38 percent when restricting to ZIP codes with at least ten potential recipients. While we find lower distance premia from larger destination ZIP codes, this is to be expected. In larger destination ZIP codes there are, mechanically, likely to be more same-firm establishments in industries downstream of the sender. The distance premium that we report describes the association with an additional *single* same-firm establishment. To have a true like-to-like comparison, it may be necessary to account for the differences across larger destination vs. smaller destination ZIP codes in the number of same-firm establishments.

[37] See https://www.bts.gov/archive/publications/commodity_flow_survey/methodology . Accessed July 20, 2019.

Table 11: Relationship between Distance, Common Ownership, and Market Shares: Sensitivity to Size of Destination Region

Dependent Variable: $\frac{X_{zi^e}}{X_z}$			
	(1)	(2)	(3)
Same-firm ownership fraction	2.828	7.109	9.300
	(0.049)	(0.132)	(0.285)
Log mileage	-0.962	-0.944	-0.946
	(0.003)	(0.004)	(0.004)
In-sample average: $(1+\text{plants} \in z)^{-1}$	0.315	0.079	0.048
Number of downstream establishments in destination	Benchmark	≥ 5	≥ 10

Notes: The first column reproduces column (2) of Table 2. Relative to the first column, we restrict to pairs of sending establishments and destination ZIP codes where there are at least five potential recipients (column 2) or at least 10 potential recipients (column 3). The sample in column (2) contains 97 million i^e–z pairs, representing the shipments of 33,000 sending establishments. The sample in column (3) contains 61 million i^e–z pairs, representing the shipments of 31,000 sending establishments.

Our fifth exercise examines the importance on the assumptions that we make about how the regression errors are clustered. Potentially, there is some unobserved spatial correlation in the extent to which individual establishments ship to individual ZIP codes. Our motivation, then, is to explore whether such spatial correlation may be leading us to overstate the precision with which we estimate the relationship between trade flows, distance, and ownership. Throughout our analysis, we have clustered errors at the level of the sending establishment. In Table 12, we explore the role of different assumptions on clustering on the resulting standard errors with an OLS specification. While it would have been ideal to re-estimate our Poisson regressions with various assumptions on clustering, this would only be feasible via a bootstrapping approach that would have taken an inordinate amount of time given the size of our dataset. In columns (2) through (4), we cluster standard errors at the industry (of the sending establishment) by destination region pair; within these columns, we choose destination regions of varying size. In columns (5) through (7), we cluster standard errors at the level of the destination region. Compared to the level of clustering in the benchmark specification, the standard errors on the same-firm ownership fraction term are lower in columns (2), (3), (5), and (6); within these columns, clustering occurs using relatively small destination regions. The standard error on our common ownership term is larger only when clustering by the destination state or by the destination state $\times$ industry (of the sending establishment) pair. Moreover, the increases in standard errors are modest, on the order of 25 percent (comparing $2.19 \cdot 10^{-3}$ or $2.27 \cdot 10^{-3}$ versus $1.78 \cdot 10^{-3}$). The standard errors on our estimate of the relationship between distance and trade flows are more sensitive to how we cluster.

Table 12: Relationship between Distance, Common Ownership, and Market Shares: Clustering

Dependent Variable: $\frac{X_{zi^e}}{X_z}$				
	(1)	(2)	(3)	(4)
Same-firm ownership fraction	0.0553	0.0553	0.0553	0.0553
	(0.001773)	(0.000941)	(0.001111)	(0.002194)
Log mileage	-0.0113	-0.0113	-0.0113	-0.0113
	(0.000084)	(0.000019)	(0.000108)	(0.000321)
Clustering	Sending Estab.	Industry$\times$ Dest. ZIP Code	Industry$\times$ Dest. County	Industry$\times$ Dest. State
Dependent Variable: $\frac{X_{zi^e}}{X_z}$				
		(5)	(6)	(7)
Same-firm ownership fraction		0.0553	0.0553	0.0553
		(0.000981)	(0.001219)	(0.002274)
Log mileage		-0.0113	-0.0113	-0.0113
		(0.000037)	(0.000317)	(0.000858)
Clustering		Dest. ZIP Code	Dest. County	Dest. State

Notes: The first column estimates columns (2) of Table 2 with an OLS rather than a Poisson specification. Across different columns, we apply different assumptions on the level at which observations are correlated. Throughout all columns, the sample contains 190 million i^e–z pairs, representing the shipments of 35,000 establishments. In all specifications, we calculate the unweighted multilateral resistance terms.

However, in our benchmark estimations, this relationship was much more precisely estimated. In sum, subject to the caveat that our benchmark specification applied Poisson regressions as opposed to the OLS regressions that we estimate here, we conclude from Table 12 that our benchmark regressions are adequately precisely estimated under a variety of assumptions about the errors' spatial correlation.

Our sixth exercise assesses whether our regression results are sensitive to the weighting of observations. Weighting may be salient in one of two ways. First, in constructing the Commodity Flow Survey, the U.S. Census Bureau over-samples larger establishments. In columns (2) and (5) of Table 13, we apply this CFS sampling weight. Weighting observations by the (inverse) sampling probability leads to both a weaker estimated relationship between trade flows and common ownership and between trade flows and distance, with no substantial net effect on the distance premium. Second, in computing our same-firm ownership fraction, we have so far computed the fraction of downstream establishments in the destination ZIP code that are commonly owned with the sending establishment. In columns (3) and (6), we instead weight destination ZIP code plants by their

Table 13: Relationship between Distance, Common Ownership, and Market Shares: Sensitivity to Weighting Methods

Dependent Variable: $\frac{X_{zi^e}}{X_z}$						
	(1)	(2)	(3)	(4)	(5)	(6)
Same-firm ownership fraction	2.828	2.505	2.299	2.941	2.594	2.428
	(0.049)	(0.000)	(0.039)	(0.047)	(0.000)	(0.038)
Log mileage	-0.962	-0.827	-0.961	-0.944	-0.831	-0.942
	(0.003)	(0.000)	(0.003)	(0.005)	(0.000)	(0.005)
Multilateral Resistance		Unweighted			Weighted	
Use CFS Sample Weights	No	Yes	No	No	Yes	No
Use payroll to weight in the same-firm ownership fraction	No	No	Yes	No	No	Yes

Notes: The first column reproduces column (2) of Table 2. The fourth column reproduces column (3) of Table 2. Throughout all columns, the sample contains 190 million i^e–z pairs, representing the shipments of 35,000 establishments.

payroll.[38] Here, the coefficient estimate on the same-firm ownership term is smaller than in our benchmark regressions by about 20 percent or 5 percent, depending on whether one applies the unweighted or weighted multilateral resistance terms. In sum, different weighting methods yield modestly different coefficient estimates.

Seventh, we re-estimate our benchmark specification, cutting the sample based on industries' export intensity. We do so because our dataset of sending establishments and destination ZIP codes necessarily excludes exports and imports. As a result, our estimate of the relationship between trade flows and common ownership may be mis-measured, particularly so for industries in which exports and imports are prevalent. According to columns (1) through (3) of Table 14, the relationship between the same-firm ownership fraction and market shares is similar for industries with high export shares (above 10 percent) and industries with low export shares (below 10 percent). Not surprisingly, the relationship between distance and trade flows is somewhat stronger for high-export industries, likely reflecting the greater tradability of these commodities.

Throughout our empirical analysis, we have excluded i^e–z pairs for which the sending establishment resides in ZIP code z. Our primary rationale for excluding these observations is that the logarithm of the mileage between i^e and z is undefined for these pairs. However, given that both market shares and the same-firm ownership fraction tend to be substantially higher than average for these types of observations, it is necessary to explore the robustness of our results to their ex-

[38] Our Online Appendix A model indicates that the same-firm ownership fraction should be computed by taking establishments' input expenditures as weights. Unfortunately, these expenditure measures are available only for a subset of industries. Since payroll may be sufficiently dissimilar from downstream establishments' expenditures, our benchmark regressions are based on unweighted averages of the same-firm ownership fraction.

Table 14: Relationship between Distance, Common Ownership, and Market Shares: Additional Robustness Checks

Dep. Variable: $\frac{X_{zi^e}}{X_z}$					
	(1)	(2)	(3)	(4)	(5)
Same-firm ownership	2.828	2.923	2.850	2.795	2.591
ownership fraction	(0.049)	(0.071)	(0.053)	(0.048)	(0.091)
Log mileage	-0.962	-0.601	-1.045	-0.962	-0.999
	(0.003)	(0.007)	(0.004)	(0.003)	(0.004)
Sample	Benchmark	High Export	Low Export	Same-ZIP Code Pairs	Sending-Estab. × Destination County

Notes: The first column reproduces column (2) of Table 2. In columns (2) and (3), we split the sample according to the export intensity of the industry of the sending establishment. The cutoff export share is 10 percent. In column (4), our sample adds i^e–z pairs in which the sending establishment resides in the destination ZIP code. In addition to the variables listed, we include as an explanatory variable an indicator, equal to 1 if i^e resides in z. Columns (2), (3), (4), and (5) contain 19 million, 171 million, 192 million, and 51 million i^e–z pairs, corresponding to the shipments made by 8,000, 27,000, 35,000, and 35,000 establishments. In all specifications, we calculate the unweighted multilateral resistance terms.

clusion from our sample. This is our eighth check. Column (4) of Table 14 demonstrates that the addition of i^e–z pairs in which the sending establishment resides in the destination ZIP code does not alter our results.[39]

Ninth, one may be concerned that establishments' shipment patterns may be spatially correlated, especially (if shippers batch long-haul shipments) among ZIP codes in faraway destinations. We explored different assumptions on the spatial correlation of errors in Table 12. As a second approach to check for the impact of spatial correlation, in column (5) of Table 14 we re-estimate our benchmark regression with counties representing destination regions. Relative to results in column (1), the coefficient on common ownership is less than 10 percent smaller. Consistent with the idea that errors are spatially correlated across ZIP codes within countries, the standard error is nearly twice as large in column (5) as it is in column (1). Nevertheless, the standard errors in column (5) are small in absolute terms. So, Table 2's main finding, namely that the coefficient on the same-firm ownership fraction is 2.5 to 3 times larger than the coefficient on distance, endures with our alternative definition for the destination region.

[39]To compute log(mileage) for these observations, we take the minimum distance over the set of observations in our baseline sample. In our Table 14 column (4) regression, we also include an indicator variable, equal to 1 for the observations for which the sending establishment resides in the destination ZIP code. The inclusion of this indicator variable implies that the coefficient estimates on the same-firm ownership fraction or the log mileage term are unaffected by our choice of imputed value for log mileage for observations for which $i^e \in z$. Since the coefficient estimate corresponding to this indicator variable is wholly dependent on the distance we assign to "within ZIP code" observations, Table 14 omits the coefficient estimate on the within-ZIP code indicator.

Table 15: Relationship between Distance, Common Ownership, and Market Shares: Control Function Estimates with Distance-Ownership Interaction Terms

Dependent Variable: $\frac{X_{zi^e}}{X_z}$			
	(1)	(2)	(3)
Same-firm ownership	2.328	2.133	2.253
ownership fraction	(0.283)	(0.401)	(0.569)
Log mileage	-0.964	-0.964	-0.965
	(0.003)	(0.003)	(0.003)
Interaction between log mileage	0.586	0.352	1.046
and same-firm ownership fraction	(0.307)	(0.375)	(0.527)
Error from the first stage:	1.122	1.312	1.188
Same-firm ownership fraction	(0.285)	(0.403)	(0.570)
Error from the first stage: interaction between	-0.289	-0.056	-0.751
log mileage and same-firm ownership fraction	(0.309)	(0.376)	(0.527)
Number of segments	1	2	3

Notes: All regressions include sending-establishment fixed effects. The first-stage regressions include log mileage, the incidental merger fraction, and the interaction between the two as covariates. The sample includes 190 million i^e–z pairs, drawing on the shipments made by 35,000 establishments. In the final row, "Number of segments" refers to the S we used when identifying which establishments were part of an incidental merger. In all specifications, we calculate the unweighted multilateral resistance terms.

In Table 3, we estimated a positive relationship between trade and the interaction of distance and common ownership. In our tenth set of exercises, we apply our two-stage control function approach to a specification with the distance-ownership interaction terms. Since this specification now includes a second endogenous variable, we require a second instrument. In addition to the incidental merger fraction — our instrument in the Table 7 first-stage regression — we include the interaction of the incidental merger fraction and log mileage as an explanatory variable in our first-stage regressions. In Table 15, we find that the effect of ownership on trade flows is larger for faraway destinations; the coefficient estimate is larger than in Table 3 but with substantially larger standard errors.

Finally, as an alternative to the two-stage control function approach, Wooldridge (1997) and Windmeijer (2000) derive the moment conditions for cases with a linear first stage and a fixed effect Poisson second stage. We apply these moment conditions and re-estimate the relationships between trade flows, distance, and common ownership. The estimates are given in Table 16, with each column applying a different definition of incidental merger establishments. The coefficients on the same-firm ownership fraction are now larger than the benchmark Poisson regression estimates, though with substantially larger standard errors. Because of the larger uncertainty surrounding the GMM estimates, we take the coefficient estimates from our two-stage control function approach to

Table 16: Relationship between Distance, Common Ownership, and Market Shares: GMM Estimates

Dependent Variable: $\frac{X_{zi^e}}{X_z}$	GMM Estimates			Baseline
	(1)	(2)	(3)	(4)
Same-firm ownership	4.660	4.051	4.095	2.828
fraction	(0.942)	(1.429)	(2.039)	(0.049)
Log mileage	-0.972	-0.972	-0.972	-0.962
	(0.005)	(0.005)	(0.005)	(0.003)
Number of segments	1	2	3	–

Notes: All regressions include sending-establishment fixed effects. The first-stage regressions also include log mileage as a covariate. The sample includes 190 million i^e–z pairs, drawing on the shipments made by 35,000 establishments. In the final row, "Number of segments" refers to the S we used when identifying which establishments were part of an incidental merger. In all specifications, we calculate the unweighted multilateral resistance terms. The last column reports our baseline results (column 2 from Table 2) without attempting to address potential endogeneity in the same-firm ownership fraction variable.

be our headline results.

D Control Function and GMM Approaches

Here, we explore the two-stage control function and GMM approaches used in Section IV.C and Online Appendix C. In particular, we specify our GMM moment conditions and perform a Monte Carlo exercise to assess the performance of our control function and GMM estimators. For this appendix only, let π_{zi^e} be our dependent variable; d_{zi} an explanatory variable; s_{zi^e} an endogenous explanatory variable; i^e the index of a sending establishment; and z the index of a destination ZIP code. There are a large number of sending establishments but a fixed set of locations Z.

Consider the following data generating process, a fixed effect Poisson model with an endogenous regressor:

$$\pi_{zi^e} \sim Poisson(\exp(s_{zi^e}\beta + d_{zi}\gamma + v_{i^e} + \varepsilon_{zi^e}))$$
$$s_{zi^e} = d_{zi}\alpha + x_{zi^e}\sigma + \eta_{i^e} + \xi_{zi^e}$$
$$\varepsilon_{zi^e} = \xi_{zi^e}\rho + \phi_{zi^e}.$$

In the final equation, ϕ_{zi^e} is independent of ξ_{zi^e}. We also assume that ε_{zi^e} is uncorrelated with $\varepsilon_{z'i^e}$ for $z \neq z'$ and that $\mathbb{E}[\phi_{zi^e}] = \mathbb{E}[\varepsilon_{zi^e}] = 0$. Finally, let x_{zi^e} denote our instrument for s_{zi^e}. With endogeneity, $Cov(s_{zi^e}, \varepsilon_{zi^e}) \neq 0$, but $Cov(x_{zi^e}, \varepsilon_{zi^e}) = 0$.

Our GMM estimator is from Wooldridge (1997) and Windmeijer (2000). Our moment condi-

Table 17: Monte Carlo Results

	Mean (1)	S.D. (2)	Mean (3)	S.D. (4)	Mean (5)	S.D. (6)	Mean (7)	S.D. (8)
Panel A: Poisson Regression, No Instruments								
β	0.050	0.019	0.050	0.012	0.050	0.015	0.050	0.012
γ	0.030	0.041	0.028	0.028	0.029	0.041	0.028	0.027
Panel B: Control Function Estimation								
β	0.010	0.021	0.010	0.014	0.010	0.017	0.010	0.013
γ	0.042	0.040	0.040	0.028	0.041	0.041	0.040	0.027
First Stage								
σ	2.000	0.003	2.000	0.002	2.000	0.002	2.000	0.002
α	0.300	0.003	0.300	0.002	0.300	0.002	0.300	0.002
Panel C: GMM Estimation								
β	0.011	0.035	0.010	0.019	0.011	0.029	0.011	0.024
γ	0.041	0.041	0.040	0.028	0.041	0.040	0.039	0.027
Sending Establishments	500		500		1000		1000	
Destination ZIP Codes	200		400		200		400	

Notes: The true values for these simulations are $\beta = 0.01$, $\gamma = 0.04$, $\alpha = 0.3$, $\sigma = 2$, and $\rho = 0.2$. The odd-numbered columns give the average parameter estimate from our 1,000 simulations. The even-numbered columns give the standard deviation across simulations.

tion is:

$$\mathbb{E}\left[x_{zi^e}\left(\frac{\pi_{zi^e}}{\exp(s_{zi^e}\beta+d_{zi}\gamma)}-\frac{\frac{1}{Z}\sum_{z'=1}^{Z}\pi_{z'i^e}}{\frac{1}{Z}\sum_{z'=1}^{Z}\exp(s_{z'i^e}\beta+d_{z'i^e}\gamma)}\right)\right]=0\ . \tag{11}$$

With the goal of examining the performance of the control function and GMM estimators that we use in Section IV.C and Online Appendix C, we perform a series of Monte Carlo simulations. In these simulations, we use the following parameter values: $\beta = 0.01$, $\gamma = 0.04$, $\alpha = 0.3$, $\sigma = 2$, and $\rho = 0.2$. With these parameter values, we simulate data on either 500 or 1,000 sending establishments and $Z \in \{200, 500\}$ destinations.

Monte Carlo results for 1,000 simulations are reported in Table 17. In Panel A, we report the estimation results from a fixed effect Poisson model, without addressing endogeneity. In this panel, our coefficient estimate on the relationship between s_{zi^e} and π_{zi^e} is biased upward. Panel B uses our two-step control function approach. In the first stage, we use an ordinary least squares regression, with fixed effects, to regress s_{zi^e} on both d_{zi} and the instrument x_{zi^e}. We then predict $\hat{s_{zi^e}}$ and obtain a residual $\hat{\xi_{zi^e}}$. Adding this residual as a control in the second-stage fixed effect Poisson model estimation, we are able to recover the true parameter values reasonably well. Similarly, Panel C indicates that our GMM estimator, based on Equation 11, allows us to recover the correct parameter values. In all three panels, the across-simulation standard deviations of the parameter are inversely related to the number of destination ZIP codes.

E Aggregate Effects

In this appendix, we apply our estimates on the prevalence of intra-firm shipments and the relationships among shipment intensity, common ownership, and distance to quantify the aggregate importance of common ownership. To perform these counterfactual exercises, we employ the models of Caliendo and Parro (2015) and Caliendo et al. (2018). An extended and aggregated version of the model we have laid out in Section II, these models incorporate input-output linkages across sectors, multiple primary inputs, and (in the case of Caliendo et al., 2018) labor mobility across regions.

To summarize the Caliendo et al. (2018) model, each region has an initial stock of land and structures. In Caliendo et al. (2018), each region is one of the 50 U.S. states. In our analysis here, closer to the geographic definition used in the earlier parts of this paper, an individual region represents either a single MSA (Metropolitan Statistical Area) or a state's non-metropolitan portion.[40] Consumers within each region work and consume a bundle of consumption goods produced by different industries. Their preferences are described by a Cobb-Douglas utility function over the goods and services consumed of each industry's commodity. Within each region-industry pair, a continuum of intermediate input producers combine (via a Cobb-Douglas production function) land and structures, labor, and material inputs to create the output. Establishments compete as a function of their own idiosyncratic productivity and the average productivity of other potential suppliers to the final good producer within each destination market; the intermediate-good-supplying establishment that is able to deliver the good at the lowest price serves the destination. This aspect of the model corresponds to the partial equilibrium model discussed in Section II. Also, within each industry and region, final good producers make a region-industry-specific bundle by combining the goods that they have purchased from intermediate input suppliers.

Below, we delineate the maximization problems faced by each region's representative consumer, each region-industry's intermediate good producing firms, and each region-industry's final good producing firms. We then present the market-clearing conditions and define the competitive equilibrium. Much of the material below can be found, in much greater detail, in Caliendo and Parro (2015) and Caliendo et al. (2018).[41]

[40] There are two reasons why we apply a geographic classification based on MSAs rather than ZIP codes. First, some of the required regional data on employees' compensation or total gross output do not exist at the finer level. Second, in computing the counterfactual equilibrium, we must repeatedly solve a system of (linear) equations of dimension equal to the $Z \cdot J$, the number of regions multiplied by the number of industries. This would be computationally challenging, to say the least, with the finer ZIP-code-based geographic classification.

[41] There is one meaningful way in which the Caliendo et al. model — and, consequently, the model used in this section — does not nest the Eaton, Kortum, and Sotelo (2012)-based model introduced in Section II: In this section,

Each region is home to a representative consumer, who inelastically supplies labor and has Cobb-Douglas preferences over the goods produced by each industry:

$$U_i = \prod_{j=1}^{J} (c_i^j)^{\xi^j} \text{ where } \sum_{j=1}^{J} \xi^j = 1.$$

These preference parameters are identical across regions. Using P_i^j to refer to the price of final good j in region i, and $I_i = \frac{r_i H_i + w_i L_i}{L_i}$ as the per capita income of households in region i, the indirect utility of households in region i equals

$$U_i = \frac{I_i}{P_i}; \text{ and where } P_i \equiv \prod_{j=1}^{J} \left(\frac{P_i^j}{\xi^j} \right)^{\xi^j}$$

equals the ideal price index in region i.

Within each region and industry, a continuum of intermediate-good-producing establishments produce using a combination of materials, structures and land, and labor. Individual establishments have idiosyncratic productivity levels, v_i^j, with the levels drawn from a Fréchet distribution with parameter θ^j. The production function for the set of establishments in region i and industry j with productivity draw v_i^j is given by

$$q_i^j(v_i^j) = v_i^j \cdot \left[T_i^j \cdot h_i^j(v_i^j)^{\beta_i} \cdot l_i^j(v_i^j)^{1-\beta_i} \right]^{\gamma^j} \cdot \prod_{k=1}^{J} \left[M_i^{jk}(v_i^j) \right]^{\gamma^{jk}}.$$

In this equation, the input choices $h_i^j(\cdot)$, $l_i^j(\cdot)$, and $M_i^{jk}(\cdot)$ of establishments in region i are func-

we revert to the more conventional representation of establishments as points on a continuum. As a result, when computing counterfactual responses to changes in trade costs, the entire response will occur through the intensive margin: A decline in trade costs will not result in pairs of regions going from zero to positive trade flows. For the goal of this section — computing the welfare effects of counterfactual changes in trade costs — the representation of firms as points on a continuum is a reasonable approximation.

In one of their counterfactual exercises, using a single-sector model, Eaton, Kortum, and Sotelo examine the change in international trade flows that would result from a uniform 10 percent reduction in cross-border trade costs. They report, "World exports rise by 43 percent due to lower trade costs, in line with results in Eaton, Kortum, and Kramarz (2011)... nearly all of this increased trade occurs within pairs of countries that were already trading, 99.9984 percent" (p. 365). On the other hand, when examining trade across MSAs (instead of countries) separately by industry (instead of aggregating across industries), the extensive margin will likely play a larger role than in Eaton, Kortum, and Sotelo's experiment.

In addition, one can rationalize the difference in formulations — a continuum of establishments in this section as opposed to a countable number in Section II — as in Gaubert and Itskhoki (2016). Gaubert and Itskhoki propose a model in which each industry has a small number of firms (since they are interested in the extent to which individual firms can explain countries' comparative advantage), but with a continuum of industries. In this section, in line with Caliendo and Parro (2015) and Caliendo et al. (2018), we apply a coarser industry definition compared to what we use in Section II. So, one may think of the sectors in this section as a collection of more finely defined industries that formed the basis of our Section II model.

tions of their idiosyncratic productivity levels. Each establishment in region i rents structures at (constant) unit price r_i, hires labor at constant unit price w_i, and purchases material inputs at constant unit prices P_i^k (for $k \in 1,2,...,J$). Assuming production functions exhibit constant returns to scale (so that $\gamma^j + \sum_k \gamma^{jk}$=1), an establishment with idiosyncratic productivity equal to v_i^j produces at constant marginal cost

$$\frac{x_i^j}{v_i^j \left(T_i^j\right)^{\gamma^j}}; \text{ where } x_i^j \equiv \left[\left(\frac{r_i}{\beta_i \gamma^j}\right)^{\beta_i} \cdot \left(\frac{w_i}{(1-\beta_i)\gamma^j}\right)^{1-\beta_i}\right]^{\gamma^j} \cdot \prod_{k=1}^{J} \left[\frac{P_i^k}{\gamma^{jk}}\right]^{\gamma^{jk}}. \tag{12}$$

For each region and industry, there is a perfectly competitive industry of final good producers, who combine the output of intermediate input producers purchased from the continua of establishments from different supplying regions, according to the following production function:

$$Q_i^j = \left[\int_{\mathbb{R}_+^Z} \left[\tilde{q}_i^j(v_i^j)\right]^{\frac{\varsigma_i^j - 1}{\varsigma_i^j}} \phi^j(v^j) dv^j\right]^{\frac{\varsigma_i^j}{\varsigma_i^j - 1}}.$$

Here, $\tilde{q}_i^j(v_i^j)$ equals the intermediate goods purchased from producers that have idiosyncratic productivity v_i^j, $\phi^j(v^j)$ denotes the joint density function of the idiosyncratic productivity levels of the producers from the Z possible origin regions, and ς_i^jequals the elasticity of substitution across intermediate good varieties. The purpose of introducing these final good producers is to cleanly characterize the price of an industry's output in each region. This price equals the final good producers' marginal cost:

$$P_i^j = \left[\int_{\mathbb{R}_+^Z} \left[p_i^j(v_i^j)\right]^{1-\varsigma_i^j} \phi(v_i^j) dv^j\right]^{\frac{1}{1-\varsigma_i^j}}. \tag{13}$$

As in Section II, each final good producer purchases from the intermediate good supplier that is able to supply the good at the lowest price. Because competition across intermediate good suppliers is perfectly competitive, the price paid by the intermediate good user equals the supplier's marginal cost multiplied by the cost of transporting the good from the supplier to the destination:

$$p_i^j(v^j) = \min_{i \in \{1,...Z\}} \left\{\frac{\omega_i^j \tau_{zi}^j}{v_i^j \left(T_i^j\right)^{\gamma^j}}\right\}.$$

The transportation cost τ_{zi}^j potentially varies by industry and reflects both the distance from i

to z and the share of good-j producing establishments in i that share ownership with downstream plants in destination z. In the case of service industries, we set $\tau_{zi}^{j}=\infty$.[42]

Caliendo et al. show that, if the idiosyncratic productivity is drawn from a Fréchet distribution, then Equation 13 is equivalent to

$$P_i^j=\left[\Gamma\left(\frac{\theta^j+1-\varsigma_i^j}{\theta^j}\right)\right]^{1-\varsigma_i^j}\cdot\left[\sum_{i=1}^{Z}\left[x_i^j\tau_{zi}^j\right]^{-\theta^j}\left(T_i^j\right)^{\theta^j\gamma^j}\right]^{-1/\theta^j}, \tag{14}$$

where $\Gamma(\cdot)$ is the Gamma function.

To complete the description of this model, the market-clearing conditions for labor, structures and land, and final goods are given by Equations 15-17, below:

$$L=\sum_{i=1}^{Z}\sum_{j=1}^{J}L_i^j=\sum_{i=1}^{Z}\sum_{j=1}^{J}\int_{\mathbb{R}_+}l_i^j(v)\,\phi_i^j(v)dv\;; \tag{15}$$

$$H_i=\sum_{j=1}^{J}H_i^j=\sum_{j=1}^{J}\int_{\mathbb{R}_+}h_i^j(v)\,\phi_i^j(v)dv\text{ for }i\in 1,2,...Z\text{ ; and} \tag{16}$$

$$Q_i^j=L_i\cdot c_i^j+\sum_{k=1}^{J}M_i^{jk}=L_i\cdot c_i^j+\sum_{k=1}^{J}\int_{\mathbb{R}_+}M_i^{jk}(v)\,\phi_i^k(v)dv\text{ for }i\in 1,2,...Z\;. \tag{17}$$

Use X_z^j to denote total expenditures on commodity j in region z. In equilibrium, the aggregate trade balance for each region, z is given by:[43]

$$\sum_{i=1}^{Z}\sum_{j=1}^{J}\pi_{zi}^jX_z^j=\sum_{i=1}^{Z}\sum_{j=1}^{J}\pi_{iz}^jX_i^j\text{ for }i\in 1,2,...Z\;. \tag{18}$$

[42] In reality, even though services tend to be relatively more difficult to transmit across space than goods, certain services are to some extent tradable. However, CFS data on goods-producing and goods-distributing establishments are uninformative about how trade barriers relate to distance and ownership for service-industry establishments. Thus, our model will only be able to explore the counterfactual trade and welfare effects of reduction to trade barriers in the goods-related industries.

[43] A simplification we make here is to impose balanced trade across regions. As Caliendo et al. (2018) document, in reality, within the United States trade imbalance is prevalent. Certain states — such as Indiana and Wisconsin — run substantial trade surpluses, while others — including Florida and Georgia — have large trade deficits. To rationalize these trade imbalances, Caliendo et al. (2018) assume that, while some fraction of a state's land and structures are owned locally, the remainder are owned nationally. States with a deficit are able to finance their consumption because they own a relatively large share of the national portfolio of structures. To match the trade imbalances, then, Caliendo et al. define state total income (equal to total final consumption expenditures) to be equal to the sum of the state's trade imbalances (as recorded in the CFS) and the state's value added (as recorded by the BEA). With our finer definition of areas, this procedure unfortunately results in negative income for certain MSAs (principally those that send large volumes of refined petroleum to other areas, such as Lake Charles, Louisiana). So, instead, we assume that all structures and land are owned locally and, correspondingly and counterfactually, that trade across regions is balanced.

One of the key differences between Caliendo and Parro (2015) and Caliendo et al. (2018) — the two papers upon which we build — relates to the treatment of primary inputs. In Caliendo et al. (2018), consumers are allowed to costlessly migrate across regions. As a result, utility is equalized across regions: $U_i = \frac{I_i}{P_i} = U$ for all i. In contrast, in Caliendo and Parro (2015) labor is completely immobile. There is some initial exogenously given allocation of labor across regions, which does not respond to changes in trade costs or technology. Below, we will apply these two alternate, diametrically opposed specifications for our counterfactual exercises.

Having specified the consumers' and producers' maximization problems and the market-clearing conditions, we now define a competitive equilibrium. This definition is taken almost directly from Caliendo et al. (2018): Given factor supplies, L and H_i, a *competitive equilibrium* for this economy is given by a set of factor prices in each region $\{r_i, w_i\}$; a set of labor allocations, structure and land allocations, final good expenditures, consumption of final goods per person, and final goods prices $\left\{L_i^j, H_i^j, X_i^j, c_i^j, P_i^j\right\}$ for each industry and region; a set of pairwise sectoral material use in every region M_i^{jk}; and pairwise regional intermediate expenditure shares in every sector, π_{zi}^j, such that (a) the optimization conditions for consumers and intermediate and final goods producers hold and all markets clear (Equation 15-17); (b) aggregate trade is balanced (Equation 18); and (c) utility is equalized across regions. Condition (c) is omitted in the specification with immobile labor.

Next, we outline the algorithm presented in Caliendo and Parro (2015) and Caliendo et al. (2018) to compute the change in equilibrium trade flows and aggregate welfare in response to a change in trade costs. As in those earlier papers, we will use Y' to refer to the counterfactual value of an arbitrary variable Y and $\widehat{Y} = \frac{Y'}{Y}$ to refer to the change in variable Y.

Step 1: Guess an initial vector of costs for the primary input (labor and land/structures) bundle: Call $\omega_i = \left(\frac{r_i}{\beta_i}\right)^{\beta_i}\left(\frac{w_i}{1-\beta_i}\right)^{1-\beta_i}$ the primary input unit price and $\hat{\omega} = (\hat{\omega}_1, \hat{\omega}_2, ..., \hat{\omega}_Z)$ the vector of changes in the primary input prices.

Step 2: Given this guess for the primary input bundles' cost changes, compute the changes in the costs of each industry-region's input cost bundles, and the final good prices in each industry-region using Equations 12 and 14:

$$\hat{x}_i^j = \left(\hat{\omega}_i^j\right)^{\gamma^k} \prod_{k=1}^{J} \left[\hat{P}_i^k\right]^{\gamma^{jk}}$$

$$\hat{P}_i^j = \left[\sum_{i=1}^{Z} \pi_{zi}^j \left[\hat{x}_i^j \hat{\tau}_{zi}^j\right]^{-\theta^j}\right]^{-1/\theta^j}.$$

Step 3: Given changes in the costs of industry-regions' input cost bundles and prices for industry-regions' final good, compute the changes in the trade shares.

The changes in trade shares are given by

$$\hat{\pi}_{zi}^{j} = \left(\frac{\hat{x}_i^j}{\hat{P}_z^j} \hat{\tau}_{zi}^j \right)^{-\theta^j}.$$

Step 4: Labor mobility condition:

In the specification with immobile labor, $\hat{L}_i = 1$ for all regions i. If, instead, we follow the Caliendo et al. (2018) algorithm, changes in the labor force of each region are given by:

$$\hat{L}_i = \frac{\left(\frac{\hat{\omega}_i}{\hat{P}_i \hat{U}}\right)^{1/\beta_i}}{\sum_{z=1}^{Z} L_z \left(\frac{\hat{\omega}_z}{\hat{P}_z \hat{U}}\right)^{1/\beta_z}} L, \text{ where}$$

$$\hat{U} = \sum_{z=1}^{Z} \frac{L_z}{L} \left(\frac{\hat{\omega}_z}{\hat{P}_z}\right) \left(\hat{L}_z\right)^{1-\beta_z}, \text{ and}$$

$$\hat{P}_z = \prod_{j=1}^{J} \left(\hat{P}_z^j\right)^{\xi_j}.$$

Step 5: Regional market clearing in final goods:

$$\left(X'\right)_z^j = \alpha^j \hat{\omega}_z \left(\hat{L}_z\right)^{1-\beta_z} I_z L_z + \sum_{k=1}^{J} \gamma^{kj} \sum_{i=1}^{Z} \left(\pi'\right)_{iz}^k \left(X'\right)_i^k.$$

Step 6 Trade balance (used in Caliendo and Parro, 2015):

$$\sum_{i=1}^{Z} \sum_{j=1}^{J} \left(\pi'\right)_{zi}^j \left(X'\right)_z^j = \sum_{i=1}^{Z} \sum_{j=1}^{J} \left(\pi'\right)_{iz}^j \left(X'\right)_i^j. \tag{19}$$

Step 6′: Labor market clearing (used in Caliendo et al., 2018):

$$\hat{\omega}_z \left(\hat{L}_z\right)^{1-\beta_z} \left(I_z L_z\right) = \sum_{j=1}^{J} \gamma^j \sum_{i=1}^{Z} \left(\pi'\right)_{iz}^j \left(X'\right)_i^j. \tag{20}$$

This equation states that shipments of commodity j can either be consumed (the first summand on the right-hand side) or used as a material input (the second summand).[44]

[44] Regarding the first summand, note that $\hat{\omega}_z \left(\hat{L}_z\right)^{1-\beta_z} I_z L_z$ equals $\hat{\omega}_z \left(\hat{L}_z\right)^{-\beta_z} I_z L_z'$. Also note that intermediate good producers' cost-minimizing choices of land/structures and labor implies that $\hat{I}_z = \hat{\omega}_z \left(\frac{\hat{H}_z}{\hat{L}_z}\right)^{\beta_z}$. Since the stock of land/structures is fixed within each region, $\hat{\omega}_z \left(\hat{L}_z\right)^{1-\beta_z} I_z L_z$ equals $I_z' L_z'$.

To update our initial guess of costs for the primary input bundle, we need one additional market-clearing condition. Caliendo and Parro (2015) and Caliendo et al. (2018) use different market-clearing conditions.

Since the trade shares (the πs), changes in each region's labor force (the Ls), and the shipments of different commodities from different regions (the Xs) are each functions of the $\hat{\omega}$ vector, failure of Equation 19 or Equation 20 implies that our guess of $\hat{\omega}$ needs to be updated.

The algorithm follows steps 2-6 until Equation 19 holds (when working through the case with immobile labor) or Equation 20 holds (when working through the case with mobile labor).

We next describe the model's calibration. Beyond the aforementioned data on same-firm ownership shares, distance measures, and shipment rates, this exercise requires data parameterizing consumers' preferences for different final consumption goods, industries' production functions, regions' initial labor and capital endowments, and the dispersion in establishments' fundamental productivity. For these parameters we follow, as much as possible, the calibration procedure outlined in Caliendo et al. (2018). We adopt an industry classification scheme with 19 goods-related and 13 service industries.[45] For this set of industry definitions and for our more coarsely defined regions, we re-compute trade flows and same-firm ownership shares from the 2007 Commodity Flow Survey. Data from the 2007 BEA Input-Output Table identify parameters related to sectoral production functions and the representative consumer's final preferences: We set γ^{jk} — the Cobb-Douglas share parameter that describes the importance of industry k's commodity as an input for production in sector j — equal to the share of industry j's expenditures that are spent on purchases of commodity k, and we let γ^{j} (the share of capital and labor in production) equal the residual share of industry j's expenditures. The preference parameter for industry j's output ξ^{j} is proportional to the industry's final consumption expenditures. The initial labor endowment L_i equals MSA i's total employment as a share of aggregate employment. (These employment figures are taken from the BEA Regional Accounts. The total labor endowment L is normalized to 1.) We compute the share of land and structures in value added for MSA i, β_i, following the procedure of Caliendo et al. (2018).[46] Our estimates of θ^{j}, which parameterize the dispersion of establishments' idiosyncratic

[45] The industries that produce or distribute goods are Mining; Food, Beverages, and Tobacco; Textiles, Apparel, and Leather; Paper Products; Printing; Petroleum and Coal Products; Chemical Products; Rubber and Plastic Products; Wood Products; Nonmetallic Mineral Products; Primary Metals; Fabricated Metal Products; Machinery; Computer and Electronic Products; Electrical Equipment; Transportation Equipment; Furniture; Miscellaneous Manufacturing; and Wholesaling. The service industries are Farms, Forestry, and Fishing; Mining Services; Utilities; Construction; Retail; Transportation Services; Publishing and Information; Finance, Insurance, and Real Estate; Professional and Business Services; Health and Education; Arts, Amusement, Accommodation, and Food Services; Other Personal Services; and Government. Caliendo et al. (2018) refer to the first set of industries as "tradable" industries and the latter set of industries as "non-tradable." While services tend to be less tradable than goods, there are certain exceptions, like Finance, Insurance, and Real Estate.

[46] That is, we begin by computing $1-\tilde{\beta}_i$ as the share of total compensation in MSA i that is paid to labor. Since the non-labor compensation equals not only payments to land and structures, but also equipment rentals, we calculate the share of land and structures as $\beta_i = \frac{\tilde{\beta}_i - 0.17}{0.83}$, where the value 0.17 reflects payments to equipment.

productivity, are taken from Caliendo and Parro (2015).[47]

For the initial and counterfactual trade costs, τ_{zi}^{j} and $\tilde{\tau}_{zi}^{j}$ respectively, we set

$$\tau_{zi}^{j} = \frac{\alpha_1}{\theta^j} \cdot \log \text{mileage}_{z \leftarrow i} + \frac{\alpha_2}{\theta^j} s_{zi}^{j} \text{ , and}$$
$$\tilde{\tau}_{zi}^{j} = \frac{\alpha_1}{\theta^j} \cdot \log \text{mileage}_{z \leftarrow i} + \kappa \frac{\alpha_2}{\theta^j} s_{zi}^{j} \text{ ,}$$

where $\alpha_1 = 0.95$ and $\alpha_2 = -1.80$ equal the values given in the second column of Table 7.

Table 18 presents the results from our counterfactual exercises for $\kappa \in \{0,1,2,3,4,5\}$. These exercises correspond to the elimination of common ownership ($\kappa = 0$), the status quo ($\kappa = 1$), or a 2-, 3-, 4-, or 5-fold increase in the share of same-firm establishments in destination MSAs.

An increase in trade costs due to the elimination of common ownership, the $\kappa = 0$ case, leads to a modest 0.2 percent decrease in real wages and a 0.1 percent drop in gross output. Given the small same-firm ownership fraction present in the data (a reduction from 0.05 percent to 0), and given that we are assessing only the effects of changes in ownership in the goods-producing and goods-distributing industries, these aggregate effects are nontrivial. There are two reasons behind this substantial multiplier effect. First, common ownership tends to be prevalent for destination-origin pairs that are close to one another — pairs over which many shipments already occur. Second, increases in trade costs propagate (via input-output linkages) throughout all industries, not only the manufacturing and wholesale industries that experience the initial decrease in productivity. Moreover, it is likely these values are lower bound estimates of the trade volume effect of eliminating common ownership, since our counterfactual calculation imposes the marginal trade effects from our estimates onto inframarginal ownership links. It is likely that the most trade-enhancing links in the economy have effects on shipment volumes considerably larger than that implied by the magnitude of our estimates.

In the subsequent rows, we compute the welfare and gross output changes that would occur if common ownership shares in destination MSAs were progressively larger. When the same-firm ownership share is five times its current value ($\kappa = 5$), the most trade-enhancing case, welfare increases by 1.2 percent and gross output by 5.6 percent relative to the initial allocation. Comparing across the $\kappa \in \{2,3,4,5\}$ cases indicates that marginal welfare gains due to the reduction in transaction costs from increasing common ownership grow non-linearly. In columns (3) and (4), corresponding to Caliendo and Parro (2015), we consider an alternative specification in which labor is immobile across regions and the share of structures and land in production equals 0. Here, counterfactual changes in gross trade flows are somewhat smaller.

In summary, our counterfactual exercises imply that increasing levels of vertical integration

[47] Caliendo and Parro (2015) did not estimate θ^j for the Furniture and Wholesaling industries. For these two industries and for the non-tradable industries, we set θ^j=5.

Table 18: Counterfactual Effects of Changing the Same-Firm Ownership Fraction

Same-firm ownership fraction	Welfare (1)	Gross Output (2)	Welfare (3)	Gross Output (4)
0 ×	-0.2	-0.1	-0.2	-0.1
1 ×	0.0	0.0	0.0	0.0
2 ×	0.2	0.1	0.2	0.1
3 ×	0.5	0.3	0.5	0.3
4 ×	0.8	1.3	0.8	1.2
5 ×	1.2	5.6	1.2	5.2
Is labor mobile?	Yes	Yes	No	No

Notes: Each row describes the counterfactual welfare and trade response, stated as percentage changes, of uniformly increasing the same-firm ownership fraction by a different factor. Welfare, as given in the first and third columns, equals the change in real wages, $d\log\left(\frac{w_i}{P_i}\right)$, averaged across all regions i.

would lead to both higher trade flows and higher welfare. We emphasize that this exercise is meant only to assess the aggregate implications of across-establishment trade costs, one of the several channels through which firm ownership patterns affect consumer welfare. We argue in our earlier work that the private benefits of vertical integration are not primarily motivated by easing the flows of physical inputs along production chains. Thus, it is possible that the figures in Table 18 understate the welfare effects of vertical integration. On the other hand, in our application of Caliendo et al. (2018)'s perfect-competition-based framework, we did not attempt to assess the effect of changing ownership patterns on markups or product availability. It is certainly possible that, through market foreclosure and other anti-competitive practices, increased vertical integration may lead to lower trade flows and consumer welfare compared to what we report in Table 18. Thus, the counterfactual exercises in this section are only a first step, albeit an important one, toward measuring the aggregate effects of alternative ownership patterns.

Appendix References

Caliendo, Lorenzo, and Fernando Parro, "Estimates of the Trade and Welfare Effects of NAFTA," *Review of Economic Studies*, 82 (2015), 1–44.

Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte, "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy," *Review of Economic Studies*, 85 (2018), 2042–96.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, "An Anatomy of International Trade:

Evidence from French Firms," *Econometrica*, 79 (2011), 1453–98.

Eaton, Jonathan, Samuel Kortum, and Sebastian Sotelo, "International Trade: Linking Micro and Macro," in Daron Acemoglu, Manuel Arellano, and Eddie Dekel, eds., *Advances in Economics and Econometrics: Theory and Applications*, Tenth World Congress (2012), 329–70.

Gaubert, Cecile, and Oleg Itskhoki, "Granular Comparative Advantage," Mimeo, 2016.

Windmeijer, Frank,"Moment Conditions for Fixed Effects Count Data Models with Endogenous Regressors," *Economics Letters*, 68 (2000), 21–24.

Wooldridge, Jeffrey M., "Multiplicative Panel Data Models without the Strict Exogeneity Assumption," *Econometric Theory*, 13 (1997), 667–78.