AUTOMATION AND MARKET DOMINANCE

Victor Manuel Bennett*

October 22, 2020

Abstract

Does the availability of new process technologies—like automation—reinforce the lead of dominant firms, or the opposite? Using novel plant-level data on automation, I show patterns consistent with endogenous automation adoption reducing market-leader share on average; particularly so in the most concentrated markets and when markets are not growing. I propose that whether leaders or laggards have more incentive to invest in the new technology depends on the balance of two effects: a cost-spreading effect and a market-stealing effect. In growing markets, the cost spreading effect dominates and process improvements entrench leaders. In more "zero-sum" markets, laggards' incentives to adopt are greater and automation becomes a force for market parity.

JEL: L11, L25, L6, O31, O33

Preliminary. Please do not cite without permission.

*Fuqua School of Business, Duke University, 100 Fuqua Dr., Durham, NC, 27708, victor.bennett@duke.edu. The author thanks the Fuqua School of Business for funding. The author also thanks the editorial team, Ashish Arora, Inês Black, Dan Gross, Susan Helper, Frank Levy, Pino Lopomo, Ryan McDevitt, Paige Ouimet, Andrea Patacconi, Rafaella Sadun, Rob Seamans, Jason Snyder, Chad Syverson, Steve Tadelis, and Nico Trachter for generously reading prior drafts; Daron Acemoglu, David Autor, Jim Bessen, Jason Furman, Bert Grider from Census, Shane Greenstein, Katheryn Russ, Mark Schankerman, Jay Shambaugh, Tim Simcoe, and John Van Reenen for conversations; Robert McElmurry from Fanuc USA, Kira Noodleman from Bee Partners, Arthur Petron from OpenAI, and Dan St. Louis from the North Carolina Manufacturing Solutions Center for generous industry background; Brad Jensen for generous access to his calculations; Rebecca Steorts, Davis Berlind, and the rest of the Steorts Lab at Duke Statistics for methodological suggestions with respect to entity resolution algorithms; attendees of seminars arranged by the Boston University/Technology Policy Research Initiative, the Israel "Governance, Technology, and Regulation conference", and Duke Fuqua, University of Utah, and Columbia Business School for comments. Taishi Sassano and Todd Hall provided exceptional research assistance and Larisa Cioaca and Divya Sebastian provided exceptional human intelligence for classification. All errors remain my own.

1 Introduction

Scholarship on the relationship between technological change and the persistence of market dominance dates back at least a century. Much of the empirical work has focused on firms' R&D investment and patentable inventions, despite the fact that R&D investment and patenting are confined to relatively few firms and industries (Bound et al., 1984; Cohen, 2010) and that in some industry studies, substantial productivity gains have been documented from adopting technology available on the open market (see Syverson, 2011, for examples).

This study empirically investigates the relationship between commonly available industrial automation capital—including, but not limited to, robots—and market dominance. While I find patterns consistent with a "cost-spreading" effect (as in Cohen and Klepper, 1996) that leads to dominant firms investing more in automation, I also find evidence of a second offsetting effect. The latter, a "market-stealing", effect that drives lagging firms to invest, can actually dominate in markets that are not growing and such that competition is more "zero-sum." The net effect in these cases is that firms' endogenous adoption of technology leads to reduced dominance of leading firms.

This empirical study was made possible by two empirical innovations. The first is that I construct data that distinguishes discrete US markets using structural estimates of the geographic reach of different industries from Gervais and Jensen (2019). Defining geographic regions is critical for measuring the gap between firms and has, to date, only been accomplished in specific industries (e.g., Bresnahan and Reiss, 1991; Syverson, 2004); the definitions I use allow for studying a much broader range of industries and allow between-market-within-industry comparisons.

My second major innovation is the construction of establishment-level panel data on investments in automation across industries.[1] These data are obtained through a novel classification algorithm separating all shipments through US maritime customs during 1994–2014 by whether or not they are expenditures on automation. A second algorithm allows linking those shipments to establishment-level performance data from the National Establishment Time Series (NETS).

These new data are this study's first contribution. Furthermore, these two innovations allow me to associate automation with changes in market dominance at the market level and, even better, investment at the establishment level with position in the market structure.[2]

This study contributes new empirical findings to a large literature on the competitive implications of technological innovations. That work studies whether and when new technology primarily benefits entrants versus incumbents and thus whether

[1]While there are some data on process technology, primarily used in the empirical work on labor market implications of automation, they are generally not suitable for this question because they are based on classifications of jobs (e.g., Autor, Levy and Murnane, 2003; Felten, Raj and Seamans, 2018), infer automation from production function residuals (e.g., Autor and Salomons, 2018), measure product development and not adoption (Mann and Püttmann, 2017; Webb, 2020), or do not allow distinguishing establishment-level adoption (e.g., Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020).

[2]A few other working papers appeared concurrently with this one using firm-level automation data, some from shipping data (e.g., Bessen et al., 2020; Dixon, Hong and Wu, 2019; Humlum, 2020; Koch, Manuylov and Smolka, 2019; Acemoglu, Lelarge and Restrepo, 2020). All of those focus on issues of labor and the labor share, making this study unique in focusing on competitive dynamics. Furthermore, those papers focus specifically on robots, which form a small share of automation equipment. Shipments flagged as robots for having the stem of "robot" in the commodity description comprise 1.8% of the automation shipments, as classified by my algorithm, and 3.1% of the dollar value. The papers using shipping data use codes from the Harmonized System (HS) codes for robotics to classify shipments, which cover only a small share of automation capital. Even expanding to the six-digit HS code from the ten-digit HS codes that correspond to robots, these shipments constitute 1.7% of automation shipments and 1.5% of the dollar value.

innovations lead to entrenchment or displacement. That literature, influenced by observations in the Schumpeterian tradition that larger incumbents are often displaced with waves of new technology (e.g., Christensen and Bower, 1996; Henderson and Clark, 1990; Gans, Hsu and Stern, 2002; Bennett, 2020), has sought to document offsetting effects that might lead newer or smaller firms to flourish under some conditions.

In addition to the literature on technologies' roles in displacement of large firms, there is a literature on the endogenous adoption of existing process technology by different types of firms.[3] For example, Åstebro (2002) investigates the effect of firm and plant size on adoption of computer-aided design (CAD) and computer numerical control (CNC) technology and finds support for the cost-spreading logic. This study suggests that the performance of competitors and whether the market is growing are important considerations in that decision. Karshenas and Stoneman (1993) also look at CNC and investigate an "order effect," among others, finding little support for it. This paper suggests that the lack of support may be due to the fact that in addition to the authors' theorized ordinal effects, cardinal difference in performance and whether markets are growing can drive the adoption decision.

These empirical findings also contribute to a recent phenomenological literature on the implications of marginal-cost-reducing process improvements for market dominance (e.g., Bessen, 2017; Autor et al., 2020; Van Reenen, 2018). These studies have

[3]There was also a stream of theoretical literature on the implications of marginal-cost-reducing process improvements for market dominance (e.g., Flaherty, 1980; Spence, 1984; Athey and Schmutzler, 2001). While the models introduced by that literature helped introduce theoretical mechanisms that should be considered, their flexibility allowed for such a dissatisfyingly wide range of outcomes (see Sutton, 1996, 1998) that the theoretical literature was largely abandoned in favor of empirical studies, of which this is one.

largely used a cost-spreading logic in arguing that dominant firms will adopt such technologies and thus become increasingly dominant. I find patterns consistent with that logic, but also finds patterns consistent with an offsetting "market-stealing" effect in markets that are not growing. This finding puts boundary conditions on the assertion that marginal-cost-reducing process improvements will lead to increasing dominance and suggests that in markets that grow less, laggard adoption may dominate leader adoption, such that automation will endogenously lead not to increasing dominance, but rather to greater competitive parity.

The documentation of the market-stealing effect, in suggesting that even technologies increasing economies of scale may not necessarily lead to entrenchment by established firms, joins recent work suggesting that new technologies, both labor-substituting (Belenzon, Bennett and Patacconi, 2019) and labor-complementing (Bennett and Hall, 2020), can increase the rate of entry of new firms.

2 Empirical investigation of competitive position and automation

The central question of the theoretical literature on marginal-cost-reducing process improvements for market dominance was whether automation would increase the dominance of leading firms (e.g., Flaherty, 1980; Spence, 1984; Athey and Schmutzler, 2001). Those models have slightly different structures, but share a common structure. In all of them, firms can make fixed cost investments in reducing marginal cost of production. The most recent of these, Athey and Schmutzler (2001), characterizes a

Markovian game in which myopic firms invest in cost reduction based on their own current marginal cost and that of their competitor, which together completely define the state. Whether "increasing dominance" of leading firms obtains in all of these models depends on whether returns to adoption are greater for leading firms.

The returns-to-investment function is not directly observable, but the empirical predictions can be operationalized by two hypotheses. The first is that dominant firms would be more likely to adopt process improvements. The second is that this tendency would be increasing in their lead. I, therefore, begin my analysis of the phenomenon by constructing data that will allow me to test those two hypotheses, which provide the structure for my initial analysis. In Section 2.1, I describe the data used to investigate these hypotheses. Foreshadowing my results, the patterns are not, on average, consistent with increasing dominance in this empirical setting. Instead, I find that leading firms' investment is, on average, decreasing in their lead, while lagging firms' investment is increasing in that gap. Furthermore, I find that these patterns of investment correspond to decreasing dominance by leading firms as a result of equilibrium investment.

In Section 3, I outline a stationary game in the spirit of the Athey and Schmutzler (2001) model that generates the patterns observed in the data, but simplified to highlight the mechanisms at play. I then derive new predictions of that model that are not predicted by the existing models. Specifically, my simple model predicts that the equilibrium effect of investment depends on whether the market is growing, a contingency not raised in prior models. I conduct additional analysis in which I find patterns consistent with this novel prediction, lending support to the mechanisms highlighted in the model described and providing a theoretical contribution to the literature on process improvements for market dominance.

2.1 Data

2.1.1 Defining markets

Much of the recent work on market structure has measured concentration at the national level (e.g., Autor and Salomons, 2018; Autor et al., 2020; Gutiérrez and Philippon, 2017), though national concentration measures do not capture the important features of market dominance (see Shapiro, 2018; Syverson, 2019, for discussion). While industry-by-geography market definition is not as ideal as would be product-by-geography definition, it is a better representation of what consumers face and may yield dramatically different results.[4]

Subnational market border definition has the added benefit of improving identification by allowing for comparison across geographic markets within time and within industry.

To date, though, there is no industry-level classification of market size that can be used to classify catchment areas instead of just assuming that all industries share the same catchment area. To construct one, I use the classification of transportability from Gervais and Jensen (2019). This work is part of a small literature that

[4]In fact, Rossi-Hansberg, Sarte and Trachter (2020) show that if CBSA, county, or ZIP code levels are a better representation of markets, the trend in concentration may actually be decreasing instead of increasing. Rinz (2018) shows the same patterns using data from the US Census Longitudinal Business Database (LBD) (Jarmin and Miranda, 2002). Handwerker and Dey (2019) confirm this finding with BLS Occupational Employment Statistics (OES) merged with the Quarterly Census of Employment and Wages (QCEW).

structurally measures the transportability of products of different industries, largely to study trade. Using differences in the distribution of population from the agglomeration of businesses, this work creates a unit-free index of industry-level product transportability (SES). The innovation in this manuscript is to calibrate that index to define geographic boundaries within the United States using industries in which the size of markets is known. For example, the catchment area for cement described by Syverson (2004) corresponds roughly to the average geographic area of a four-digit ZCTA level. Because the cement industry code receives an *SES* score of .25, I define all industries with scores at or below that level to the 4-digit ZCTA-level. Local broadcast media (television and radio) receive *SES* scores of .5. Because MSAs were developed to represent the markets reached by these media, I define the catchment area for all industries with scores in $(.25, .5]$ as the MSA level. Higher *SES* scores, like those for movie production, indicate national (or greater) markets.

In summary, I classify industries into one of three categories: local ($SES \leq .25$, defined at the 4-digit ZCTA-level),[5] regional ($.25 < SES \leq .5$, defined at the MSA-level)[6], and national ($.5 > SES$). For much the analysis below, I restrict myself to local and regional markets, as having other markets within the industry-by-year is critical for identification and the market positions in markets in which foreign producers may compete would be mis-measured.

[5] All the results presented below are robust to defining local markets at the three-digit ZCTA level.

[6] Establishments outside of MSAs are assigned to a synthetic MSA at the three-digit ZCTA level and results are robust to their exclusion.

2.1.2 Contributing datasets

National Establishment Time Series (NETS) The primary source of establishment-level data is the *Manufacturing Sample* of the National Establishment Time Series (NETS), which is a panelized version of annual Dun & Bradstreet publications. The NETS has been used extensively by academics (e.g., Rossi-Hansberg, Sarte and Trachter, 2020; Neumark, Wall and Zhang, 2011), and its properties have been compared to those of Census administrative records by Barnatchez, Crane and Decker (2017) and Haltiwanger, Jarmin and Miranda (2013).[7]

Barnatchez, Crane and Decker (2017) suggest that comparability to administrative records can be improved by aggregating D&B establishments within a geography. However, their specific recommended aggregation often aggregates establishments in different industries. Given the importance of industry classification to the phenomenon of interest here, the results presented follow the spirit of aggregation and aggregate establishments with the same name, city, and state. I also follow the suggestion from Barnatchez, Crane and Decker (2017) to remove a list of industries and restrict attention to establishments with employee counts above 10 and below 1000.

The full set of industries included and their representation in the aggregate data is provided in Appendix A.2.3.

Port Import/Export Reporting Service (PIERS) The Port Import/Export Reporting Service (PIERS) data are provided by market research company IHS-Markit and comprise the population of bills-of-lading from US ports during the sam-

[7] I conduct an additional validation exercise focused on the revenue measures using Compustat data in Appendix A.6.4

ple period. These data have been used in economics, but predominantly for studying international trade (e.g., Hummels, 2007). To the best of my knowledge, they have not been used to study establishment-level outcomes and the recipient addresses for shipments have not been used by economists.

The data include (a) a text description of the commodity shipped, counts of the item in the shipment, and the shipment size and weight (b) the names and addresses of the shipper and recipient (c) classifications of the goods for tariff purposes (including the six-digit Harmonized System (HS) category); and c) details of the vessel bringing the shipment.

I use these data to measure the flow of industrial automation equipment to individual US establishments. There is general consensus among professionals working on automation that the vast majority of equipment comes from overseas,[8] but I verify the robustness of my results empirically in Section A.6.4.

A series of purpose-built machine-learning (ML) algorithms were used to classify commodity descriptions as industrial automation or not. For ease of interpretation, the results presented below are based on a simple heuristic derived from the token loadings, but results using three distinct ML algorithms are presented in section A.6.5.

A commodity is classified as automation equipment if any of the following hold:

1. Includes a stem[9] of "machine" but not stems of tokens "sewing", "coffee",

[8]Focusing only on robotics, Richard Blake said in Forbes: *"Makers of robots are almost entirely based outside the U.S. (e.g., Japan's Fanuc, Switzerlands ABB, South Koreas Doosan and Denmarks Universal Robots)"* (Blake, 2019). This is corroborated in the comparison of firms order "robots" to those ordering more automation capital in Appendix A.2.2

[9]In natural language processing and information theory, stemming is "a procedure to reduce

"walking", "washing", "fax" or "facsimile",

2. Includes a stem of "robot" but not the stem of the token "toy",

3. Includes a stem of "automatic".

This classification algorithm has the advantage of being broader than the ill-defined term "robots", whose associated shipments make up a small fraction of automation equipment and may bias results toward firms with broad enough product offerings to purchase general purpose arms and gantries from firms like Fanuc, Kuka, Universal, and ABB and away from specialized automation equipment—such as bottle fillers, boarding machines, automatic looms, and painting rigs—which unambiguously automate tasks.

I then convert the flow of shipments to a stock by computing the depreciated value of shipments received in the last 15 years, with shipments depreciating by 20 percent each year. This value was chosen to match accounting conventions, though the results are qualitatively robust to alternate depreciation rates.

2.1.3 Resulting aggregate data

The two datasets described above are algorithmically matched on name using a procedure described in Appendix A.2.1.[10]

The data resulting from the matches have observations at the establishment-by-year level with columns indicating the industry (six-digit NAICS), geographic

all words with the same stem to a common form" for parsimonious string matching (Lovins, 1968). For example, "mach" is the stem for "machine", "machinery", and "machines".

[10] I also demonstrate robustness to the particular choice of matching algorithm in that section and, in A.6.6, show that estimation on the matched sample is actually biased against the results found below.

market, sales, employment, and automation capital measured in dollars as both a flow and stock. Because the phenomenon of interest is automation, I remove industries in which no expenditure of capital is measured during the entire data series by any establishment, thus restricting data to the risk set. As described in the introduction, I also restrict attention to industries in which catchment areas are subnational, to allow for identification using multiple markets, and restrict attention to settings in which foreign goods may imply mismeasured competitive position. I also remove all monopolist markets given, my focus on the implications of competitive imbalance among extant firms.

The resulting data cover 1,375,024 records in 25,943 industry-by-markets in 115 industries in years the 1995–2014. These records represent 746,532 unique establishments, as defined above. These establishments represent 97,930 unique firms, as defined by unique headquarters D&B identifiers in the NETS data. 29.4 percent of entries are in local markets and the rest are regional. Table 1 presents summary statistics for establishment-by-year matched data.[11]

Table 1: Summary statistics from matched data

	Mean	Std. dev.	Min.	Max.
Value of automation capital shipments in year	3314.813	(180816)	0	7.06e+07
Stock of automation capital shipments in year	11730.54	(541629)	0	2.08e+08
Sales in year ($M)	8.614493	(85.6418)	0	45496
Employees in year	66.02781	(104.0592)	11	1000
Sales leader [0,1]	.2106741	(.4077875)	0	1
Observations	1375024			

The following section describes the analysis performed on these data.

[11] In the appendix, Figure 11 displays the geographic distribution of the data within the US at the 2014 FIPS county-level.

2.2 Analysis

2.2.1 Preliminary empirical analysis

As described above, the empirically observable constructs from the literature on marginal-cost-reducing investment and market dominance are (a) The market share of the leading firm, (b) the gap to a trailing firm, and (c) investment.

Here, I operationalize g, the difference in revenue between the leader (indexed by 1) and the second-place firm, normalized by the leader's revenue, as:

$$gap_i = \frac{revenue_1 - revenue_2}{revenue_1}. \tag{1}$$

The relationship of interest, therefore, is between the stock of automation capital and this *gap*. The reduced-form empirical model is therefore:

$$\begin{aligned} &Ln(Automation\_Capital\_Stock_{it} + 1) = \\ &\qquad \alpha + \beta_1(gap_{it-1}) + \beta_2(gap_{it-1} \times Leader) + \beta_3 Leader \\ &\qquad + \nu(Controls_{it-1}) + \gamma_i + \eta_{mt} + \epsilon_{it}. \end{aligned} \tag{2}$$

To account for secular trends at the industry k level, I include fixed effects at the year–by–six-digit-NAICS level, η_{kt}. I also include fixed effects at the establishment level to account for time-invariant variation due to idiosyncratic features of a particular establishment, γ_i. The time-varying controls are the log of 1 plus the focal establishment's sales at time $t-1$ and the log of 1 plus count of employees, also at time $t-1$. Standard errors are allowed to cluster at the market-by-year level to account for correlation within market-by-period.

While the reduced-form Equation 2 represents the relationship I would like to study and Table A3 presents the results of estimating that equation, that table is relegated to Appendix A.4 as the coefficients are biased (Nickell, 1981). To see this, note that *gap* is a function of automation investments in the prior period, so a simplified version of the estimation equation can be written as follows

$$
\begin{aligned}
y_{it} \equiv Ln(Automation\_Capital\_Stock_{it} + 1) = \qquad (3) \\
\alpha + \beta_1(gap_{it-1}) + \epsilon_{it} = \\
\alpha + \beta_1 f(Ln(Automation\_Capital\_Stock_{it}) + (\nu_i + \zeta_{it}),
\end{aligned}
$$

where $\epsilon_{it} = \nu_i + \zeta_{it}$ and ν_i is a stochastic and unobserved individual-specific time-invariant effect which allows for heterogeneity in the means of y_{it}.

In this form, one can see that the estimation equation represents a situation of dynamic panel data that Nickell (1981) showed to be biased because ν_i is correlated with both the dependent variable, $Ln(Automation\_Capital\_Stock_{it} + 1)$, and the lagged version, $Ln(Automation\_Capital\_Stock_{it-1} + 1)$, violating the requirement of conditional exogeneity for the OLS estimator to be unbiased.

Including a time-invariant fixed effect γ_i eliminates the ν_i component of the error because the estimated value of γ_i is ν_i, so it drops out. When the number of observations N is large relative to the number of time periods T, however, even the fixed-effects/within-groups estimator is biased.

The lagged dependent variable, after subtracting the fixed effect, can be written:

$$y_{i,t1} - \frac{1}{T-1}(y_{i1} + ... + y_{it} + ... + y_{i,T1}). \tag{4}$$

The error term, after subtracting the fixed effect, can be written:

$$\zeta_{it} - \frac{1}{T-1}(\zeta_{i2} + ... + \zeta_{i,t1} + ... + \zeta_{iT}). \tag{5}$$

One can see that the term $\frac{-y_{it}}{T-1}$ in Equation 4 is correlated with ζ_{it} in Equation 5 and the term $\frac{-\zeta_{-1}}{T-1}$ in Equation 5 is correlated with $y_{i,t1}$ in Equation 4. Nickell (1981) showed that this means that even the fixed-effects/within-groups estimator is inconsistent when T is not sufficiently large.

There are two solutions to the issue of bias in the dynamic panel data estimator. Both require instrumenting for the lagged dependent variable. In this case, if the instrument for the gap meets the exclusion restriction—that is, it is not correlated with lagged investment by the focal establishment—the bias is removed.

One approach is to use a generalized method of moments (GMM) estimator that makes structural assumptions about the order of autoregression in the data generating process. In Appendix A.6.1, I demonstrate the results of estimating the systems dynamic panel estimator introduced by Blundell and Bond (1998) as an improvement over the one introduced by Arellano and Bond (1991).

My preferred approach, shown below, is to use a different instrument, driven by assumptions on the competitive process.[12] Here, I assume that gap_{it} is a function

[12]The direction of estimates between the systems dynamic panel data estimator and of the one

of (a) some idiosyncratic time-invariant component x_m with a common mean at the industry k level X_k, (b) the initial marginal costs of the establishments in the market $mc_{j0} \forall j \in M$, and (c) the cumulative investments of the establishments in the focal market:

$$gap_{it} \equiv g(x_m, mc_{j0} \forall j \in M, d_{j1} \forall j \in M ... d_{jt-1} \forall j \in M).$$

The assumption of the common mean seems plausible given our knowledge of industries outside the data. For example, telecommunications markets have different gaps between leading and lagging firms, but tend to have large gaps because they share the properties of having large fixed costs relative to marginal costs and tend not to be differentiated, though that varies across markets. Professional services tend to have smaller gaps because they share the property of having large marginal costs relative to fixed costs and tend to be more differentiated, though again that varies by market.

Under this common mean assumption, the average gap across other markets in the industry is correlated with the gaps in the focal market, but not with any of the lagged automation capital investments in the focal market. This instrument therefore removes the Nickell (1981) bias.[13] In the estimations in Table 2 below, I use the Herfindahl-Hirschman Index of other markets in the industry instead of the gap between leader and laggard because it is common across establishments in the

discussed here are the same, providing confidence in the approach.

[13]One threat to the exclusion restriction is if firms with establishments in multiple markets make investment decisions for all establishments or if establishments respond to choices made by sibling establishments of their competitors in other markets. Both cases introduce a correlation between the idiosyncratic market components x_m within the industry, re-introducing the Nickell (1981) bias. In section A.6.6, I demonstrate robustness to accounting for this by estimating results only on single-establishment firms.

industry, but has the same desirable properties.

Column 1 in Table 2 presents the results of estimating Equation 2, which suggest that investment is actually increasing in the gap for laggards and decreasing for leaders. In Column 2, I verify that this result is robust to accounting for lagged investment.

Note that because two values are instrumented for in these estimations, the main effect of gap_{it} and its interaction with the *leader* indicator, the recommended indicator of the strength of the instrument in the first stage is the Kleibergen-Paap rk Wald F-statistic. Using the thresholds recommended by Stock and Yogo (2005), the instruments are strong enough that the Wald would reject at less than 10 percent if the true power were 5 percent.

The differences between these results and those in the endogenous regressions in Table A3 support the assumed structure of the endogeneity, which biases the results toward suggesting increasing dominance. Accounting for that endogeneity and removing the bias changes the result to suggest decreasing dominance.[14]

To aid interpretation of the results and allow for a more flexible response function, Figure 1 shows the results of plotting the regression-predicted marginal effect of the market share gap in dollars on the range of possible gaps.

[14] Appendix Table A.3.1 shows the results of analysis of market-level outcomes, using a different identification strategy, yielding results consistent with those here.

Table 2: Establishment-level effect of competitive asymmetry on automation investment

	(1)	(2)
VARIABLES	Ln(1+ Automation capital stock value)	
Revenue gap [lag]	11.20***	2.052**
	(3.197)	(0.869)
Leader × Revenue gap [lag]	-19.30***	-3.551**
	(5.525)	(1.502)
Leader [lag]	7.720***	1.423**
	(2.203)	(0.599)
Ln(1+Sales) [lag]	2.116***	0.392**
	(0.601)	(0.163)
Ln(1+Employees) [lag]	-0.0107	-0.00220
	(0.0257)	(0.00726)
Ln(1+Automation capital stock value) [lag]		0.842***
		(0.00275)
Observations	1,358,765	1,358,765
Model	2SLS	2SLS
Establishment FE	Yes	Yes
NAICS6 × Year FE	Yes	Yes
Markets	Subnational	Subnational
Kleibergen-Paap rk Wald F	7.24	7.24

Note: Errors clustered at market-by-year level. Using the critical values from Stock and Yogo (2005), the instruments are strong enough that the Wald would reject at less than 10 percent if the true power were 5 percent. *** p<0.01, ** p<0.05, * p<0.1

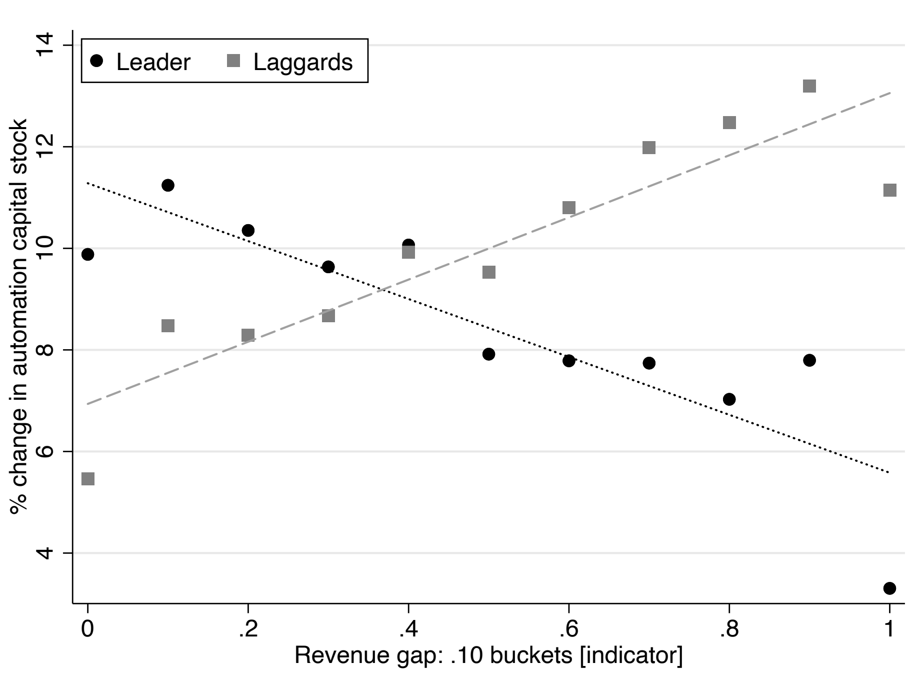

Figure 1: Marginal effect of the competitive asymmetry on dollars of automation investment

3 Illustrative example of a model under asymmetric Cournot competition generating the uncovered patterns

Consider a setting with two firms $i \in \{L, H\}$ in two stages. At the beginning of the first stage, both firms draw a random marginal cost of production.[15] Without loss of generality, I represent the low-cost firm's marginal production cost by $c_L = c > 0$ and the high-cost firm's cost by $c_H = cb$, where $b > 1$.[16]

Taking into account the commonly known initial marginal costs of production, each firm simultaneously decides a level $d_i \geq 0$ of automation capital to acquire at per-unit price f. This automation capital will reduce that marginal cost of production, the resulting marginal cost for firm i being $c_i = c_i(d_i)$ where $c' < 0$.

In the second stage, the firms play a simple asymmetric-cost Cournot duopoly game. They face an aggregate demand such that the prevailing market price is $p = A - Q$, where $Q = \sum_{i \in \{L,H\}} q_i$ is aggregate industry output.

In the Cournot duopoly game, firms choose their individual quantities to maximize individual profit, $\pi_i = [p - c_i(d_i)]q_i - fd_i$.

In the first stage, they choose their investment level to maximize profit considering

[15]The stationary structure of the model and random initial marginal cost of production are intended as analogues to Athey and Schmutzler (2001) and to highlight the mechanisms of interest. The structure is intended to provide a tractable and simple representation of the observed phenomena and not to suggest that other factors, including internal organizational considerations and additional competitive factors, are not at play in the real world.

[16]I test whether patterns are consistent with the Cournot result that exogenous reduction in marginal cost of production will increase sales quantity in Appendix A.6.2. The coefficient of an instrumental variables regression of sales on automation expenditure—which represents exogenous investment—is positive and significant.

the equilibrium quantities that will be produced in the subsequent stage.

The familiar unique equilibrium quantities in the second stage are:

$$q_i^*(d_i) = \frac{1}{3}(A + c_j(d_j) - 2c_i(d_i)). \tag{6}$$

Given Equation 6, the first-stage profit function for firm i is:

$$\begin{aligned}
\pi_i(d_i) &= [A - q_i(d_i) - q_j(d_j) - c_i(d_i)]q_i(d_i) - fd_i \\
&= [A - \frac{1}{3}(A + c_j(d_j) - 2c_i(d_i)) - \frac{1}{3}(A + c_i(d_i) - 2c_j(d_j)) - c_i(d_i)] \\
&\quad \cdot\frac{1}{3}(A + c_j(d_j) - 2c_i(d_i)) - fd_i \\
&= \frac{1}{9}(A + c_j(d_j) - 2c_i(d_i))^2 - fd_i.
\end{aligned}$$

From this equation, it is simple to verify that the derivative of the profit function with respect to investment and then quantity q is positive, $\frac{\partial \pi_i}{\partial d_i \partial q_i} > 0$. This means that, all else equal, the local returns to investment are increasing in quantity produced. This is a familiar result due to the increasing gap between the average cost and marginal costs of production. That effect in which returns to investment are greater for leading firms—which has a long tradition among innovation scholars, often referred to as the "cost-spreading effect" after Cohen and Klepper (1992, 1996)—is the increase in the intensive margin of profit. When this effect dominates, we would expect increasing returns to scale and increasing dominance in the Athey and Schmutzler (2001) sense. The demonstration of this effect is important for understanding the mechanisms at play, so I return to it below.

Empirical predictions in terms of marginal costs of production are not particularly useful given that marginal cost of production is not observable by the empiricist. It is, therefore, helpful to reframe the model in terms of quantities which are observable. I therefore define a variable describing the gap in revenue between the firms g as follows:

$$g = \frac{pq_L - pq_H}{pq_L}. \tag{7}$$

Note that g ranges from zero, when quantities produced are equal, to unity as the high-cost producer's output approaches zero. Note also that the common price causes price to drop out.

Because the quantities produced are smooth continuous functions of b, they can be inverted, allowing the isolation of b as a function of the unique revenue gap it implies:

$$b = \frac{-A(g) + 2c(g) - 3c}{c(g-3)}. \tag{8}$$

Substituting for b in the profit functions for the two firms allows me to solve for cross-partials and characterize the equilibrium using the monotone comparative statics approach. The relevant cross-partials are:

$$\frac{\partial \pi_L}{\partial d_L \partial d_H} = -\frac{4}{9}\frac{\partial c_L}{\partial d_L}\frac{\partial c_H}{\partial d_H} < 0 \tag{9}$$

$$\frac{\partial \pi_H}{\partial d_L \partial d_H} = -\frac{4}{9}\frac{\partial c_L}{\partial d_L}\frac{\partial c_H}{\partial d_H} < 0 \tag{10}$$

The fact that both cross-partials are negative indicates that investment expenditures are strategic substitutes, meaning that the dominant response by a focal firm to the other firm increasing expenditure on capital is to reduce expenditure.

Interestingly, in this game, those expenditures depend on b, the gap in costs before expenditure. Specifically, there is a class of games in which the greater the original gap in costs, which is one-to-one with b, the larger the investment by the laggard and the smaller the investment by the leader, meaning that automation is actually reducing the performance gap.

Proposition 1. *Post-investment market dominance of the leader is non-increasing in g_{pre}, defined as what g would be absent expenditure, with equilibrium expenditure on marginal cost reducing capital, $\{d_L, d_H\}$, when the return to investment is such that:*

$$\frac{4}{9}(2\frac{\partial c_H}{\partial d_H}\frac{\partial c_H}{\partial b} - (A + c_L(d_L) - 2c_H(d_H))\frac{\partial c_H}{\partial d_H \partial b}) < 0. \tag{11}$$

Proof. The returns to expenditure by the leading firm are increasing for all values of $b > 1$:

$$\frac{\partial \pi_L}{\partial d_L \partial b} = -\frac{4}{9}\frac{\partial c_L}{\partial d_L}\frac{\partial c_H}{\partial b} > 0. \tag{12}$$

Combined with the fact that investments are strategic substitutes, that implies that when $\frac{\partial \pi_H}{\partial d_H \partial b} < 0$, equilibrium expenditure for the laggard, d_H^*, will be increasing and equilibrium expenditure by the leader, d_L^*, will be decreasing for all $b > 1$. This occurs if and only if condition 11 is met.

Because g is monotonic and one-to-one with b, that condition also implies that d_L is non-increasing in g and d_H is non-decreasing. Laggard spending exceeding leader spending implies that post-investment costs are weakly closer than pre-spending costs, implying increased parity. $\square$

This effect highlights a mechanism that, to my knowledge, has not been highlighted in the literature: a "market-stealing effect." The incentive to invest depends on how much market share would be gained through the investment. Market stealing is nonmonotonic in the cost difference between firms. As the cost gap increases, the market share increases, but at a decreasing rate:

$$\frac{\partial}{\partial b}\frac{q_L - q_H}{q_L} = \frac{3(A-c)c}{(A+(b-2)c)^2}. \tag{13}$$

Essentially, a firm with very high market share has very little market left to gain through cost decreases.

That market-stealing effect is analogous to the effect described by Arrow (1962), but for process improvements. Arrow (1962) noted that accounting for market share— the extensive margin of profit—is critical in a competitive model. In that paper, Arrow refers to the threat of product innovations cannibalizing a leader's own share, reducing that leader's incentive to invest in product innovations.

Furthermore, the extensive margin of the total market $Q = q_L + q_H$ is decreasing in the cost gap, $\frac{\partial Q}{\partial b} = -\frac{c}{3}$, so not only does the marginal increase in share decrease, but the increase in the whole from which that share comes decreases. Thus, increasing expenditure yields diminishing returns for leading firms and increasing returns for lagging firms, as the cost gap increases.[17]

From those equations, one can also see that the market-stealing effect depends on the elasticity of demand to quantity. In a growing market, it is possible that both the leading and lagging firms gain market—though not market *share*—by investing.

[17] I provide a simple real-world example case study in Appendix A.1.

In this case, investments by the two firms may not be strategic substitutes, and the cost-spreading effect, described earlier, dominates the market-stealing effect.

For example, in Klepper's 1996 model, the market expands and each firm serves a share of the new market equal to its share in the prior period. In that setting, cost reduction returns no new share for laggards, guaranteeing that the cost-spreading effect dominates.

The mode of competition, highlighted by Athey and Schmutzler (2001), is also a crucial determinant of the equilibrium outcome.

In the asymmetric-cost Cournot depicted above, both firms produce, but in non-linear proportions based on the difference in their costs. In Cournot, the returns to cost advantages are concave: as the difference between costs increases—indicating that the leader is more dominant pre-investment—the quantity returns to decreasing marginal cost decrease. This means that (a) the cost-spreading effect grows more slowly for leaders and more quickly for laggards and (b) the market-stealing effect decreases more quickly for leaders and increases more quickly for laggards. For some investment technologies, this implies that leader investment is decreasing, and laggard investment is increasing, in the leaders' pre-investment advantage.

In traditional undifferentiated Bertrand, for example, the entire market goes to the cost leader at the second-place firm's cost (or ε below). In that case, the leader earns additional profits over the entire market from cost reduction. Laggards earn nothing from cost reduction unless it is pivotal to capturing the entire market. This means that the cost-spreading effect will dominate in all regions with the possible exception of "neck-and-neck" competition where the cost of investment required to capture leadership may be low enough that share-stealing dominates and the laggard becomes the leader.

The empirical results are, on average, consistent with the asymmetric cost Cournot predictions, so I use this stylized representation of competition to derive additional predictions.

It should be noted that this model does not capture all of the possible effects at play. For example, in industries with network effects, returns to scale might come from demand-side returns to scale instead of production returns. The model also does not include intertemporal strategic investment behavior such as deterrence behavior. As with any model, one must abstract from particular features and concentrate on others. The particular mechanisms I highlight here are the tradeoffs between cost-spreading and market-stealing. Determining whether a model that highlights relevant features is useful requires generating predictions that were not what the model was designed to represent and seeing if those predicted patterns are present in the data.

One specific prediction of the Cournot-based model that has not been illustrated in the literature is that whether marginal-cost reducing innovation drives increasing dominance or the opposite depends on whether or not the market is growing. In summary, the model suggests that in markets that are not growing, we would expect to see the pattern we saw in the Section 2.2.1 aggregate analysis, where endogenous adoption of automation reduces dominance. In markets that are growing sufficiently, we would predict that the cost-spreading effect would dominate, and dominance of leading firms would increase.

3.1 Novel prediction of the model

In this section, I provide additional analysis to investigate whether the predictions of the model which are not those for which the model was designed nevertheless obtain in the data. If they do, it adds support to the theoretical suggestion that this is a helpful representation of the dominant forces.

Figure 2 plots regression predicted levels of automation capital expenditure as a function of lagged sales for three classes of markets: those in which total sales shrank from the last year, those that grew less than fivefold, and those that grew more than fivefold. The figure shows that predicted investment is effectively equivalent for all but the fastest-growing markets. Those markets that are growing—the markets most similar to those depicted in Klepper (1996)—will likely demonstrate very little market-stealing. In those cases, the net result of the two effects would, theoretically, be dominated by cost-spreading. We would, therefore, expect investment to increase much more with sales, which we do see.

Having seen patterns consistent with cost-spreading in the settings in which we expect to see it at full strength, I move on to market-stealing. In settings with market-stealing, we expect lagging firms to invest more, relative to leaders, when they have more to gain.

The y-axis in Figure 3 plots expected laggard spending minus expected leader spending—generated from the regressions modeled in Column (1) of Table 2. The x-axis shows the revenue *gap*.[18]

[18]Market growth in Figure 3 is based on whole-market revenue growth. This is my preferred specification because it allows growth to differ between industries in a geography. To validate the robustness of this approach, Appendix A.6.3 shows an analogous graph, but with growth defined

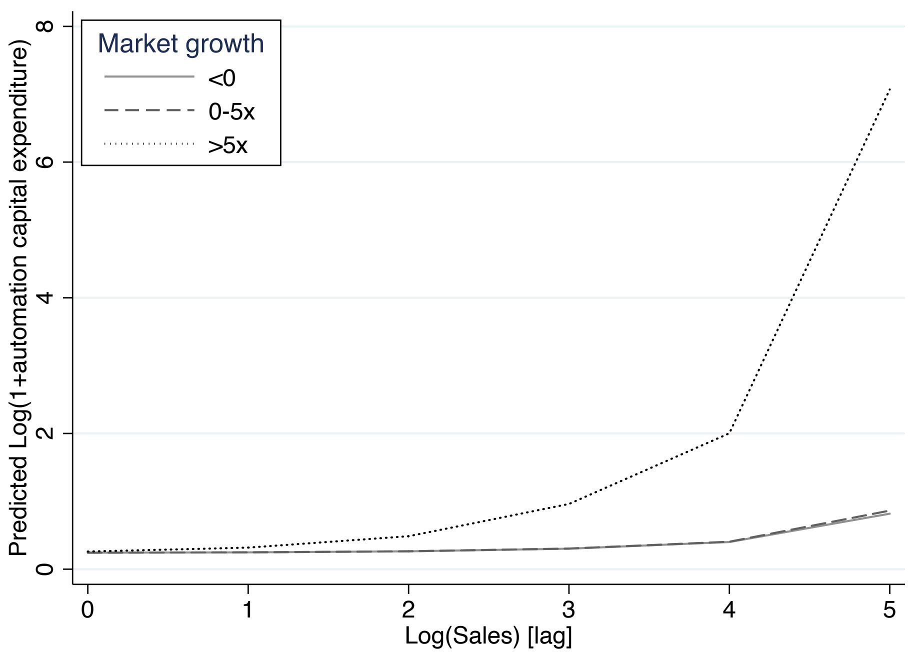

Figure 2: Cost-spreading: Firm scale and investment in automation capital

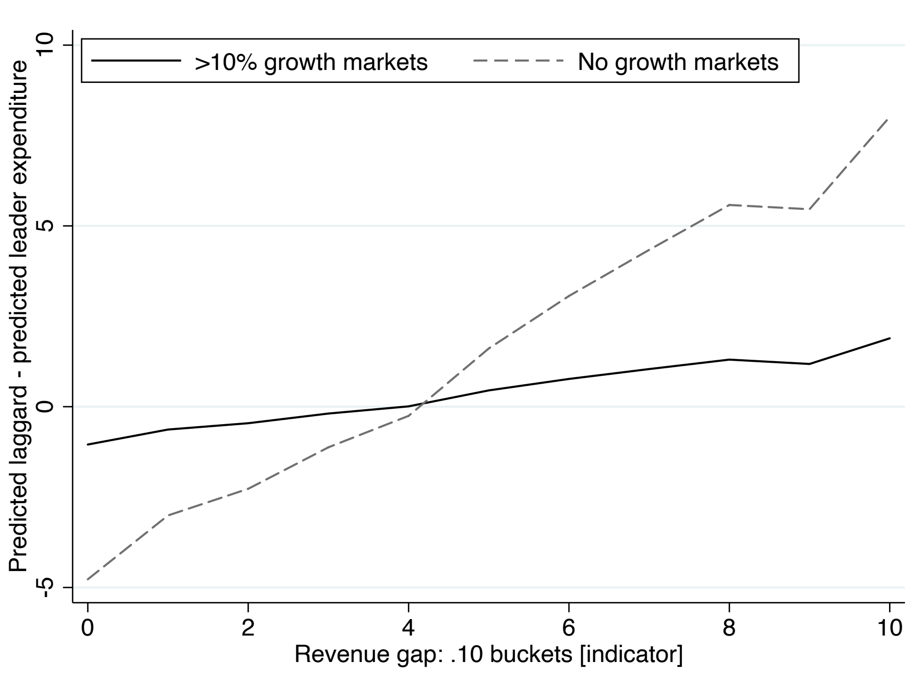

Figure 3: Market growth and the balance of cost-spreading and market-stealing effects

The two lines represent the relationships for growing and non-growing markets, respectively. The plot shows an increase in laggard spending relative to leader spending, until it exceeds leader spending, for markets that are not growing. Observing that pattern in the aggregate depends on the preponderance of growing versus shrinking markets.[19]

These results support the suggestion that the preliminary results above are due to the balance of the theorized effects in Section 3.

In the appendices, I describe robustness tests performed to validate these analyses.

4 Discussion

One should be cautious in interpreting the above results, given that some of the novel features of this study come with limitations.

Empirically, the novel process of defining market geographic bounds is not a perfect representation of the desired product-level cross-price elasticity of demand. But, it is a major improvement over national-level computations of concentration.

In addition, the data used here represent a particular definition of automation capital. I find this definition to be more satisfying than the very specific category of "robot" for the question at hand, but the fact that automation equipment is not clearly defined philosophically means that empirical studies will have to make

as population growth in the relevant MSA. Results are consistent with the alternate approach.

[19]It is critical to distinguish markets from industries. Whether sales increase more for leading firms than for laggards is the critical dimension, and thus the relevant construct is the local market and not global production.

operational choices to represent the concept.

Finally, the novelty of the data means they will benefit from future verification. While I conduct several analyses to confirm validity in Appendix A.6.4, some questions remain. For example, the revenue measures in the NETS data have not received as much validation as the employment measures have. I provide some initial validation in Appendix A.6.4 that firm-level aggregate revenue patterns are highly correlated with accounting data from Compustat, but I look to researchers with access to independently collected revenue data for private companies to continue to investigate the validity of the NETS measure.

My simple model, like any model, abstracts some potentially relevant features of the setting. For example, this model does not address additional potential competitive factors—like potential network effects or strategic inter-temporal behavior or internal features of the firm such as individual incentives or organizational frictions—that could impact adoption. Because the model generates novel predictions that are supported by the data, I believe it changes thinking on the subject and is a valuable contribution to the discussion of marginal-cost-reducing investment and dominance by leading firms.

Acknowledging these limitations, the results are quite robust and have interesting implications for understanding the heterogeneous role of process technology improvements in market dominance.

A Appendices

A.1 Case study of endogenous automation increasing market parity

To make ideas concrete, I describe the particular case of custom cable manufacturers. A critical component of cables is the housing for the printed circuit board (PCB) in the connector at the end of the cable that routes the wire lines coming from one connector to the other and sometimes provide limited signal processing. Custom cable manufacturers have establishments, therefore, assigned to the the bare printed circuit boards industry NAICS code 334412. Figure 4 depicts a rendering of the PCB in a video graphics array (VGA-standard) monitor cable.

Figure 4: Rendering of circuit board in VGA cable from ISC Custom Cables and Electronic Inserts (ISC Custom Cables, 2019)

Custom housings for the PCBs in connectors can be made through a number of processes, including high–scale-economy processes like injection molding and urethane casting or lower–scale-economy processes like CNC machining or additive/3D printing. The market for noncommodity cables is largely regional because of the need for collaboration between vendors and customers, which allows the comparison of markets. Figure 5 compares three regional cable markets in the years around 2012—markets identified as A, B, and C for anonymity required by the data provider—using [measures depicting] only coarser Herfindahl-Hirschman Indices (HHI).

All three markets had 10 or fewer regional producers during this window and none had any measured investment in automation equipment in either 2010 or 2011. The largest producers in each market held substantial market share, as indicated the 2009 HHI depicted in Figure 5.

In 2012, the third-largest producer in market C, denoted c, invested nearly \$50,000 in a vertical injection molding machine, allowing it to produce specialized cables at lower average costs.

Establishment c's sales were higher in 2013 and 2014 and none of the other establishments in the market had dramatic sales increases. The end result is that establishment c closed the gap with the leader in market C, as indicated in Figure 5, while the less-concentrated markets actually got more concentrated.

With this example in mind, I move to the data used to study this phenomenon broadly in the US economy.

A.2 Data appendices

A.2.1 Description of matching procedure

The greatest challenge of working with these data is a process called "entity resolution." The PIERS data are not matched to entities and variations in spelling and shortening of names can make detecting that two observed names correspond to the same latent entity extremely difficult. Matching those latent entities to NETS

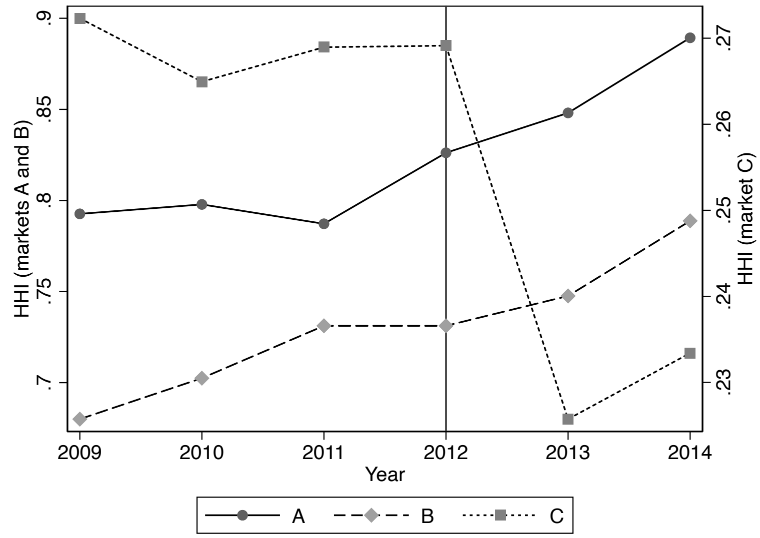

Figure 5: Case study of HHI in three comparable markets

establishments is a second challenge.

The technique I use for resolving entities in the current data is a machine learning algorithm from a class of algorithms introduced by Fellegi and Sunter (1969), but much improved upon since then. These algorithms are supervised learning techniques in which one first constructs the Cartesian product of records, yielding data comprising of the full set of candidate matches. One then calculates distance metrics between the candidates for each pair. A sample of the candidates are scored as matches or non-matches and an empirical model is estimated of the likelihood of a match as a function of the distance between the observations.[20] Those weights are applied to the distance measures in the unscored data to classify them by estimated probability of a match. For shipments with positive probability of match to multi-

[20]Steorts et al. (2019) provides an excellent overview of recent advances in techniques used to improve the computability of matches in large data samples.

ple NETS records, the maximum likelihood record is selected. Finally, a confidence cutoff is established and those above the cutoff are classified as matches.

In my implementation, I began by hand-correcting city and state names to match valid geographic identifiers from US TIGER data. This allows blocking the Cartesian product of records on city and state, without which the project would be computationally infeasible. Then, *firm name* from the NETS data, *recipient name* from the PIERS data, and addresses from both are used to compute distances. Two distances were computed: string distance between names and string distance between the street component of the address. The particular string distance is bigram distance with simple weight penalties assigned to bigrams based on their baseline prevalence in the block.[21] As above, empirical models included logit, random forest, lasso, and ridge regression. To allow for as much flexibility as possible, similarity scores were divided into 10 buckets each, with each bucket representing 10 percent (i.e., scores from .1 to .2). Let $name\_s\hat{i}m\_score$ represent a vector of indicator variables for each bucket of the name similarity score and similarly define $street\_s\hat{i}m\_score$ and $disam\_nam\hat{e}\_sim\_score$. The empirical model used to determine the cutoff is:

$$\begin{aligned} I[Match_i] = \alpha &+ \hat{\beta}_1(name\_s\hat{i}m\_score) + \hat{\beta}_2(street\widehat{\_sim}\_score) \\ &+ \hat{\beta}_3(name\widehat{\_sim}\_score \times street\_s\hat{i}m\_score) + \hat{\beta}_4(disam\_na\widehat{me\_}sim\_score) + \epsilon \end{aligned} \tag{14}$$

[21]String distance is computed as $\frac{1}{(min[s1,s2])}\sum_{m\in M}\frac{1}{f_m}$, where M is the set of shared grams between the two strings, s_i indicates the number of grams in the relevant string, and f_m represents the empirical frequency of the focal gram.

and the cutoff for the main results was set at probability of .7, though results are robust to, at least, between .6 and .95.

To further demonstrate robustness, Section A.6.6, discusses the results of simulating alternate placebo matching techniques and shows that the results of the present manuscript are unlikely to be obtained from alternate matches.

A.2.2 "Automation" versus "Robot"

In the introduction, I note some of the concurrent studies studying automation, all but this one studying questions related to labor. Most of those studies investigate a narrower class of capital: "robots". In this section I compare the firms which from the present data set measured as having ordered automation capital based on the full set of stems described in Section 2.1 appearing in the commodity description to those classified exclusively by the stem of the word "robot".

This comparison serves to demonstrate the value of a broader definition of automation capital by showing how many more firms automate, generally, than those who specifically use robotics. The comparison also shows two biases from using measures based only on robots:

1. Among some classes of firms, some of those who automated are marked as not having done so because they didn't order robots, biasing results toward zero,

2. Across the sample, those having ordered robots have different product market breadth and sales volumes, which could result in measured effects associated with robots being biased away from true automation in either direction, depending whether larger or smaller firms have greater effects.

As noted in the introduction, robots represent a small fraction of orders: 1.8% of automation shipments, as classified by the algorithm described in Section 2.1, and 3.1% of the dollar value. Figure 6 depicts the kernel density of firm-level[22] average yearly sales. Because sales values are extremely skewed, the measure is winsorized at the top 5% and presented on a log scale. Note that the dataset constructed for robot orders is constructed of firms with higher sales, with both robot-ordering and non-robot-order firms having higher sales than firms ordering general automation capital. This accords with the footnote in Section 2.1.2 above noting that because robots are generally re-programmable—serving more varied purposes, and also require programming, they tend to be ordered by larger firms with more wider product lines.

The bias of robots is further corroborated by comparing the employment of firms ordering robots to those ordering automation capital generally.

Figure 7 shows, again, that firms in industries which had ordered robots are larger, with those ordering robots having dramatically higher employment (indicated by a right-ward shift of the kernel density).

Firms ordering robots also tend to be in industries with different catchment areas. Figure 8 shows that firms making no orders in both datasets have roughly similar catchment areas. Those ordering, however, are much less likely to be in local markets, and much more likely to be in regional.

The fact that roughly half of the firms depicted in Figure 8 are in national markets suggests that national measures of automation, including those in the IFR (used, for

[22]Firm-level measures are computed by aggregating establishment-level measures to the DUNS number of the headquarters reported in the NETS data

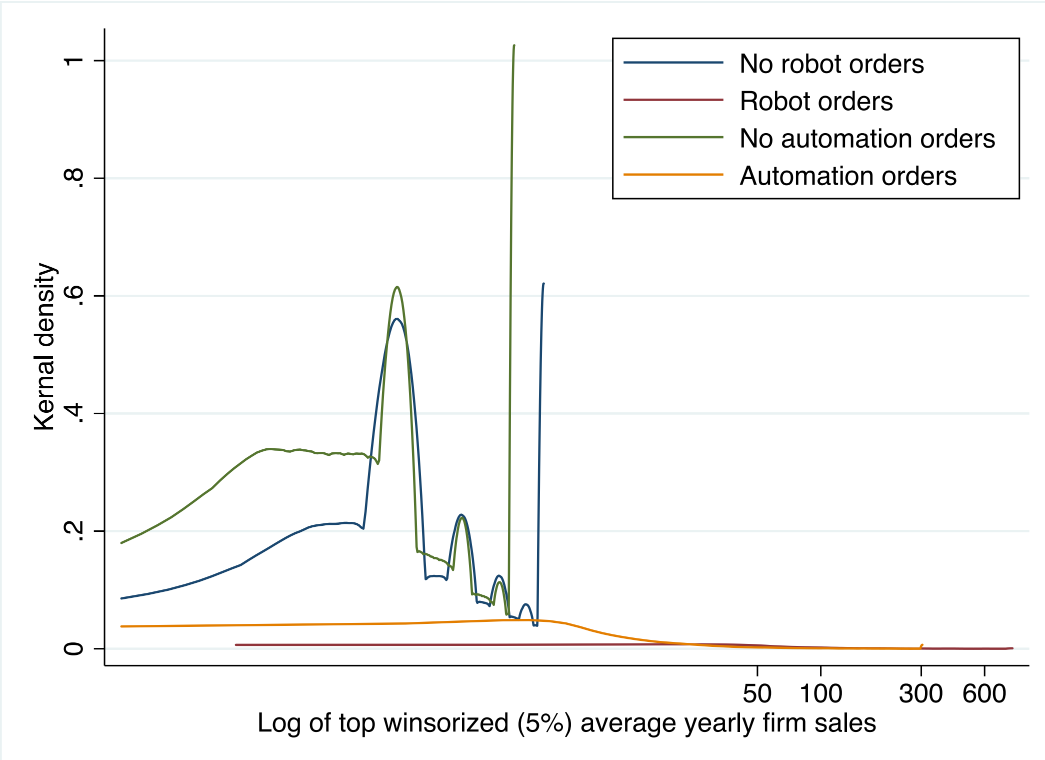

Figure 6: Differences in average yearly sales volume between firms ordering automation capital and firms ordering robotics

example, in Acemoglu and Restrepo (2018) and Graetz and Michaels (2018)) would be poor representations of product market competition, and furthermore, that they would be differentially bad for automating and non-automating firms, and even for those automating with robots versus more general automation capital. This fact and the broader base of automation capital captured by my measure illustrate the utility of the innovations in this manuscript.

A.2.3 Included industries

Table A1 shows the six-digit NAICS industries represented in the merged data.

Table A1: Included industries (NAICS six-digit)

NAICS 6-digit	Estabs	Total auto. cap. inv. ($k)	Unique markets	NAICS 6-digit	Estabs	Total auto. cap. inv. ($k)	Unique markets
311211	652	197995.8	147	332431	615	34829.1	111
311511	1160	8130.99	203	332510	3611	48257.09	276
311612	2262	20875.36	272	332618	4093	219374.5	340
311811	3613	3615.27	358	332710	52283	498749.1	862
311812	10444	103774.6	570	332721	1986	32759.99	175
311991	672	1889.93	117	332722	1937	73349.37	189
311999	5142	31204.42	360	332811	1182	14485.03	131
312111	3812	52203.01	419	332812	5832	55382.22	390
312113	622	3641.45	158	332813	12127	188174	482
321113	4538	6024.51	1050	332911	1452	136548.5	170
321114	1150	26643.87	222	332999	9047	47587.38	498
321214	644	679.34	152	333511	4153	152838.2	264
321911	23514	26917.62	1312	333515	4046	827599.3	304
321912	1055	6819.76	246	333911	2301	164150.2	247
321918	6199	38783.86	526	333994	998	9586.7	145
321920	8095	23315.71	761	333999	9725	412353.2	480
321992	3074	1997.08	503	334220	9831	56511.64	431
321999	14335	17991.85	1064	334290	2699	5516.49	239
322211	2518	152882	388	334412	3839	59877.26	202
322212	368	8994.65	74	334416	1254	61905.29	151
322299	2070	71226.65	196	334418	264	3062.76	45
323113	15535	93192.79	848	334419	8423	228637	355
323117	1049	37330.48	141	334513	4228	153466.7	283
324121	2828	3395.17	351	334515	5193	42140.12	270
324122	958	2336.46	133	335313	1791	106200.4	182
325199	5088	28876.16	375	335314	4071	88369.28	285
325314	690	38	151	335929	413	13822.59	90
325411	2299	17404.4	208	335931	1336	37330.46	161
325510	4302	14832.95	314	335999	13731	148585.7	551
325520	2862	39386.33	219	336211	2514	44116.78	322
325910	1347	13197.18	139	336360	1423	28122.07	190
325991	226	8942.93	63	336370	1367	208358.9	138
326111	1121	48762.86	131	336413	4060	405333.4	268
326113	1685	126224.3	187	337110	22606	72293.02	1191
326122	478	13015.45	119	337127	5788	29042.36	394
326140	1182	23129.8	165	337215	7156	49443.71	486
326150	1281	125028.7	175	337910	1537	16519.54	184

Table A1: Included Industries (NAICS 6-digit) [continued]

NAICS 6-digit	Estabs	Total auto. cap. inv. ($k)	Unique markets	NAICS 6-digit	Estabs	Total auto. cap. inv. ($k)	Unique markets
326160	384	51028.23	87	337920	3738	2711.54	252
326191	583	23346.34	112	339112	8720	43704.51	347
326199	21314	881095.8	981	339113	12224	4289134	576
326299	2479	102423.8	249	339920	22520	64051.34	1127
327215	6605	55024.23	419	339950	61875	65861.35	871
327310	990	45472.58	153	339991	1201	25987.42	153
327320	10546	34992.24	812	339999	93650	468049.7	2408
327331	1790	17603.98	408	511110	28377	73800.32	882
327332	217	9646.96	69	511120	31104	130661.5	785
327390	10496	67902.67	755	511140	4502	4490.51	321
327420	4073	2561.63	278	511199	69869	40087.04	1457
327991	4588	36945.39	394	517110	17092	30083.78	819
327999	524	8678.18	94	517210	10939	5653.73	747
331222	202	101.84	48	541219	767	5812.07	186
332111	1863	139163.4	221	541310	447	744.04	123
332311	3228	20340.31	432	541380	2419	25440.31	388
332312	13351	110629.3	885	541712	5018	86507.83	283
332313	2434	60163.53	266	541860	1007	14106.27	132
332321	4054	18848.59	323	541910	903	1081.21	102
332322	11758	133735.3	742	561450	246	1943.64	55
332323	6799	22602.77	362	561492	27	9.75	12

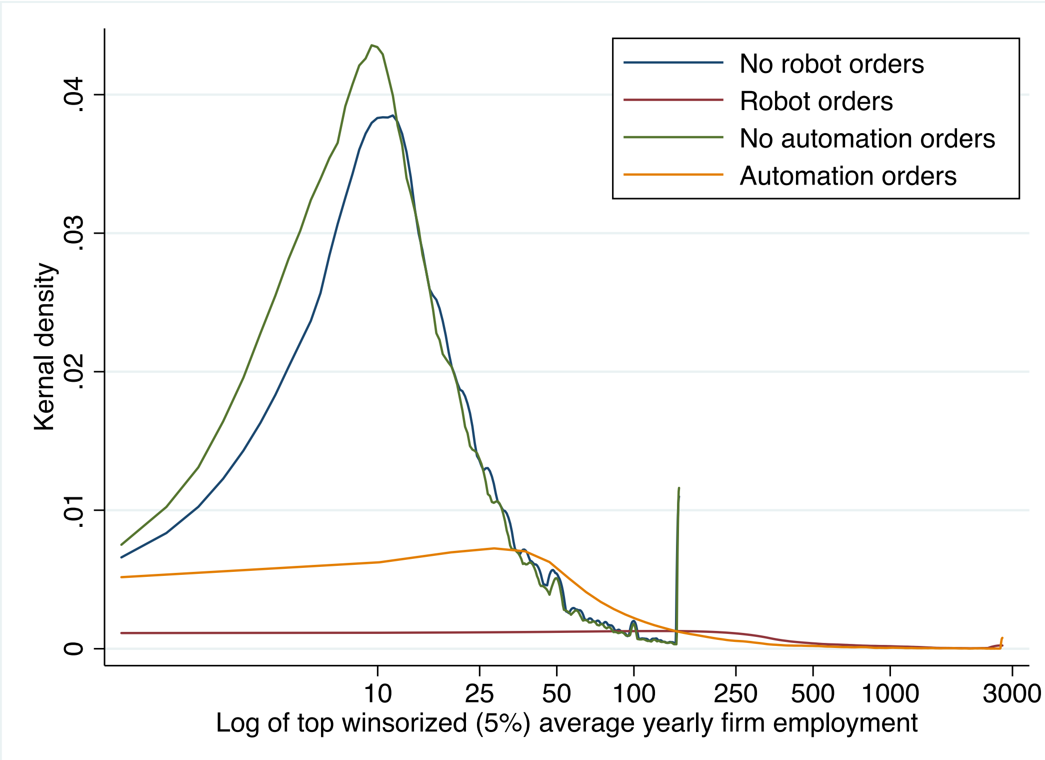

Figure 7: Differences in average yearly employment between firms ordering automation capital and firms ordering robotics

A.2.4 Categories of NETS firms removed before matching

Figure 9 presents the regular expressions used to remove trading companies, distributors, and shippers who may not be using purchases themselves. Universities are also removed because manufacturing production is not their central business model.

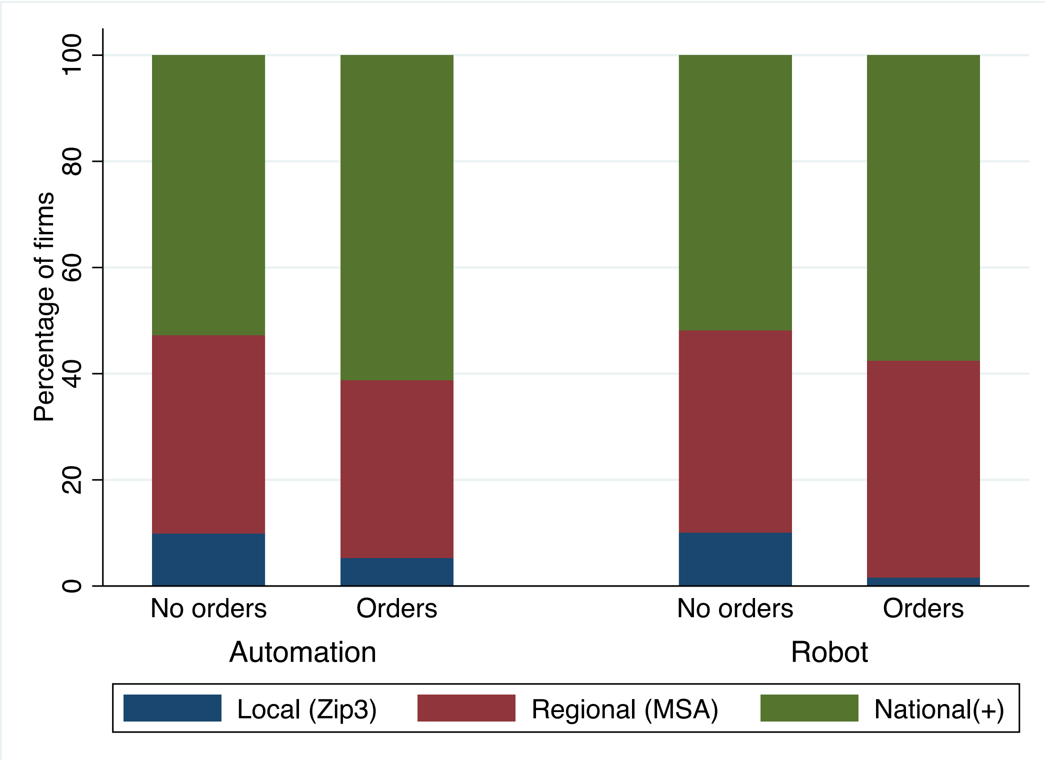

Figure 8: Differences in catchment areas of firms ordering robots versus those ordering more general automation

A.3 Analysis appendices

A.3.1 Market-level parity changes the implications of automation

The implication of the above establishment-level results is that markets in which automation is available will become less concentrated as a result of strategic investments by market participants. Table A2 presents the results of analyses at the market level. Columns 1 and 2 present the results of a regression of market-level HHI on lagged HHI with the sample split by whether any automation investment was made in the period. Both models include fixed effects at the market and NAICS6-by-year-

```
drop if regexm(name,"DI?ST(RIBUT)?(ION|ER|OR)?S?")
drop if regexm(name,"IMPO(RT)?(ER)?S?")
drop if regexm(name,"EXPO(RT)?(ER)?S?")
drop if regexm(name,"(AIR )?EXP(RESS)?(\sLINE)?R?S?")
drop if regexm(name,"FO?R?WA?R?D(I?N?G|E?RS?)?")
drop if regexm(name,"LOGIST?I?C?S?")
drop if regexm(name,"INTE?R?CONTINENTAL")
drop if regexm(name,"TRA?NSP?O?R?T?A?T?I?O?N?")
drop if regexm(name,"SH[I]?P(PIN)?G( LINE)?[S]?")
drop if regexm(name,"CO?NTA?I?NE?R( ?LINE)?S?")
drop if regexm(name,"(AIR )?FRE?I?G?H?T\s?(LINE)?R?S?")
drop if regexm(name,"TRA?DI?N?G")
drop if regexm(name,"CU?STO?MS?( BRO?KE?RS?)?")
drop if regexm(name,"BRO?KE?R(S|A?GE)?")
drop if regexm(name,"CONSOLIDAT(O|E)RS?")
drop if regexm(name,"TRA?DI?N?G?")
drop if regexm(name,"EXPEDIT(O|E)RS?")
drop if regexm(name,"CA?RGO")
drop if regexm(name,"(INTER)?MARINE")
drop if regexm(name,"((AIR ?(AND|&)? ?SEA)|(SEA ?(AND|&)? ?AIR))")
drop if regexm(name,"INTERMODAL")
drop if regexm(name,"MARITIME")
drop if regexm(name,"AD[UA]*NA(S|AL)?E?S?")
drop if ustrregexm(name,"UNIV(\b|E?RSI?TY)?")
drop if strmatch(name,"PANALPINA")
drop if regexm(name,"OVERSEAS")
drop if regexm(name,"OCEAN SERVICE")
drop if regexm(name,"TRUCKING")
```

Figure 9: Regular expressions used to remove NETS firms before matching

levels. Errors are allowed to cluster at the market level.

As above, HHI is instrumented for using the HHIs of other markets in the same NAICS6-by-year. I use a split sample because the desired comparison is markets able to make their equilibrium investment versus those unable to make their investment, rather than markets with different levels of investment. The differences between the estimated coefficients on lagged HHI are different at levels beyond 0.26 percent suggesting that concentration is less persistent in automating industries.

Columns 3 and 4 present the results of an alternate estimation strategy: HHI regressed on lagged HHI instrumented by the second lag using the dynamic panel data estimator detailed in Blundell and Bond (1998), which addresses downward bias in the original Arellano-Bond estimator (Arellano and Bond, 1991). For ease of interpretation and computation, NAICS6-by-year and market-level fixed effects are accounted for by demeaning the dependent variable.

Column 5 shows the results of a regression including all markets with automation capital stock and HHI interacted. Automation is instrumented with the exchange rate between the US and the relevant capital-producing country as in Section A.6.2.[23]

As in the other columns, the results suggest that there is a secular trend toward concentration, but that automation actually works against that trend in all but the most concentrated markets. For the most concentrated markets, the aggregate effect is indistinguishable from zero at conventional levels. That accords with the model predictions that automation will reduce concentration for markets that were not already extremely concentrated.

[23]Exchange rates have been used as exogenous price shifters in several studies, including Bertrand (2004).

Table A2: Market-level effect of competitive asymmetry on automation investment

VARIABLES	(1) HHI	(2) HHI	(3) HHI	(4) HHI	(5) HHI
Demeaned HHI [lag]			0.696***	0.737***	
			(0.0117)	(0.00153)	
HHI [lag]	0.571***	0.669***			1.229***
	(0.0319)	(0.00359)			(0.205)
Ln(1+Automation capital stock) [lag]					-0.0451
					(0.0282)
HHI × Ln(1+Automation capital stock) [lag]					-1.165***
					(0.406)
Ln(Market sales) [lag]	-0.000510	-6.41e-06	-0.0262***	0.0453***	
	(0.00549)	(0.000645)	(0.00330)	(0.000807)	
Establishments [lag]	-2.97e-05**	-3.89e-05**	0.000111**	0.000331***	
	(1.30e-05)	(1.62e-05)	(4.72e-05)	(4.65e-05)	
Observations	8,129	352,180	10,305	353,481	362,244
Model	2SLS	2SLS	ABBB	ABBB	2SLS
Market FE	Yes	Yes			Yes
NAICS6 × Year FE	Yes	Yes			
Sample	Automated	Non-automated	Automated	Non-automated	All
Markets	Non-national	Non-national	Non-national	Non-national	All
Number of markets			3,920	28,472	
First-stage F	408.97	4671.06			73.00
NAICS6 × Year FE					Yes
NAICS6 × Year demeaned			Yes	Yes	
Market demeaned			Yes	Yes	

Errors in Columns 1 and 2 are allowed to cluster at the NAICS-by-year level. *** p<0.01, ** p<0.05, * p<0.1. Coefficients on lagged HHI in Columns 1 and 2 are different at<.26% level. Coefficients in Columns 3 and 4 are different at the .06% level

Panel 1 of Figure 10 shows the results of a regression predicting the time t HHI of a market as a function of its time $t-1$ HHI depending on whether there is automation capital expenditure in the period. Panel 2 shows the gap between the two. As described above, the results are consistent with the suggestion that automating markets decrease their concentration by more when they were previously more concentrated.

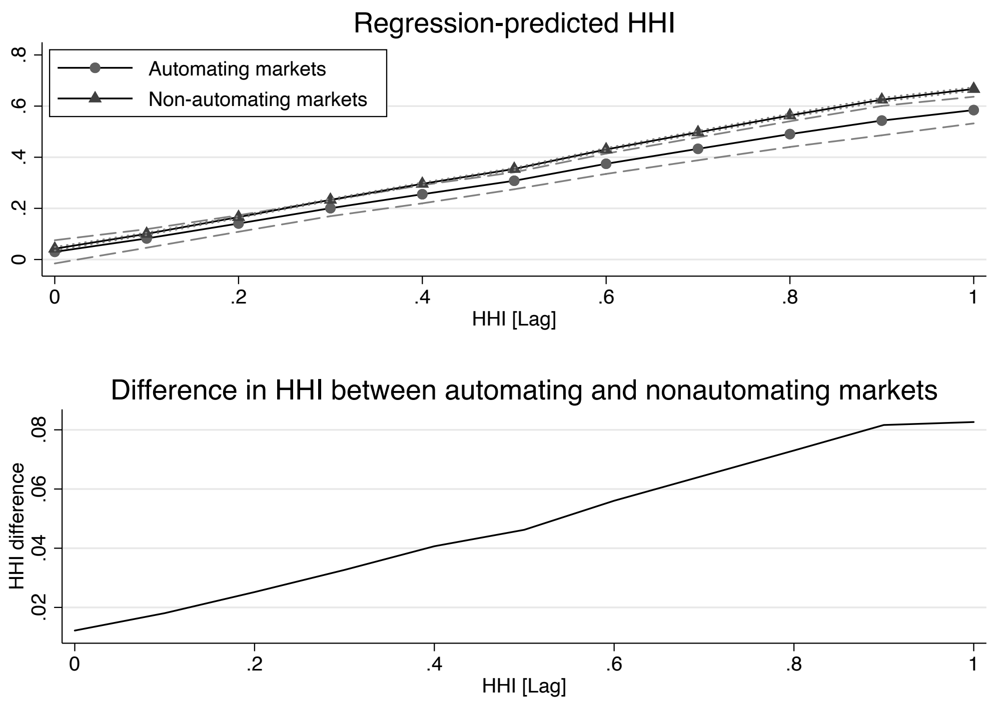

Figure 10: Robustness of results to alternate depreciation rates for computing stocks

A.4 Basic OLS and first-stage regressions for instrumented regression

Table A3 presents the basic OLS versions of instrumented regressions in Table 2.

Table A3: Basic OLS versions of instrumented regressions in Table 2

	(1)	(2)
VARIABLES	Ln(1+ Automation capital stock value)	
Revenue gap [lag]	0.0147	0.00348
	(0.0103)	(0.00521)
Leader × Revenue gap [lag]	0.00671	0.00203
	(0.0160)	(0.00836)
Leader [lag]	0.0202**	0.00601
	(0.00918)	(0.00505)
Ln(1+Sales) [lag]	0.0153***	0.00679**
	(0.00579)	(0.00285)
Ln(1+Employees) [lag]	0.0700***	0.0125***
	(0.00667)	(0.00336)
Ln(1+ Automation capital stock value) [lag]		0.842***
		(0.00274)
Constant	-0.0488**	0.00466
	(0.0225)	(0.0113)
Observations	1,358,765	1,358,765
R^2	0.633	0.880
Model	OLS	OLS
Establishment FE	Yes	Yes
NAICS6 × Year FE	Yes	Yes
Markets	Subnational	Subnational

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Errors clustered at market-by-year level.

Table A4 presents the first stages of the 2SLS instrumented regressions in Table 2.

Table A4: First stages of instrumented regressions in Table 2

VARIABLES	(1) Revenue gap	(2) Revenue gap × Leader	(3) Revenue gap	(4) Revenue gap × Leader
Mean HHI in other industry markets	-142.5***	-82.39***	-142.5***	-82.39***
	(1.092)	(0.731)	(1.092)	(0.731)
Mean HHI in other industry markets × Leader	-0.118***	-0.0387***	-0.118***	-0.0387***
	(0.0104)	(0.0102)	(0.0104)	(0.0102)
Leader [lag]	0.0534***	0.416***	0.0534***	0.416***
	(0.00514)	(0.00499)	(0.00514)	(0.00499)
Ln(1+Sales) [lag]	-0.130***	0.0336***	-0.130***	0.0336***
	(0.000888)	(0.000477)	(0.000888)	(0.000477)
Ln(1+Employees) [lag]	0.0104***	0.00176***	0.0104***	0.00176***
	(0.000961)	(0.000562)	(0.000961)	(0.000562)
Ln(1+ Automation capital stock value) [lag]			0.000168	8.16e-05
			(0.000135)	(8.90e-05)
Constant	67.77***	38.65***	67.77***	38.65***
	(0.512)	(0.343)	(0.512)	(0.343)
Observations	1,358,765	1,358,765	1,358,765	1,358,765
R^2	0.799	0.928	0.799	0.928
Model	OLS	OLS	OLS	OLS
Establishment FE	Yes	Yes	Yes	Yes
NAICS6 × Year FE	Yes	Yes	Yes	Yes
Markets	Subnational	Subnational	Subnational	Subnational

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Errors clustered at market-by-year level.

Table A3 presents the basic OLS versions of instrumented regressions in Table

A8.

Table A5: Basic OLS versions of instrumented regressions in Table A8

VARIABLES	(1)	(2)
	Ln(*Sales*)	
Ln(1+Auto. capital stock value)	0.00123***	0.00196***
	(0.000294)	(0.000471)
Ln(1+Sales) [lag]	0.632***	0.632***
	(0.00299)	(0.00299)
Ln(1+Employees) [lag]	0.108***	0.108***
	(0.00332)	(0.00332)
Ln(1+Auto. capital stock value) [lag]		-0.000926*
		(0.000477)
Constant	0.177***	0.177***
	(0.00944)	(0.00944)
Observations	1,276,403	1,276,403
R^2	0.959	0.959
Model	OLS	OLS
Establishment FE	Yes	Yes
Market × Year FE	Yes	Yes
Markets	Subnational	Subnational

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Errors clustered at market-by-year level.

Table A4 presents the first stages of the 2SLS instrumented regressions in Table A6.

A.5 Geographic representation of automation investment patterns

Figure 11 shows the geographic distribution of the matched data.

Table A6: First stages of instrumented regressions in Table A8

VARIABLES	(1) Ln(1+Auto. capital stock value)	(2)
Imputed relevant exchange rate	-0.00209***	-0.000554**
	(0.000463)	(0.000232)
Ln(1+sales) [lag]	0.0345***	0.00877***
	(0.00615)	(0.00304)
Ln(1+employees) [lag]	0.0457***	0.0105**
	(0.00851)	(0.00439)
Ln(1+Auto. capital stock value) [lag]		0.838***
		(0.00301)
Constant	0.210***	0.0628**
	(0.0591)	(0.0301)
Observations	1,276,403	1,276,403
R^2	0.708	0.902
Model	OLS	OLS
Establishment FE	Yes	Yes
Market × Year FE	Yes	Yes
Markets	Subnational	Subnational

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Errors clustered at market-by-year level.

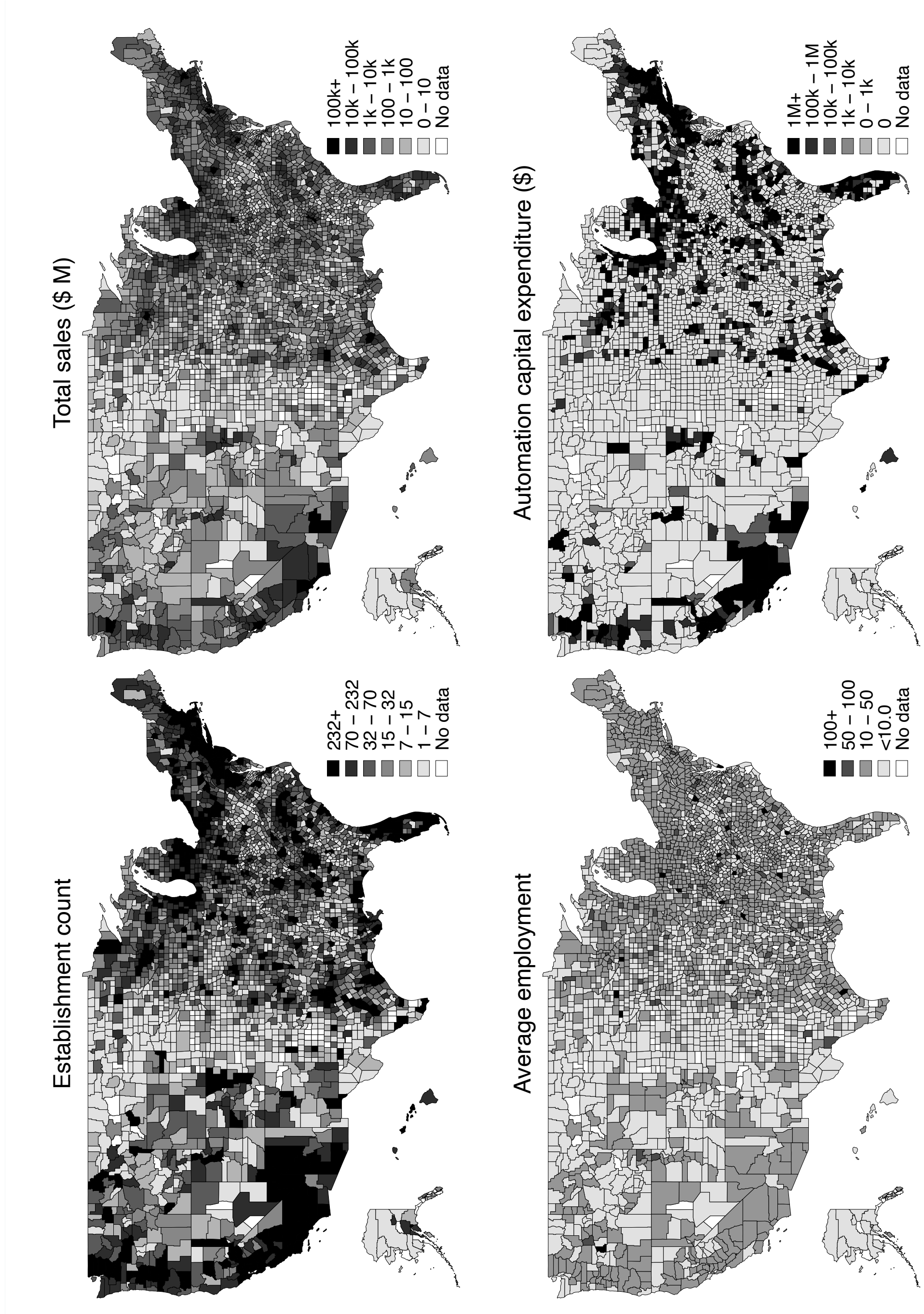

Figure 11: Geographic distribution of matched data

A.6 Robustness appendices

A.6.1 Robustness to using systems dynamic panel data (DPD) estimator instead of 2SLS

In Section 2.2 above, I note that there are two approaches to dealing with the Nickell (1981) bias, both of which rely on instrumenting for the lagged dependent variable. The results in Table 2 are derived with my preferred instrument, but here I show the robustness of those results to the systems dynamic panel data estimator introduced by Blundell and Bond (1998) as an improvement to the one introduced by Arellano and Bond (1991).

To make these estimates comparable to the estimates in Section 2.2, I demean the dependent variable by levels at which which Table 2 has fixed effects.

Table A7 shows the results indicating the expected positive autocorrelation between automation capital investment and, consistent with the above results:

1. A positive main effect on investment by leading firms.
2. A positive main effect on the gap between leading and lagging firms.
3. A negative interaction on the indicator for leadership and the gap of sufficient magnitude to make the net effect of leadership decrease with the gap.

These results add confidence that the above results are not driven by the particular choice of estimator.

Table A7: Systems dynamic panel data estimator replication of main result in Table 2

VARIABLES	(1) $Ln(Automation\_capital\_stock\_value + 1)$
$Ln(Automation\_capital\_stock\_value_{t-1})$	0.898***
	(0.000567)
Ln(1+Sales) [lag]	0.0194***
	(0.00514)
Ln(1+Employees) [lag]	0.128***
	(0.00688)
Revenue gap [lag]	0.0897***
	(0.0125)
Leader × Revenue gap [lag]	-0.0991***
	(0.0152)
Leader [lag]	0.0118***
	(0.00415)
Constant	-0.580***
	(0.0271)
Observations	1,375,481
Unique establishments	147,584
Model	ABBB
Establishment FE	Yes
NAICS6 × Year FE	Yes
Markets	Subnational

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors computed using the finite sample correction from Windmeijer (2005).

A.6.2 Validating that exogenous investment increases sales

A fundamental component of the model is that if an establishment exogenously invests in automation capital, its marginal cost of production decreases, making sales increase. In this subsection, I demonstrate patterns consistent with that assumption. Isolating exogenous investment from equilibrium investment requires instrumenting for the decision to invest.

Automation is instrumented for using the likely exchange rate faced by establishments if they were to order. Because country of origin is logically only available for establishments that have ordered, I first regress exchange rate on logged sales, logged employees, and indicators for NAICS6 industry and year for the sample of establishments that did order. I then predict that exchange rate for the entire sample. This effectively gives the expected exchange rate that would be faced by an establishment of a given size—measured by sales or employment—in a given industry in a given year for the product it is likely to buy.

The results of the analysis appear in Table A8, which demonstrates that, as predicted, exogenous investment in automation capital increases revenue.

These reports support the assumption of automation increasing sales absent a competitive response.

A.6.3 Robustness of comparison of results for growing and nongrowing markets

Figure 3 above shows a pattern consistent with the prediction of the model in Section 3 that in growing markets the cost-spreading effect will play a larger role and markets

Table A8: Causal impact of automation on sales

	(1)	(2)
VARIABLES	*Ln*(*Sales*)	
Ln(1+Auto. capital stock value)	5.562***	20.99**
	(1.247)	(8.814)
Ln(1+Auto. capital stock value) [lag]		-17.58**
		(7.386)
Ln(1+Sales) [lag]	0.498***	0.505***
	(0.0448)	(0.0821)
Ln(1+Employees) [lag]	-0.273***	-0.238
	(0.0921)	(0.162)
Observations	1,276,403	1,276,403
Model	2SLS	2SLS
Establishment FE	Yes	Yes
Market × Year FE	Yes	Yes
Markets	Subnational	Subnational
Instrument	Relevant exchange rate	Relevant exchange rate
Cragg-Donald Wald F	20.39	5.70

Note: Errors clustered at market-by-year level. Using the critical values from Stock and Yogo (2005), the instruments are strong enough that the Wald would reject at less than 10 percent if the true power were 5 percent. *** p<0.01, ** p<0.05, * p<0.1

that are more zero-sum market-stealing can play a greater role. Market growth for that figure was computed base on total market sales growth year-over-year. I prefer that specification because it allows different growth levels across industries within geography. One might be concerned, however, that market growth is driven by automation. To verify that is not driving the observed patterns, Figure 12 shows an analogous graph, but with growth computed at the geography level and determined by population growth in the relevant MSA.

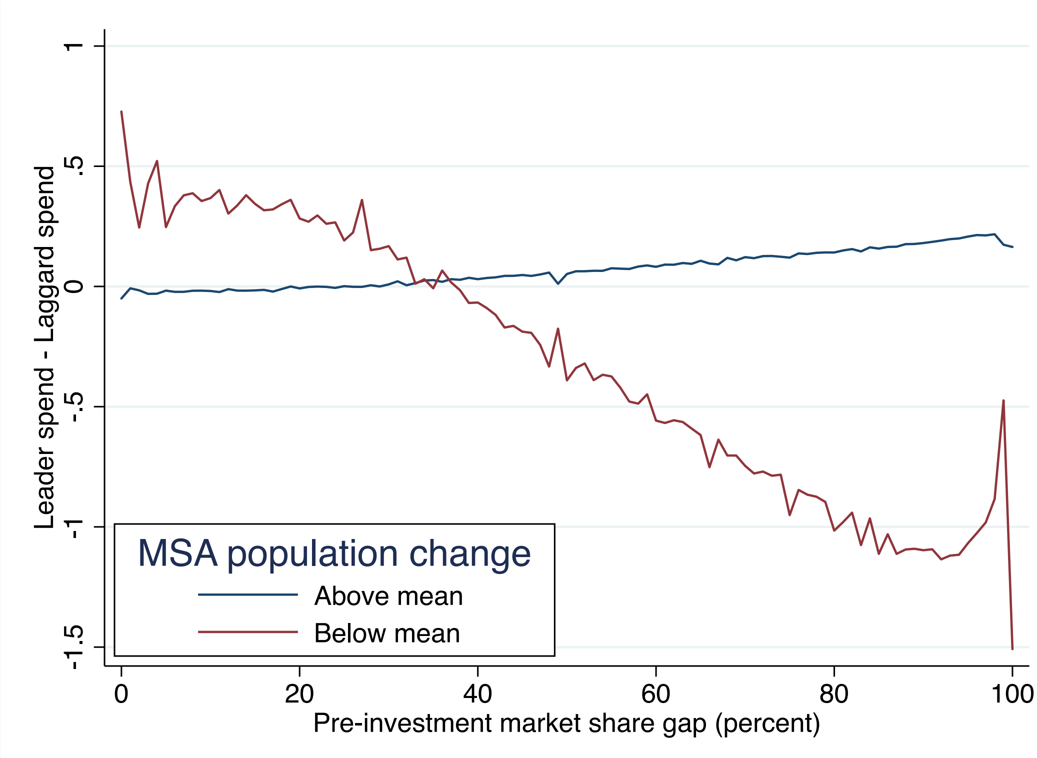

Figure 12: Difference between laggard and leader spending as a function of pre-investment *gap*: Growing and nongrowing markets
Note: MSA population growth data are available from Census for years 2010–2016. Mean MSA population growth is .4 percent for that period.

A.6.4 Validating the components of the data

In this section, I take several approaches to validating the PIERS and NETS data. For NETS, I perform a preliminary validation of the revenue measures by matching aggregate NETS revenue for public firms to their accounting data in Compustat and comparing the patterns. For PIERS, the first approach I take is to compare the measure to independently collected measures of similar constructs. The second is to demonstrate robustness of the above results to other PIERS-based measures derived from other classifiers.

Comparing the NETS-based revenue measure to independently collected measures While studies have compared NETS to employment data (Haltiwanger, Jarmin and Miranda, 2013; Barnatchez, Crane and Decker, 2017), decidedly less attention has been paid to revenue data. While the revenue data has been used in industry to assess the credit-worthiness of businesses and relied upon by scholars (Rossi-Hansberg, Sarte and Trachter, 2020) whose results have been replicated in Census data (Rinz, 2018; Handwerker and Dey, 2019), comparing the data themselves to other sources has not been done to date. While comparable data at the establishment level for comparison exist only from Census and are not available to all researchers, I use a novel sample of public firm data from Compustat hand-matched to NETS to compare the data. NETS data are aggregated to the headquarters DUNS number, which is then matched with Compustat data.

In Table A9, I correlate the aggregated NETS establishment revenues at the firm-by-year level for years 1989–2015 to the Compustat *Net Income*, designated `nis` in

Table A9: Correlation between COMPUSTAT revenue and NETS sales for matched sample

	[1]	[2]
[1] NETS aggregated sales	1	
[2] COMPUSTAT Net Income (*nis*)	0.9133***	1
Observations	725	

Compustat.

While levels may differ considerably for many reasons, including international revenue for public firms in Compustat, the correlation is sufficiently high to add confidence in changes in the NETS revenue data.

Comparing the PIERS-based measure to independently collected measures

International Federation of Robotics (IFR) data The primary dataset that has been used by economists (e.g., Graetz and Michaels, 2018) to study automation comes from the International Federation of Robotics (IFR). As those data are only intended to represent "robots", they capture a much narrower phenomenon than the "automation capital" described here. The IFR data, furthermore, are presented at the country-by-year level, not allowing for identification of even market-level effects, let alone the establishment-level phenomena studied here. Nevertheless, it is interesting to compare the time trends between the IFR data and the novel PIERS-based measure introduced here to corroborate validity.

Table A10 presents the pairwise correlations between the stock and flow measures

Table A10: Correlation between country-by-year IFR data and PIERS-based measure

	(1)	(2)	(3)	(4)
(1) Mean automation capital stock	1			
(2) Total automation capital stock	0.997***	1		
(3) Total automation capital flow	0.834***	0.852***	1	
(4) Mean automation capital flow	0.846***	0.858***	0.997***	1
(5) US robot density (IFR)	0.956***	0.946***	0.804***	0.820***

derived from PIERS and the US robot density measure from IFR. The data are very highly correlated, suggesting that IFR data might be a reasonable proxy for aggregate automation and that the PIERS-based measure has some external validity.

Industry-level elasticity of substitution (σ) The recognition of the substitutability of labor for capital has been important at least since Hicks (1932) and was formalized in the production function by Arrow et al. (1961) as σ. While much of the literature on σ has focused on aggregate economy-level substitution (summarized in Chirinko, 2008), some work has recently focused on estimating industry-level differences in that substitutability.

Theoretically, one might think of σ as representing the latent "automatability" of labor tasks in a production model, suggesting that investment in automation equipment should be greater in higher-substitutability industries. This proposition provides an opportunity to further validate the PIERS measure. In Figure 13, I show the correspondence between the industry-level measure of σ from Chirinko and Mallick (2017) and the PIERS automation capital measure.

To generate the figure, I regress market-by-year–level automation capital stock on four-digit-NAICS–level σ estimates, controls for market-by-year sales, market-by-

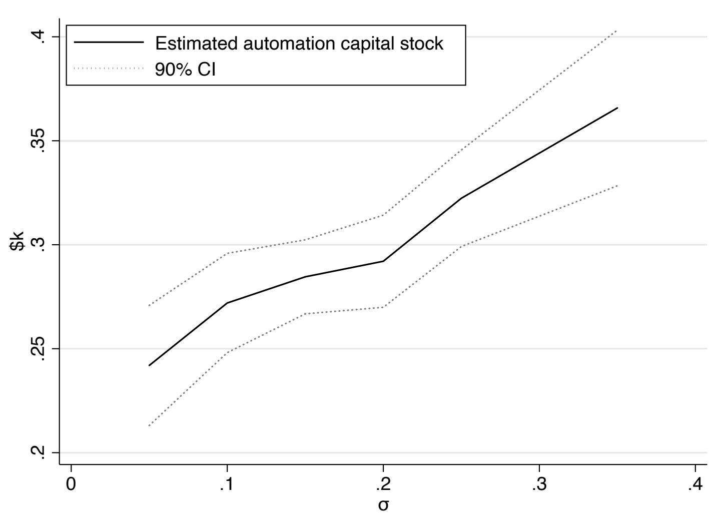

Figure 13: Validating the PIERS measure by correlating automation capital stock to industry-level σ from Chirinko and Mallick (2017)

year HHI, and fixed effects at the MSA and year levels. I allow errors to cluster within market. I then predict the estimated capital stock and collapse the data into buckets of σ by .05. Figure 13 thus shows regression-smoothed relationship between the values.

Because Chirinko and Mallick (2017) estimate σ using factor share shifts from KLEMS data, this is a completely independent measure of the phenomenon, so the correspondence between the two values adds support for the measure.

Validating against reported machinery capital expenditure in US Census Quarterly Survey of Plant Capacity (QPC) Another independently collected measure of investment in automation can be derived from the US Census Quarterly Survey of Plant Capacity. Beginning in 2008 (with the exception of 2012), respondents were asked to select from a list of candidate reasons why they had excess plant capacity. Those reasons include depressed demand as well as "Machinery Capital Expenditures." Publicly available data are available at the industry-by-quarter level, so I was able to aggregate the percentage of respondents who noted "Machinery Capital Expenditures" to the industry-by-year level and correlate that with the industry-by-year value of automation capital shipments derived from the PIERS-based measure.

Table A11: Correlation between NAICS3-by-year reported machinery capital expenditure in US Census Quarterly Survey of Plant Capacity and PIERS-based measure

	Value of automation capital shipments in year	
QPC respondent % with "Machinery Capital Expenditures"	Correlation	p-value
(1) Sum of quarters	0.291***	0.3%
(2) Mean of quarters	0.274***	0.5%

Note: The relevant question was not asked in all years, so years included are 2008–2011 and 2013–2014. 2008–2001 data were requested through a Freedom of Information request from Census in March 2020. Data provided are at the level of the fiscal quarter, so Rows 1 and 2 provide different within-year aggregations.

Table A11 shows that the correlations between the QSPC-based measure and the PIERS-based measure are strong and significant. Because the QSPC measure includes all machinery expenditure and not only maritime shipments, it adds confidence in the PIERS-based measure.

Validating investment measure against public company accounting measures from Compustat As a third confirmatory check of the viability of the PIERS-based measure, I compare it to self-reported accounting measures for public firms in Compustat. To do so, I first take a random sample of 300 firms present in Compustat listed on American exchanges during the time period and classified as being in manufacturing industries (NAICS 31–33). Of those, I found 200 with corresponding DUNS numbers listed in the Mergent Intellect Database. I summed all establishment-year PIERS records to the matching headquarters DUNS number and compare my PIERS-based automation capital flow and stock values with the two accounting line items that would represent automation capital: "Capital Expenditure" and "Property, Plant, and Equipment — Machinery and Equipment at Cost." Table A12 shows raw correlations between the accounting measures and the PIERS-based measure.

Table A12: Raw correlations between PIERS-based measure and Compustat accounting measures

	[1]	[2]	[3]	[4]
[1] Automation Capital Flow	1			
[2] Automation Capital Stock	0.653***	1		
[3] PPE — Machinery and Equipment at Cost	0.167***	0.302***	1	
[4] Capital Expenditures	0.159***	0.243***	0.879***	1
Observations		606		

Table A13 shows conditional correlations between the accounting measures and the PIERS-based measure removing firm and year fixed effects.

The strength of the associations on such limited data suggests support for the PIERS-based measures, including at the firm level.

Robustness of results to accounting for industry dollar percentage of equipment captured in PIERS A final potential concern about the PIERS-based measure is that, because it only captures equipment purchased from overseas,

Table A13: Conditional correlations between PIERS-based measure and Compustat accounting measures

	(1)	(2)	(3)	(4)
VARIABLES	Capital Expenditures		Property, Plant, and Equipment: Machinery and Equipment at Cost	
Automation capital (flow)	5.27e-05***		0.000228*	
	(1.92e-05)		(0.000123)	
Automation capital (stock)		5.46e-05***		0.000380***
Constant	252.4***	237.4***	1,840***	1,722***
	(8.697)	(9.308)	(58.06)	(61.05)
Observations	593	593	524	524
R^2	0.788	0.794	0.780	0.792
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Note: Sample constructed by (a) selecting a random sample of Compustat firms, (b) matching those to DUNS numbers at the Central Index Key (CIK) level, (c) matching those to headquarters DUNS numbers in NETS, and (d) collapsing establishment-level automation measures to the headquarters. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

it may provide a biased estimate of the behavior of establishments in industries that can source equipment domestically.

To verify whether this is the case, I begin by collecting industry-level data on capital expenditure from the US Census' Annual Survey of Manufacturers (ASM). For the years 2006–2014 (excepting 2007 and 2012), the Census broke out and made available in the results of instrument 31GS101 the average self-reported "Capital expenditures on all other machinery and equipment." For each industry for which these data were available, I computed the ratio of the PIERS-based measure to the self-reported total expenditure. I interpret this ratio as capturing the extent to which the type of equipment purchased by the industry in that year was being purchased domestically. I then reestimate the regressions of interest from Table 2, but with the sample split based on the industry correlation with with the ASM measure. If the results were biased by differential behavior by those industries for which the PIERS-based measure was a less-accurate measure of automation capital, that would manifest in qualitatively different results.

On the contrary, the results are statistically indistinguishable at conventional levels ($p < 42.58$). This suggests that while the PIERS-based measure captures less of the automation capital for some industries, there is little reason to believe that the competitive drivers of buying domestically produced equipment are sufficiently different to invalidate the results.

A.6.5 Robustness to alternate classifiers

The results above were the product of a heuristic classifier that represented the factor loadings from a series of supervised learning algorithms. To implement that classification, a set of entries were hand-coded by at least two research assistants. Where there were disagreements, the tie was broken by the author. Empirical models were then run predicting the configurations of tokens across the population that indicated a commodity was what I called "automation capital." The empirical models included simple logit, random forest, lasso, and ridge regression. All produced qualitatively similar factor weightings. For concreteness and ease of description, the main results presented here are based on a deterministic routine derived from the loading on token configurations.

Table A14: Robustness of results to accounting for industry dollar percentage of equipment captured in PIERS

	(1)	(2)
VARIABLES	Ln(1+ Automation capital stock value)	
Revenue gap [lag]	5.219***	8.418**
	(1.892)	(3.858)
Leader × Revenue gap [lag]	-9.699***	-14.16**
	(3.474)	(6.383)
Leader [lag]	3.981***	5.341**
	(1.441)	(2.392)
Ln(1+Sales) [lag]	0.872***	1.842**
	(0.315)	(0.852)
Ln(1+Employees) [lag]	0.0535**	0.0637
	(0.0232)	(0.0394)
Observations	182,567	223,599
R^2	-0.282	-1.035
Model	2SLS	2SLS
Establishment FE	Yes	Yes
NAICS6 × Year FE	Yes	Yes
Sample	Bottom quartile ASM correlation industries	Top quartile ASM correlation industries
Markets	Subnational	Subnational
Kleibergen-Paap rk Wald F	7.60	3.16

Errors clustered at market-by-year level. Using the critical values from Stock and Yogo (2005), the instruments are strong enough that the Wald would reject at<10 percent if the true power were 5 percent for both columns. *** p<0.01, ** p<0.05, * p<0.1.

To demonstrate the robustness of the above results to other plausible classifiers, I categorized the data using three additional classifiers and produced qualitatively similar results, presented in Table A15.

The first alternate classifier is derived from the text in the patent abstracts from

Table A15: Robustness of results to alternate classifiers

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Ln(1+ Automation capital stock)					
Revenue gap [lag]	10.74***	2.059***	6.296***	1.408***	12.44***	1.107
	(2.422)	(0.751)	(1.248)	(0.460)	(3.144)	(0.866)
Leader × Revenue gap [lag]	-18.34***	-3.539***	-10.77***	-2.426***	-20.94***	-1.844
	(4.175)	(1.294)	(2.155)	(0.794)	(5.356)	(1.476)
Leader [lag]	7.395***	1.425***	4.311***	0.966***	8.442***	0.742
	(1.677)	(0.519)	(0.859)	(0.316)	(2.152)	(0.593)
Ln(1+Sales) [lag]	2.042***	0.403***	1.179***	0.267***	2.384***	0.227
	(0.454)	(0.141)	(0.232)	(0.0854)	(0.589)	(0.162)
Ln(1+Emp) [lag]	-0.0144	-0.00634	-0.000415	-0.000100	0.00418	0.0164*
	(0.0254)	(0.00837)	(0.0121)	(0.00473)	(0.0308)	(0.00905)
Ln(1+ Auto. capital stock) [lag]		0.850***		0.860***		0.851***
		(0.00205)		(0.00269)		(0.00173)
Observations	1,499,874	1,499,874	1,407,927	1,407,927	1,533,408	1,533,408
Measure	Webb	Webb	Sandler	Sandler	Nof	Nof
Model	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Establishment FE	Yes	Yes	Yes	Yes	Yes	Yes
NAICS6×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Markets	Subnational	Subnational	Subnational	Subnational	Subnational	Subnational
Kleibergen-Paap rk Wald F	13.03	13.01	20.03	19.95	10.50	10.50

Errors clustered at market-by-year level. Using the critical values from Stock and Yogo (2005), the instruments are strong enough that the Wald would reject at less than 10 percent if the true power were 5 percent. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

the patents scored as robotics patents in Webb (2020). I collected a vocabulary of every token in every patent classified as "industrial robot"-related. I then created a dataset of all the unique token stems in all the commodities in the PIERS data and marked those which were also present in one of the patent abstracts from Webb (2020). I then regressed an indicator of whether the stem was present in a "robot" patent abstract on indicators for each token. The resulting stem-level coefficients can be interpreted as the degree to which the focal token was indicative of being in a “robot” patent abstract. Keeping those whose coefficients were estimated to be different from zero at the one-percent level, I then computed a commodity-level score as the sum of the scores of the token in the commodities’ descriptions. I then scored all commodities with a positive regression score as “automation capital” and reestimated the results from Table 2. Columns 1 and 2 in Table A15 show qualitatively similar results.

I then sought to classify commodities using terms not explicitly classified for that purpose. To do so, I used the top two books in the structured categories of Google’s Book Search that had been published by 1999 in the subject “Technology & Engineering Automation”: Sandler (1999) and Nof (1999). I extracted the “Common Words and Terms” from those two books and used them for the classifying regression described above. Once tokens were classified, I classified commodities as described above and rebuilt the data using these new classifications.

Though the classifiers I term “Webb”, “Sandler”, and “Nof” are all disjoint, the results depicted in Table A15 are all qualitatively similar, suggesting the results are not simply a product of my term classifier.

A.6.6 Robustness of estimates to estimation and data construction strategy

Having found support for the measures, I investigate the robustness of the results to the estimation and data construction strategy.

Accounting for the exclusion restriction in the main regressions As noted above, the exclusion restriction in my 2SLS regressions requires that the focal market's structure only correlates with other markets' lagged automation investments through the structures of those other markets. This restriction would be violated if multi-plant firms made a multi-market automation investment decision together. While exclusion restrictions cannot be tested, I can verify whether the above results hold on the subset of the data for which this is not a concern—single-plant firms.

Table A16 presents the results of reestimating the main analysis on only single-plant firms. These results are qualitatively similar to the main results, suggesting that violation of the exclusion restriction does not drive the main results.

Likelihood of these results with random matching To verify that the results are not driven by the automated matching estimator, I estimate the same model on simulated placebo data generated through random matches of shipments to establishments within the same state. The within-state simulation is intended to replicate the within-state blocking in the matching routine.

Figure 14 shows the distribution coefficients generated through 200 replications of the random matching. The coefficients generated by the actual matcher—those appearing in Table 2—are denoted by a dotted line. The coefficients for all three estimated parameters approximate a normal distribution, as expected. The mean of those distributions is statistically different from zero for all three parameters of interest estimated.

Those distributions suggest that the matching process actually biases the effects away from significance for the coefficients testing the hypothesized effects. This adds confidence to the direction of the sign and suggests that the magnitude of the effect may be even stronger than the results in Table 2 suggest.

Table A16: Robustness to using only single-plant firms

VARIABLES	(1)	(2)
	Ln(1+ Automation capital stock value)	
Revenue gap [lag]	5.026***	0.764**
	(1.050)	(0.385)
Leader × Revenue gap [lag]	-9.108***	-1.398**
	(1.904)	(0.698)
Leader	3.700***	0.570**
	(0.769)	(0.282)
Ln(1+Sales) [lag]	0.994***	0.156**
	(0.204)	(0.0749)
Ln(1+Employees) [lag]	0.0895***	0.0189***
	(0.0116)	(0.00483)
Ln(1+ Automation capital stock value) [lag]		0.843***
		(0.00351)
Observations	902,106	902,106
R^2	-0.414	0.669
Model	2SLS	2SLS
Establishment FE	Yes	Yes
NAICS6 × Year FE	Yes	Yes
Sample	Single-plant firms	Single-plant firms
Markets	Subnational	Subnational
Kleibergen-Paap rk Wald F	20.74	20.66

Note: Errors clustered at market-by-year level. Using the critical values from Stock and Yogo (2005), the instruments are strong enough that the Wald would reject at less than 10 percent if the true power were 5 percent. *** p<0.01, ** p<0.05, * p<0.1.

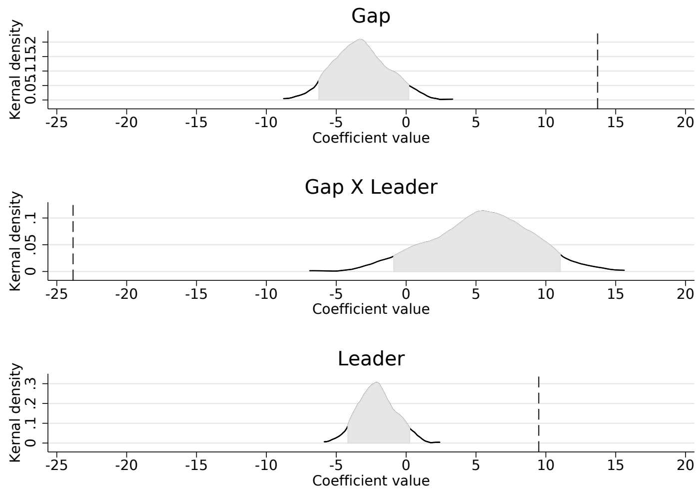

Figure 14: Robustness of results to placebo matching

References

Acemoglu, Daron, and Pascual Restrepo. 2018. "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review*, 108(6): 1488–1542.

Acemoglu, Daron, and Pascual Restrepo. 2020. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy.*

Acemoglu, Daron, Claire Lelarge, and Pascual Restrepo. 2020. "Competing with Robots: Firm-Level Evidence from France." *AEA Papers and Proceedings*, 110: 383–88.

Arellano, Manuel, and Stephen Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *Review of Economic Studies*, 58(2): 277–297.

Arrow, Kenneth. 1962. "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity: Economic and Social Factors.* 467–492. National Bureau Committee for Economic Research and the Committee on Economic Growth of the Social Science Research Councils.

Arrow, Kenneth J, Hollis B Chenery, Bagicha S Minhas, and Robert M Solow. 1961. "Capital-labor substitution and economic efficiency." *Review of Economics and Statistics*, 43(3): 225–250.

Åstebro, Thomas. 2002. "Noncapital investment costs and the adoption of CAD and CNC in US metalworking industries." *RAND Journal of Economics*, 672–688.

Athey, Susan, and Armin Schmutzler. 2001. "Investment and Market Dominance." *RAND Journal of Economics*, 32(1): 1–26.

Autor, David, and Anna Salomons. 2018. "Is automation labor-displacing? Productivity growth, employment, and the labor share." *Brookings Papers on Economic Activity*, 1–87.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The fall of the labor share and the rise of superstar firms." *Quarterly Journal of Economics*, 135(2): 645–709.

Autor, David H, Frank Levy, and Richard J Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics*, 118(4): 1279–1333.

Barnatchez, Keith, Leland Dod Crane, and Ryan Decker. 2017. "An assessment of the National Establishment Time Series (NETS) database." *FEDS working paper.*

Belenzon, Sharon, Victor Manuel Bennett, and Andrea Patacconi. 2019. "Flexible Production and Entry: Institutional, Technological, and Organizational Determinants." *Strategy Science*, 4(3): 175–249.

Bennett, Victor Manuel. 2020. "Changes in persistence of performance over time." *Strategic Management Journal*, 41(10): 1745–1769.

Bennett, Victor Manuel, and Todd Hall. 2020. "Software Availability and Entry." *Strategic Management Journal*, 41: 950–962.

Bertrand, Marianne. 2004. "From the Invisible Handshake to the Invisible Hand? How Import Competition Changes the Employment Relationship." *Journal of Labor Economics*, 22(4): 723–765.

Bessen, James. 2017. "Information Technology and Industry Concentration." *Boston University Working Paper.*

Bessen, James E, Maarten Goos, Anna Salomons, and Wiljan Van den Berge. 2020. "Automatic Reaction-What Happens to Workers at Firms that Automate?" *Boston Univ. School of Law, Law and Economics Research Paper.*

Blake, Richard. 2019. "Smaller Manufacturers That Need A Helping Hand Add Robotic Arms." *Forbes.*

Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, 87(1): 115–143.

Bound, John, Clint Cummins, Zvi Griliches, Bronwyn H. Hall, and Adam B. Jaffe. 1984. "Who Does R&D and Who Patents?" *R&D, Patents, and Productivity*, 21–54. University of Chicago Press.

Bresnahan, Timothy F, and Peter C Reiss. 1991. "Entry and competition in concentrated markets." *Journal of Political Economy*, 99(5): 977–1009.

Chirinko, Robert S. 2008. "σ: The long and short of it." *Journal of Macroeconomics*, 30(2): 671–686.

Chirinko, Robert S, and Debdulal Mallick. 2017. "The Substitution Elasticity, Factor Shares, and the Low-Frequency Panel Model." *American Economic Journal: Macroeconomics*, 9(4): 225–53.

Christensen, Clayton M, and Joseph L Bower. 1996. "Customer power, strategic investment, and the failure of leading firms." *Strategic Management Journal*, 17(3): 197–218.

Cohen, Wesley M. 2010. "Fifty Years of Empirical Studies of Innovative Activity and Performance." In *Handbook of the Economics of Innovation*. Vol. 1, Chapter 4, 129–213. Elsevier.

Cohen, Wesley M, and Steven Klepper. 1992. "The anatomy of industry R&D intensity distributions." *American Economic Review*, 773–799.

Cohen, Wesley M, and Steven Klepper. 1996. "Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D." *Review of Economics and Statistics*, 78(2): pp. 232–243.

Dixon, Jay, Bryan Hong, and Lynn Wu. 2019. "The Employment Consequences of Robots: Firm-Level Evidence." *Available at SSRN 3422581*.

Fellegi, Ivan P, and Alan B Sunter. 1969. "A theory for record linkage." *Journal of American Statistical Association*, 64(328): 1183–1210.

Felten, Edward W, Manav Raj, and Robert C Seamans. 2018. "A Method to Link Advances in Artificial Intelligence to Occupational Abilities." *AEA papers and proceedings*, 108: 54–57.

Flaherty, M Therese. 1980. "Industry structure and cost-reducing investment." *Econometrica*, 1187–1209.

Gans, Joshua S., David H. Hsu, and Scott Stern. 2002. "When Does Start-Up Innovation Spur the Gale of Creative Destruction?" *RAND Journal of Economics*, 33(4): 571–586.

Gervais, Antoine, and J Bradford Jensen. 2019. "The tradability of services: Geographic concentration and trade costs." *Journal of International Economics*, 118: 331–350.

Graetz, Georg, and Guy Michaels. 2018. "Robots at Work." *Review of Economics and Statistics*, 100(5): 753–768.

Gutiérrez, Germán, and Thomas Philippon. 2017. "Declining Competition and Investment in the US." *National Bureau of Economic Research Working Paper*.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large versus Young." *Review of Economics and Statistics*, 95(2): 347–361.

Handwerker, Elizabeth Weber, and Matthew Dey. 2019. "Megafirms and Monopsonists: Not the same employers, not the same workers." *Bureau of Labor Statistics Working Paper*.

Henderson, Rebecca M, and Kim B Clark. 1990. "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms." *Administrative Science Quarterly*, 35(1): 9–30.

Hicks, J. 1932. *The Theory of Wages.* Macmillan.

Humlum, Anders. 2020. "Robot Adoption and Labor Market Dynamics." *Princeton University Working Paper*.

Hummels, David. 2007. "Transportation costs and international trade in the second era of globalization." *Journal of Economic Perspectives*, 21(3): 131–154.

ISC Custom Cables. 2019. "Insert Molding for PCBs, Sensors, Switches, Devices, Embedded Electronics." [Online; last accessed January 8, 2020].

Jarmin, Ron, and Javier Miranda. 2002. "The Logitudinal Business Database."

Karshenas, Massoud, and Paul L Stoneman. 1993. "Rank, stock, order, and epidemic effects in the diffusion of new process technologies: An empirical model." *the RAND Journal of Economics*, 503–528.

Klepper, Steven. 1996. "Entry, Exit, Growth, and Innovation over the Product Life Cycle." *American Economic Review*, 86(3): 562–583.

Koch, Michael, Ilya Manuylov, and Marcel Smolka. 2019. "Robots and firms."

Lovins, Julie Beth. 1968. "Development of a stemming algorithm." *Mech. Translat. & Comp. Linguistics*, 11(1-2): 22–31.

Mann, Katja, and Lukas Püttmann. 2017. "Benign Effects of Automation: New Evidence from Patent Texts." *SSRN*.

Neumark, David, Brandon Wall, and Junfu Zhang. 2011. "Do small businesses create more jobs? New evidence for the United States from the National Establishment Time Series." *Review of Economics and Statistics*, 93(1): 16–29.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica*, 49(6): 1417–1426.

Nof, S.Y. 1999. *Handbook of Industrial Robotics. Electrical and Electronic Engineering*, Wiley.

Rinz, Kevin. 2018. "Labor Market Concentration, Earnings Inequality, and Earnings Mobility." *CARRA Working Paper Series*.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter. 2020. "Diverging Trends in National and Local Concentration." *NBER Macroeconomics Annual 2020, volume 35*. University of Chicago Press.

Sandler, B.Z. 1999. *Robotics: Designing the Mechanisms for Automated Machinery. A Solomon Press book*, Elsevier Science.

Shapiro, Carl. 2018. "Antitrust in a time of populism." *International Journal of Industrial Organization*, 61: 714–748.

Spence, Michael. 1984. "Cost reduction, competition, and industry performance." *Econometrica*, 101–121.

Steorts, Rebecca, Brenda Betancourt, Andee Kaplan, Neil Marchant, and Beidi Chen. 2019. "Record Linkage Tutorial."

Stock, James, and Motohiro Yogo. 2005. *Asymptotic distributions of instrumental variables statistics with many instruments.* Cambridge University Press.

Sutton, J. 1998. *Technology and Market Structure: Theory and History.* MIT Press.

Sutton, John. 1996. "Game theoretical models of market structure." *Address to the World Congress of the Econometric Society, Tokyo, 1995.*

Syverson, Chad. 2004. "Market structure and productivity: A concrete example." *Journal of Political Economy*, 112(6): 1181–1222.

Syverson, Chad. 2011. "What determines productivity?" *Journal of Economic Literature*, 49(2): 326–65.

Syverson, Chad. 2019. "Macroeconomics and market power: Context, implications, and open questions." *Journal of Economic Perspectives*, 33(3): 23–43.

Van Reenen, John. 2018. "Increasing Differences between firms: Market Power and the Macro-Economy." *MIT Working Paper*.

Webb, Michael. 2020. "The Impact of Artificial Intelligence on the Labor Market." *Available at SSRN 3482150*.

Windmeijer, Frank. 2005. "A finite sample correction for the variance of linear efficient two-step GMM estimators." *Journal of Econometrics*, 126(1): 25–51.