Econometrica Supplementary Material

SUPPLEMENT TO "NEXUS TAX LAWS AND ECONOMIES OF DENSITY IN E-COMMERCE: A STUDY OF AMAZON'S FULFILLMENT CENTER NETWORK"

(*Econometrica*, Vol. 91, No. 1, January 2023, 147–190)

JEAN-FRANÇOIS HOUDE
Department of Economics, University of Wisconsin-Madison

PETER NEWBERRY
Department of Economics, Terry College of Business, University of Georgia

KATJA SEIM
Department of Economics and School of Management, Yale University

APPENDIX A: DATA

A.1. *Expenditures for the Representative Household*

IN THIS APPENDIX, we provide details on the spending data and construction of our measures of online spending by shopping mode and of offline expenditure for the representative household in county i and year t. The primary source for our online spending data is the comScore Web Behavior Database, which tracks the online purchasing and browsing activity of a random sample of internet users. With their permission, comScore records any activity on the users' registered computer, including that of other household members, and also collects the zip-code and demographic characteristics of participating households. For a given household, comScore records every online order during its time in the sample. For each order, we observe the seller's domain, the date and time of the order, the product category of item(s) purchased, the list price for each individual item, and a 'basket total', representing the order's total cost, including shipping and taxes. Coverage of the sample is reported in the first three columns of Table AI.

We begin the process of constructing our spending variables by removing transactions for product categories (defined by comScore) that Amazon does not compete in, for example travel and dating services. Next, we classify each of the remaining domains into one of the three online modes, where the classification depends on the retailers' offline footprint and, therefore, their tax liabilities. Appendix Table AII lists the top ten domains each that fall into the non-Amazon categories of sellers. Then, we aggregate the 'basket total' across all transactions through mode k for household h in year t. This results in annual 'household expenditures' in each mode.

The next step is to account for the fact that we only observe 'household expenditures' for households that made at least one online purchase in year t. To do this, we supplement the comScore data with survey data from Forrester Research, Inc.'s "North American Technographics Online Benchmark Survey." Among other information, the survey records for the period 2006 to 2007 and 2010 to 2014 whether a responding household indicates having made an online purchase over the three months prior to the data collection, together with the age, income, race, and zip code of the respondent. Patterns in

Jean-François Houde: houde@wisc.edu
Peter Newberry: pwnewberry@uga.edu
Katja Seim: katja.seim@yale.edu

[Correction added on 6 February 2023, after first online publication: The copyright line was updated.]

TABLE AI

COMSCORE SAMPLE COVERAGE AND SPENDING BY RETAIL CHANNEL.

Year	House-holds (k)	Coverage (%)		Average Expenditure ($) [Orders]			
		Counties	House-holds	Offline	Amazon	Taxed	Non-taxed
2006	38.3	84	99	5341	62 [2]	510 [13]	486 [11]
2007	39.5	85	99	5474	86 [3]	588 [14]	502 [11]
2008	20.2	71	96	5520	106 [4]	569 [14]	450 [10]
2009	15.9	64	95	4912	129 [4]	619 [15]	514 [11]
2010	15.4	66	95	4662	195 [7]	687 [16]	569 [12]
2011	17.3	68	96	4792	286 [9]	865 [20]	542 [12]
2012	18.2	67	96	4494	378 [12]	1015 [23]	553 [12]
2013	16.5	62	95	3816	491 [16]	1062 [24]	693 [15]
2014	12.0	54	92	3846	604 [19]	1107 [25]	742 [16]
2015	16.3	64	95	3219	798 [25]	1250 [28]	853 [18]
2016	24.8	72	97	3272	1040 [32]	1272 [28]	1031 [22]

Note: County and household coverage are the percentage of U.S. counties and households residing in the comScore data. Expenditures and orders are the average across households in the given year.

the Forrester data suggest an increasing take-up of e-commerce (see column (1) in Table AIII).

We employ a linear probability model to project each Forrester respondent's propensity to purchase online on household demographic categories (race, age, and census region of head of household), household income, and time trends. We then use the estimates of the model to predict the probability that a household in the comScore data made an online purchase in a given year, based on their demographics. For the years 2008 and 2009, we use linear interpolations for a given demographic group based on the predictions in 2007 and 2010. Using the predicted online purchase propensities from this model, we calculate 'expected annual expenditures' on each shopping mode for the households in the comScore sample, assuming that the propensity to purchase online applies to all modes equally.

TABLE AII

TAXED AND NON-TAXED COMPETITORS.

Sales Rank	Taxed	Non-Taxed
1	walmart.com	ebay.com
2	jcpenney.com	qvc.com
3	bestbuy.com	dell.com
4	macys.com	yahoo.net
5	apple.com	hsn.com
6	homedepot.com	overstock.com
7	victoriassecret.com	fingerhut.com
8	target.com	amway.com
9	staples.com	newegg.com
10	gap.com	orientaltrading.com
Total	102	292

Note: Table displays top 10 domains in terms of 2013 expenditures in the comScore data that we define as taxed and non-taxed.

TABLE AIII

UNSCALED HOUSEHOLD ONLINE PURCHASING.

	Forrester Offline Only (%)	Average Expenditures w/out Scaling			
		Amazon ($/year)	Taxable ($/year)	Non-Taxable ($/year)	All ($)
2006	55.6	18.2	99.8	95.1	213.0
2007	60.8	19.8	102.8	87.7	210.3
2008	–	21.5	99.8	78.9	200.3
2009	–	25.3	81.5	67.8	174.6
2010	32.1	34.1	76.7	63.6	174.3
2011	23.0	52.0	88.3	55.4	195.6
2012	23.9	84.1	85.0	46.3	215.5
2013	27.2	78.9	81.0	52.8	212.7
2014	24.6	77.3	85.2	57.1	219.6
2015	23.0	98.0	107.0	73.1	278.1
2016	22.6	116.2	98.5	79.8	294.5

Note: Expenditures are the average across households. Offline Only denotes the share of respondents who answered no to the question whether they had shopped online in the previous three months in the Forrester Technographics Survey.

Next, we aggregate across households to derive the expenditure of a representative household in county i in year t on the three online shopping modes. To do this, we first calculate the average expenditures for demographic group z in county i and year t for mode k. We then apply demographic sampling weights that measure the relative prevalence of demographic group z in the comScore data and data from the Census to derive expenditures for mode k for a representative household in county i and year t.[S1,S2]

Finally, we address the intensive-margin bias in comScore spending introduced by the firm not recording the full universe of household online activity across devices. We assume that under-reporting in comScore online spending is uniform across counties and scale up each representative household's spending by a year and shopping-mode varying scale factor. We determine the scale factor by matching the average household spending calculated using our sample to the average spending per household on Amazon calculated using Amazon's annual reports and spending on the other modes calculated using the U.S. Census Bureau's quarterly e-commerce retail sales reports. Specifically, we multiply the expected expenditures for each representative household by the scaling factor and then calculate the average household spending across the United States on each mode. We search for the scaling factors for each mode and year where the resulting household averages match the averages from the supplemental sources.

In order to construct a spending figure from Amazon's annual financial statements that is comparable to the comScore data, we only include the reported sales from the "Media" and "Electronics and Other General Merchandise" categories in North America. An additional issue in matching the spending data is that the financial statement excludes sales

[S1]De los Santos, Hortaçsu, and Wildenbeest (2012) compared the sample of comScore users in 2002 and 2004 to the Computer Use Supplement of the Current Population Survey and found that the sample generally compares well with the population of online shoppers.

[S2]The sampling weights are constructed based on the relative number of households that fall into different demographics bins in the comScore sample and in the population. The population data come from the American Community Survey ('ACS-5 year') from 2009 to 2016. To extrapolate data back to 2006, we assume a county-level constant growth rate in population belonging to each bin between 2000 and 2009, where the 2000 data come from the decennial Census.

tax paid and includes only the royalties and fees earned off of third-party sales. The comScore data, on the other hand, include sales taxes and the full revenue from third-party sales. To account for this when determining the scaling factors, we adjust the comScore data to exclude taxes and to include only the portion of third-party sales that Amazon retains. We calculate the latter using Amazon's royalty rate on third-party transactions (see description in Section A.7 below) and data on the evolution of the share of total sales accounted for by third-party sellers over the sample period reported in the 2018 annual report cover letter.

The right panel of Table AI summarizes the average household spending by shopping mode. Spending on all online modes has increased substantially over time as people substitute away from offline shopping, with Amazon displaying the most pronounced growth. By 2016, the average household spending is about $1000 on Amazon, while it spends about the same amount on non-taxed competitors. For taxed competitors, the spending is higher at $1300. Finally, households spend about $3300 at offline retailers. In brackets, we display the number of orders for the online modes, which are calculated by dividing the household spending by the price index calculated in Appendix A.6. The orders have roughly the same growth patterns as spending.

Figure A1 demonstrates the growth in Amazon's share of online and all retail spending over time. These numbers are calculated using data from Amazon's financial statements and reports by the US Census bureau, which is why we have data points outside of our comScore sample. Amazon's online share remains relatively stable, and even decreases, until about 2006. Thereafter, we see a rapid increase, culminating at over over 40% by 2018. Amazon's share of total retail has similar patterns, reaching about 5% in 2018.

We explore heterogeneity in spending on Amazon across demographics and geography in Table AIV, where we break down average annual household spending on Amazon across urban and rural counties, wealthy and non-wealthy counties, and the counties belonging to the four different Census regions. Overall, the growth rates are similar across demographic groups and regions, but there exist significant cross-sectional differences in spending based on region, income, and whether or not the consumer lives in a city.

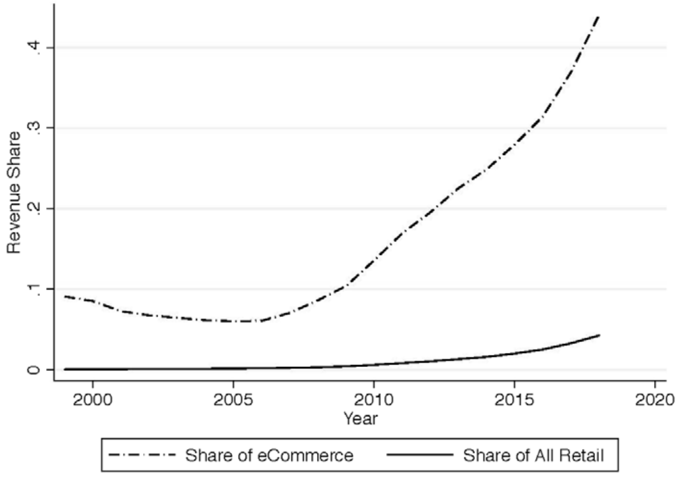

FIGURE A1.—Amazon's growth. Notes: Calculations are made using reported sales in Amazon's financial statements and reported online and total retail sales reported in the U.S. Census Bureau's quarterly e-commerce retail sales reports.

TABLE AIV

AVERAGE SPENDING ON AMAZON, BY TYPE OF COUNTY ($/YEAR).

			Income					
Year	Urban	Rural	High	Low	Northeast	Midwest	South	West
2006	70	55	87	57	80	54	52	74
2007	98	76	121	79	105	79	76	100
2008	120	99	131	102	131	93	99	129
2009	152	118	175	124	155	126	117	161
2010	212	197	255	189	234	187	180	238
2011	322	273	372	267	360	275	263	327
2012	423	367	464	367	436	384	353	447
2013	534	505	597	487	558	498	468	595
2014	656	665	752	615	721	603	589	774
2015	883	793	981	765	949	755	798	910
2016	1125	1020	1219	989	1159	1019	1007	1181

Note: We define rural counties as counties with a population density of less than 500 residents per square mile and low-income counties as counties with average household income below $80,000. Region is as defined by the U.S. Census Bureau.

Finally, we report the unscaled average household spending data in Table AIII and compare them to the scaled data in Table AI. For Amazon, the pattern of increasing household spending matches the scaled data, but the magnitude of the growth is smaller. This is in line with the intuition that the share of 'missing' transactions increases over time due to the take-up of mobile purchasing. Modes 2 and 3 generally show more of a U-shaped or flat pattern, which is largely inconsistent with what we see in the data from the U.S. Department of Commerce. Again, this is likely because of the data we are missing from mobile or other-computer transactions for these modes.

Note that the scaling is mode-year specific, so that we preserve the rich variation in spending across geographies and demographics from the comScore and Forrester data. This variation identifies the consumer response to sales tax, while the scaling acts to correct the *level* of spending and revenue by mode, which is important in identifying the supply side of the model. Therefore, the scaling exercise mostly serves to capture the degree of Amazon's growth relative to the other modes in line with the data observed in the reports.

A.2. *Employment*

The employment data come from industry sources and Amazon's financial statements and press releases. We construct a cross-section of the number of employees at a subset of fulfillment and sortation centers in 2017. Our main source of information is MWPVL and the establishment survey YTS (Your-economy Time Series).[S3] MWPVL reports the target number of employees for a subset of facilities. We match the list of facilities and the panel of Amazon establishments from YTS. This survey provides employment data for most facilities, but the overlap is not perfect. In addition, although the YTS data are annual, the data exhibit very little adjustments over time, and so we use the more recent cross-section (2017). When two facilities are covered by YTS and MWPVL, we use an average of the two estimates. Finally, for a small number of facilities that are unmatched,

[S3]See https://wisconsinbdrc.org/data/

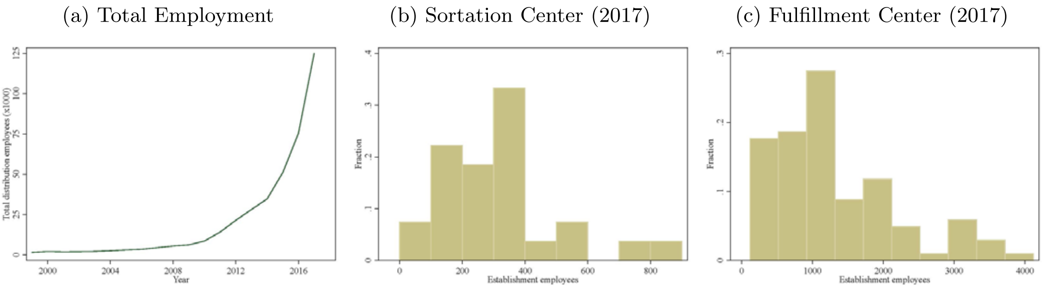

FIGURE A2.—Employment by facility type and year. Notes: The left graph is the time-series of overall fulfillment employment. The center and right graphs are the distributions of employment across sortation centers and fulfillment centers in 2017, respectively.

we use data from Reference USA which includes information on employees for a limited set of Amazon facilities.[S4] Overall, we use employment information for 131 out of the 163 facilities active in 2018 (including 26 sortation centers).

We also observe the total number of fulfillment employees in 2017, 125 thousand, from an Amazon press release.[S5] We combine the latter and information from Amazon's financial statements to get an estimate of the total number of fulfillment employees for the other years in our sample. We observe the logistic cost share of total cost and the total number of employees in Amazon's financial statements for our entire sample. With these data and the number of logistic employees from the press release, we construct the ratio of the cost share of logistics to the employment share of logistics in 2017. Assuming this ratio is constant over time allows us to back out the number of logistics employees for all of the years in our sample.

Figure A2 provides a summary of these data. Figure A2a demonstrates that the increase in fulfillment employment coincides with the accelerating expansion of the network starting around 2010. The histograms in Figures A2b and A2c show that both types of facilities have a significant amount of variation in employment in 2017, with the level of employment being much higher at fulfillment centers.

A.3. *Sales Tax*

Our tax rates come from Thomson Reuters' Tax Data Systems, which provide state, county, and local tax rates for the years 2006–2016. We add these rates together and compute the average tax rate across municipalities in each county for each year of the sample. Because we do not observe rates before 2006 or after 2016, we assume that the tax rates from 1999 to 2005 are equal to the tax rate in 2006 and tax rates from 2017 to 2018 are equal to the tax rate in 2016. Table AV summarizes the tax rate data. The household weighted average sales tax rate varies between 6.75% and 7.13%, with a standard deviation of between 1.5% and 1.6% across counties in every year. In addition, tax rates vary across time, as between 42 and 68% of households live in counties that experience tax rate changes from year to year and all counties experience at least one tax rate change over our sample period. Finally, higher population counties tend to have higher tax rates, as demonstrated by the final column.

[S4]See http://www.referenceusa.com/Home/Home

[S5]See https://press.aboutamazon.com/news-releases/news-release-details/amazon-now-hiring-over-120000-jobs-us-holiday-season

TABLE AV
TAX RATES.

Year	Ave Tax	StDev Tax	% HHs w/ Change	Corr(#HHs, Tax)
2006	6.75	1.54	55.33	0.09
2008	6.82	1.51	67.47	0.10
2010	7.03	1.66	67.97	0.13
2012	6.96	1.56	64.60	0.11
2014	7.02	1.56	42.13	0.11
2016	7.13	1.59	42.69	0.10

Note: The average and standard deviation of tax are moments from the distribution of the county-level average tax rates. HHs with change is the percentage of households that lived in a county where the average sales tax rate changed that year. Corr is the county level correlation between households and the average sales tax rate.

This tax rate is assumed to apply to all transactions at brick-and-mortar and taxed online retailers. Amazon transactions are taxed, and the local sales tax rate applies, if the consumer lives in a state where Amazon collects taxes. As mentioned in the text, nexus tax laws would suggest that this occurs when Amazon operates a facility in the consumer's state. However, it is not always the case that these two events coincide perfectly, and here, we provide anecdotal and empirical evidence that the gaps in timing are due to negotiation between state governments and Amazon.

We begin by providing a few pieces of anecdotal evidence of this connection. First, in the 1999 through 2011 financial statements, Amazon states that "a successful assertion by one or more states or foreign countries that we should collect sales or other taxes on the sale of merchandise or services could result in substantial tax liabilities for past sales, decrease our ability to compete with traditional retailers, and otherwise harm our business." (1999, page 15) There are similar quotes in later statements about the repercussions of states "requiring [Amazon] to collect taxes where [they] do not" (e.g., 2017, page 12). This suggests that Amazon considered sales tax to be a first-order issue impacting their bottom line.

A second piece of anecdotal evidence comes from documented negotiations between specific states, such as Nevada and Texas, where the debate between Amazon and the state governments over locating in the state centered around sales tax and Amazon's physical presence in the state. Amazon also chose to shut down its affiliate program in Illinois in order to avoid sales tax when the state changed the nexus laws to include affiliates. These examples provide suggestive evidence of the importance of the relationship between sales tax and Amazon's strategic decisions.

There are occasions, mostly toward the end of our sample, where Amazon began to charge sales tax before opening a facility. However, to the best of our knowledge, in all but one of these cases (NY), Amazon had plans to build a facility in that state soon after the onset of sales tax collection.[S6] Of course, it could be that Amazon decided to build a facility only after conceding on the sales tax issue, but this seems unlikely given the nature and history of nexus laws. States have unsuccessfully fought for years to change the nexus laws to include companies without a physical presence in their jurisdiction, like Amazon. However, Amazon did start to charge sales tax in all U.S. states in 2017, irrespective of physical presence. This appears to be mostly a move for good publicity and/or to put

[S6]See, for example, the situation in New Jersey: https://www.nj.com/news/2012/05/amazoncom_to_begin_collecting.html pressure on federal regulators to change the nexus laws, as Amazon had already been charging sales tax to over 90% of the U.S. population by this time.

The final pieces of anecdotal evidence are the observed location decisions of Amazon. Specifically, placing facilities on the western border of Nevada, the southern border of New Hampshire, the southern border of Wisconsin, and in Delaware suggest that sales tax played an important strategic role in determining their locations.

Given these anecdotes, we argue that the change in sales tax obligation is triggered by entry, but that this change may only take effect after negotiations with state officials. The way that we think about these negotiations is based on the observed changes in the time between entry and sales tax collection over time.

We observe that in the early period of our sample, there was often a significant lag between the date of entry and the change in sales tax laws. For example, Amazon did not start collecting sales tax in Pennsylvania until nine years after it opened its first fulfillment center in the state. This was a time when Amazon had significant bargaining power with the states. One reason is the lower demand in early time periods, so it was not as important for Amazon to be close to its customers. Additionally, in early periods, they were not located in very many states, meaning there may be competing options of other, possibly lower sales-tax, states. For both of these reasons, Amazon's threat point (the value of opening in a different state) was higher in the early periods. An example of this comes from Nevada in 2011, where a law that would have forced Amazon (and other online retailers) to collect sales tax failed due (partially) to concerns that Amazon would close its fulfillment center in the state.[S7] Later, Amazon negotiated a deal with the state to start collecting sales tax in 2014.[S8] In another example in the same year, Amazon closed a fulfillment center and abandoned plans for expansion in Texas due to a dispute about paying uncollected taxes. These two examples show Amazon using its leverage in order to gain preferential tax treatment in the early years of our sample.

However, as Amazon continued to expand their network, their bargaining power with states lessened. State governments knew that, given the level of demand, it was important for Amazon to be close to customers and that there were not many tax-friendly fallback states remaining. Because of this, we see Amazon agreeing to charge sales tax quickly after the opening of the first facility in a state and sometimes even before.

To formalize this argument, we use data on the time between entry and sales tax. First, we present Figure A3, which shows, for each first entry into a given state, the time between entry and tax collection. This demonstrates a negative relationship between date of first entry and the lag in tax collection dates. Second, we run a regression where the dependant variable is the time between first entry into a state and the beginning of tax collection. The regressors are a linear time trend, a dummy variable indicating that the first entry in that state occurred post-2012, and the minimum tax rate in a nearby state (i.e., same census region) that has yet to have a fulfillment center. The time trend is intended to capture the reduction in bargaining power due to the increase in demand, while the post-2012 dummy variable captures a discrete shift in bargaining strategy at this time period. The tax variable is included to proxy for the value of the 'outside option'. Table AVI displays the results, which show that the lag gets shorter post-2012 and for each additional year. Additionally, the more attractive the outside option (i.e., lower tax rates), the higher the lag in collection.

[S7]See https://www.reviewjournal.com/uncategorized/taxation-committee-drops-internet-sales-tax-amendment/?ref=894

[S8]See https://lasvegassun.com/news/2012/apr/23/nevada-reaches-agreement-amazon-collection-sales-t/

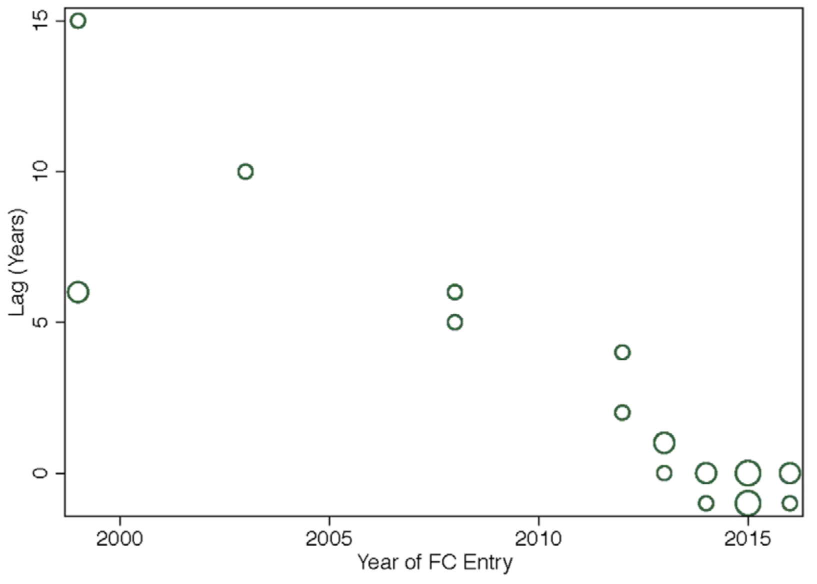

FIGURE A3.—Tax date lags. Notes: Each point represents a year of first entry into the state and the corresponding lag or lead in the implementation of the collection of sales tax. The size of each point indicates the number of facilities with that entry and lag.

We take this into account in the supply side model and in the counterfactuals by assuming a deterministic tax abatement schedule that depends on the entry year. We believe that this schedule is able to capture the most important aspect of the negotiations (the

TABLE AVI

TAX LAG REGRESSION.

Variables	(1) Tax year—Entry year
Minimum tax rate in unoccupied states from the same region	-41.80^{**}
	(17.05)
State entry year	-0.418^{***}
	(0.107)
1(Entry year > 2012)	-2.295^{*}
	(1.251)
Constant	847.1^{***}
	(214.9)
Observations	23
R-squared	0.861
Mean dep. variable	2.217
SD dep. variable	4.112
Mean tax variable	0.0393
SD tax variable	0.0225

Note: Standard errors in parentheses. The dependent variable is the first year of tax collection minus the year of the first entry into a state. Minimum tax rate is the smallest tax rate of states in the same region that do not yet have a facility. Regions are defined by the U.S. Census Bureau.

time-series) and that a more comprehensive bargaining model between the states and Amazon, although interesting, is out of the scope of this paper.

A.4. *County Characteristics*

We explain our procedure for collecting and imputing the county-level data in our analysis. We separate the sections by the different categories of data we collect.

Demographics

The estimation of the model requires observing the demographic characteristics of the representative household and the number of households in each county in the United States for the years 1999 to 2018. The characteristics we include are income, age, and race. To measure the income of the representative household, we include the average household income in a county as well as the share of households that have household income above 100 thousand. To measure age, we include the share of households with a head of the household above 35 years old and to measure race, we include the share of households with a head of the household who is black, Asian, or another race.

These variables, as well as the total number of households, are available during census years (2000 and 2010) for every county, but between census years, we rely on the Census's American Community Survey (ACS). The drawback of the ACS is that the data are only available for high-population counties for the years 2006–2008 via the 'ACS-1 Year' (depending on the variable, missing percentage ranges from 41% to 75%). The 'ACS-5 Year', on the other hand, is available between 2009 and 2016 for all counties. At the time when the data were collected, the 2017 and 2018 'ACS-5 year' were not yet available. To summarize, we observe the necessary variables (1) for the complete set of counties in 2000 and from 2009 to 2016, (2) for only large counties from 2006 to 2008, and (3) for no counties in 1999, 2001–2005, and 2017–2018.

Therefore, we impute demographics and the number of households for the missing counties between 2006 and 2008 and all counties in 1999 and between 2001 and 2005 using the following procedure. We regress each county-level demographic variable or number of households on state $\times$ year fixed effects and county fixed effects:

$$y_{it} = \mu_i + \tau_{\text{state}(i),t} + e_{it}.$$

This regression is estimated using observations for all counties during the census years, as well as non-missing ACS counties for other years. We use the predicted value of this regression to impute the data for missing counties from 2006 to 2008.

In order to use the model to predict the data for all counties in 1999 and 2001–2005, it is necessary to have values of the state/year fixed effects for these years. We assume a constant annual growth rate of these effects between 1999 and 2006, which we are able to determine using the estimates of the fixed effects for 2000 and 2006.

Finally, in order to predict the variables for 2017 and 2018, we assume a constant annual growth rate of each demographic variable and the number of households from 2015 to 2018. We calculate the growth rate using the data from 2015 to 2016.

Retail Establishments

Our measure of offline competition is the count of local retail establishments. We separate these into large and small retailers by the number of employees (greater or less than 50). For these counts, we use data from the County-Business Pattern (CBP) between 1999 and 2016 for all counties. Similarly to the demographic data, we assume a constant growth from 2015 to 2018 in order to predict the offline competition variables.

Retail Wage

To estimate the county-level average retail wage, we use the annual wage in the retail sector observed in the BLS Quarterly Census of Employment and Wages from 1999 to 2016. In each year, there are around 37% counties with no wage data and between 0.01% and 0.03% of counties are outliers (wage out of range $10,000–$50,000). To impute the wage for these counties, we calculate the weighted average wage of the twenty nearest neighboring counties that have data, using an inverse distance weight. The data are then extrapolated to 2017 and 2018 using the same procedure as the one used with the demographics, households, and retail establishments.

Rent

We observe rents for distribution and industrial establishments at the MSA level for 2010–2018 for 47 MSAs from the Regional Economic Information System (REIS), and for industrial only establishments between 2006 and 2010. Assuming the MSA-level ratio of industrial to distribution rent stayed constant over the period, we use both data sets to predict MSA-level distribution rent from 2006 to 2018.

We then proceed in three steps to form a county-level measure of rent for the years 1999–2018. In the first two steps, we use the REIS and other data to predict rents for all MSAs from 1999 to 2018, and in the third step, we use these predictions to estimate a county-level measure. The process proceeds as follows:

1. Using the 2006–2018 data for the observed MSAs, we regress the (log) average rent on the MSA-level CPI, which is available for all years between 1999 and 2018 from the BLS, while controlling for MSA fixed effects. We use the predicted values of this regression to predict rent prior to 2006 for the observed MSAs.
2. To predict rent for the remaining MSAs, we regress (log) average rent on the following MSA characteristics observed in the BLS and census demographic data and aggregated to the county level (in log): population, pop. density, employment rate, industrial employment rate, office employment rate, nb of of employees, nb of establishments, median house value, mean income, labor force participation rate, and year and census division fixed effects. We use the predicted value from this regression to predict rents for the missing MSAs for all years.
3. To predict rent at the county level, we first assign the MSA rent measure to all counties located with 20 miles of the MSA centroid. To predict outside of this radius, we regress MSA rents on the characteristics of counties located within the radius, where we include the following county characteristics: median house value, land size, population density, wage, and median income. The counties outside of the radius are then assigned the predicted value of this regression. Essentially, we extrapolate outside the MSA core using the observed relationship between housing cost and density and average distribution rent.

We note that our measure of rent does not account for the negotiations between property owners and Amazon or any market power Amazon wields in such negotiations. To the extent that Amazon's bargaining power varies across locations, it may lead to biased estimates of our shipping cost parameters. For example, if Amazon's bargaining power were higher in larger markets, then the shipping cost estimate would be overestimated; if bargaining power were higher in smaller, more remote markets, then the shipping cost estimate would be underestimated. It is unclear to us how significant the variation in bargaining power across markets might be.

A.5. *Total Retail Spending*

We use micro data from the Consumer Expenditure Survey (CEX) between 1999 and 2016 to construct a measure of average retail at the county-year level. The micro-data sample includes repeated cross-sections of roughly 60,000 households reporting annual spending on categories covered by Amazon and other general merchandise retailers. We choose categories based on their Universal Classification Code (UCC). The categories cover items that can roughly be put in to the following larger categories: beauty supplies, household items, electronics, apparel and accessories, office/school supplies, books, pet supplies, and sporting goods. A detailed description of which UCC codes are included is available upon request.

We use these data and an imputation approach similar to Blundell, Pistaferri, and Preston (2008) to construct our retail spending for the representative consumer. The starting point is a linear regression relating spending with household characteristics and time/region fixed effects:

$$y_{it} = X_{it}\beta_t + \mu_{r(i),t} + \epsilon_{it},$$

where X_i includes indicators for age and income groups, education, employment status, rate and family composition, and μ is a region/year fixed effect. Note that the regression is estimated separately for each year.

The second step measures the mean of X_{it} for each county/year. We use the aggregate census tables described in the previous section for this task. Let $\bar{X}_{j,t}$ denote this average for county j in year t.

We then use the estimates $(\hat{\beta}_t, \hat{\mu}_t)$ to calculate the conditional expectation of (log) annual spending in each county:

$$\bar{y}_{jt} = \bar{X}_{jt}\hat{\beta}_t + \hat{\mu}_{r(j),t}.$$

Finally, we transform this conditional expectation into levels assuming that spending is distributed according to a log-normal distribution:

$$\text{Spending}_{jt} = \exp\bigl(\bar{y}_{jt} + \hat{\sigma}_\epsilon/2\bigr).$$

Since the log-normal distribution is sensitive to outliers, we windsorize the distribution of $\bar{y}_{jt}$ by truncating the values to the 99.9% percentile. To form data for 2017 and 2018, we extrapolate assuming a constant annual growth rate as in the previous sections.

A.6. *Prices and Variety*

In this section, we describe how we construct the price and variety indices. We first discuss the price index. We calculate the average price of goods purchased over the course of a year for the representative household in each county using the comScore data. Specifically, we calculate the average price of goods purchased on Amazon for each household and then use population weights to aggregate to the county level. Therefore, the construction of this is the same as the construction of the weighted average spending discussed in Section A.1. To smooth any remaining noise in the pricing data and to make out-of-sample predictions, we regress the county-level prices on a linear time trend and the share of households in the county who have a head of the household that makes over $100,000 in income. Including the latter accounts for the fact that higher income households may be buying different things on Amazon.

TABLE AVII

PRICE REGRESSIONS.

Dep Var	P	Log(P)	IQR	Log(IQR)
Linear Time Trend	0.436***	0.009***	1.970***	0.084***
	(0.044)	(0.001)	(0.263)	(0.008)
Share of Pop w/ Inc > 100k	0.069	0.581***		
	(1.566)	(0.038)		
Obs	17,338	17,338	11	11
R-Sq	0.006	0.021	0.862	0.922

Note: Standard errors in parentheses. The first two columns are price regressions using county-level data and the last two columns are variety regression using annual data.

The results of these regressions are in the first two columns of Table AVII. In the first column, the dependent variable is in levels, while in the second column, it is in logs. Both specifications indicate that there is a positive trend in the prices on Amazon over time, and the log specification shows that higher income households buy more expensive items. The time trend in the log-linear model suggests that prices increase by just under 1% each year. We use the results of these models to predict the transacted prices for each county and each year. These predictions are in the first two columns of Table AVIII. Both specifications predict a 1999 price of about \$25. The prices then increase until 2018, with the increase being slightly more for the linear model.

We now turn to the variety index. We construct our measure of variety as the ratio of the interquartile range to the median price in each year, where the median price is calculated in the same way as the average price. As product availability on Amazon does not vary by location, we do not utilize the variation in prices at the county level, which represents only the variation in products purchased by a single household (or county) and not necessarily the products available on the platform. Instead, we utilize the variation in prices across the entire United States, using all transacted prices observed in the comScore data in a given year.

Similarly to the price index, we use a regression to smooth the data and to make out-of-sample predictions. However, here we only include the time trend because the dependent variable is at the year level rather than the county/year level. The results are displayed

TABLE AVIII

PREDICTED PRICE MOMENTS.

	Predicted Price Index		Predicted IQR	
Year	Linear	Log	Linear	Log
1999	25.01	25.16	−0.73	8.05
2002	26.32	25.87	5.18	10.36
2005	27.63	26.71	11.09	13.34
2008	28.94	27.81	17.00	17.17
2011	30.24	28.66	22.91	22.10
2014	31.55	29.67	28.82	28.45
2017	32.86	30.51	34.73	36.63
2018	33.30	30.77	36.70	39.85

Note: The predictions of the log models account for the variance of the residual.

in the third and fourth columns of Table AVII. Despite the fact that we only have 11 observations, the model fits the data well. The estimates show that the variation in prices is increasing over time, with the log-linear model indicating that the IQR increases by about 8% each year. We again use the results to predict the IQR across all years from 1999 to 2018. The last two columns of Table AVIII indicate that the IQR increases substantially over time. While the linear model predicts a negative value in the early years, the middle years (during our sample) and the overall increase are comparable across specifications. In our remaining analysis, we use the predictions of the log-linear model for prices and IQR, normalizing the IQR by predicted average price in constructing the aggregate variety index. We index the resulting standardized IQR to 1 in 2018.

A.7. *Marketplace Spending and Approximating the Gross Margin*

Recall that $\bar{\mu}$ is the weighted average margin between purchases directly through Amazon and purchases through third-party sellers:

$$\bar{\mu}_t = s_t^{3PS}\mu^{3PS} + \left(1 - s_t^{3PS}\right)\mu^{own},$$

where s_t^{3PS} is the share of marketplace sales observed in the 2018 annual report (cover letter). We use the annual reports of Amazon from 1999 to 2018 along with outside sources to determine the values of μ^{3PS} and μ^{own}, which we assume are fixed over time.

We begin with our approximation of μ^{own}. Amazon's annual report displays their gross margin each year (e.g., 2013, page 26), but there are three problems with using this number directly. First, this includes the cost and revenue of third-party sales. Second, this margin includes some of the costs that we will estimate such as labor, land, density, and shipping costs. Third, there are additional costs that we want to include that are not included in 'Cost of Goods Sold', for example, robotics, electricity, etc.

Instead, we calculate the margin as

$$\mu^{own} = \frac{\text{Amazon Sales} - (\text{Cost of Goods Sold} + \text{Other Costs} - \text{Shipping Costs})}{\text{Amazon Sales.}}$$

We observe 'Amazon Sales' in their annual report (e.g., 2016, p.26). Specifically, Amazon reports their total revenue in 'Media' and 'Electronics and Other General Merchandise' categories in North America, which is roughly equivalent to the revenue we predict from our model. Revenue from Canada and Mexico is not included in the comScore data, but this comprises a small share of total revenue.[S9] This figure includes both direct product sales as well as 'service sales', which include revenue from third-party sales. Therefore, we net out the third-party revenue using the share of third-party transactions and the margin earned on third-party transactions, μ^{3PS}, which we discuss below.

Amazon also reports the total 'Cost of Goods Sold', which is composed of wholesale costs and shipping costs. We compute the cost of goods sold for North America by multiplying the total cost by the ratio of sales from North America to total sales. Because we estimate shipping costs, we need to subtract these costs from the margin. Therefore, we net out the 'Shipping Costs' that are observed in the annual report (e.g., 2016, p. 25). Again, we adjust these by the share of sales that are from North America.

[S9]According to S & P Capital Platform's segment analysis of Amazon, nearly 98% of North American revenue came from the United States in 2017.

The ‘Other Costs’ variable should not include any land or labor costs that we are going to estimate. So we collect the ‘Fulfillment Cost’ from the annual report, which “consist of those costs incurred in operating and staffing our North America and International fulfillment and customer service centers and payment processing costs” (e.g., 2016, p. 27), and net out preliminary estimates of the costs of labor, land, and density from our model. We adjust this by the share of North American sales and the share of third-party sales in order to get ‘Other Costs’ for U.S. Amazon transactions only.

The calculated value of μ^{own} varies from year to year, with a high of about 0.2 in the early years and a low of about 0.09 in 2011. We use the average across all years, approximately 0.15, and assume that this is the margin that Amazon realizes on direct transactions, net of any costs we estimate.

Next, we turn to the margin Amazon earns on third-party transactions, which includes both a royalty rate and seller fees such as membership fees and stocking fees. It is not necessary for us to observe these separately, as we only care about the overall margin on third-party sales. In order to determine μ^{3PS}, we use the amount of revenue Amazon earned from third-party sellers from 2014 to 2018, which, along with Amazon’s total revenue and the share of third-party sales, allows us to back out μ^{3PS} for these years. The revenue from third-party retailers is observed starting in the 2016 annual report, which reports this measure 2 years retroactively (p. 68). The backed-out margin ranges from 0.32 to 0.35 during this time frame, so a reasonable estimate may be 0.335. However, because it is likely more costly for Amazon to receive, manage, and ship inventory from a third-party retailer, we use a conservative estimate of $\mu^{\text{3PS}} = 0.3$.

Therefore, the weighted average margin is given by

$$\bar{\mu}_t = s_t^{\text{3PS}} 0.30 + \left(1 - s_t^{\text{3PS}}\right) 0.15.$$

APPENDIX B: Estimation

B.1. *Projecting Spending*

Here, we describe the steps to project demand outside of our sample using additional data. First, we collect data on the variables in C_{it} and Z_{it} for the years and counties outside of our sample from the U.S. Census Bureau. Second, to predict the county fixed effects, we use the estimates of a linear regression of the estimated in-sample fixed effects on a large number of county characteristics. This auxiliary regression fits the data well with an R-squared of 0.89. We also use a regression with time-trend covariates to predict the census division-year fixed effects. We use the predictions of the county fixed effect and the census division-year regressions for both in-sample and out-of-sample data, which smooths the spending data to account for possible measurement error.

Finally, in order to predict the mode-year fixed effects outside of our sample, we bring in aggregate spending data from Amazon’s annual reports and the U.S. Department of Commerce. Specially, we can use the estimates of the demand model and the projections discussed above to predict total yearly spending for a given set of fixed effects. We find the values of the fixed effects such that these predictions match the information from the aggregate spending data.

B.2. *Instrument Construction*

Here, we describe how we construct the instruments. Recall that the instruments must be orthogonal to the measurement error from the demand model, but correlated with the components of the profit difference. Therefore, we construct measures of these components that are not a function of the estimated demand model.

The first variable, a shifter of the shipping distance component, is the total weighted shipping distance difference, where instead of using the number of orders predicted by the demand model as the weight, we use the population of the county:

$$\hat{X}_d^{j,j'} = \sum_{t=t(j)}^{t(j')} \beta^t \left(\sum_{i=1}^{I} \sum_{l=1}^{L} \text{Pop}_{it} \hat{\Omega}_{i,l}\big(N_t|\mathbf{a}^0\big) d_{il} - \text{Pop}_{it} \hat{\Omega}_{i,l}\big(N_t|\mathbf{a}^{j,j'}\big) d_{il} \right).$$

The function $\hat{\Omega}_{i,l}(N_t|\mathbf{a})$ represents the estimated O-D matrix under a network strategy $\mathbf{a}$.

The second, a shifter of the vertically integrated orders component, is the weighted number of vertically integrated orders, where again we use county population instead of the estimated number of orders:

$$\hat{X}_{vi}^{j,j'} = \sum_{t=t(j)}^{t(j')} \beta^t \left(\sum_{i=1}^{I} \sum_{l=1}^{L} \text{Pop}_{it} \hat{\Omega}^{\text{sc}}_{i,l}\big(N_t|\mathbf{a}^0\big) - \text{Pop}_{it} \hat{\Omega}_{i,l}\big(N_t|\mathbf{a}^{j,j'}\big) \right).$$

Shifters of gross profit differences include the differences in the two average input prices (wages and rent):

$$\Delta\text{Input prices}^{j,j'} = \sum_{t=t(j)}^{t=t(j')} \beta^t \frac{1}{n_t^0} \big(\text{IC}_t\big(N_t|\mathbf{a}^0\big) - \text{IC}\big(N_t|\mathbf{a}^0\big)\big),$$

where $\text{IC}_t(N_t|\mathbf{a})$ is the sum of the input cost (either wage or rent) across active clusters under strategy $\mathbf{a}$. Note the input cost of a cluster is the average across the facilities in that cluster, if those facilities are in different counties. When all facilities within a cluster are in the same county, then it is just the input cost in that county.

The shifter of the density cost is difference in average population density:

$$\Delta\text{Density}^{j,j'} = \sum_{t=t(j)}^{t=t(j')} \beta^t \frac{1}{n_t^0} \big(\bar{\text{Dens}}_t\big(N_t|\mathbf{a}^0\big) - \bar{\text{Dens}}\big(N_t|\mathbf{a}^0\big)\big),$$

where $\bar{\text{Dens}}_t(N_t|\mathbf{a})$ is the sum of the population density across active locations.

Finally, as an additional shifter of the gross profit difference, we construct the difference in the population-weighted average tax rate between strategies $\mathbf{a^0}$ and $\mathbf{a}^{j,j'}$:

$$\Delta\text{Tax}^{j,j'} = \sum_{t=t(j)}^{t=t(j')} \beta^t \left(\sum_{i=1}^{I} \frac{\text{Pop}_{it}}{\text{Total pop}_t} \cdot \big(\tau_{it}\big(\mathbf{a}^0\big) - \tau_{it}\big(\mathbf{a}^{j,j'}\big)\big) \right),$$

where $\tau_{it}(\mathbf{a})$ is the tax rate charged on Amazon in county i under strategy $\mathbf{a}$.

Table BI formally defines the indicator variables that we use as instruments in the estimation, as well as the number of permutations associated with each one. The last two columns report the number of swaps used for each moment. Overall, we use 79% of all available swaps in the estimation. The (+) and (−) signs beside each entry indicate the sign of the first and second variables generating the trade-offs. Recall that differences are expressed relative to the rejected option of opening facility j' early (and delaying j).

TABLE BI

DEFINITION OF MOMENT CONDITIONS.

Trade-offs	Variables	Nb. Swaps: $Z_r^{j,j'} = 1$	
Distance & Taxes	$\hat{X}_d^{j,j'}$ & $\Delta\widehat{\text{Tax}}^{j,j'}$	83 (−/+)	1857 (+/-)
Distance & Cost	$\hat{X}_d^{j,j'}$ & $\Delta\widehat{\text{Input prices}}^{j,j'}$	648 (−/+)	2282 (+/-)
VI & Taxes	$\hat{X}_{vi}^{j,j'}$ & $\Delta\widehat{\text{Tax}}^{j,j'}$	133 (+/+)	967 (−/-)
VI & Cost	$\hat{X}_{vi}^{j,j'}$ & $\Delta\widehat{\text{Input prices}}^{j,j'}$	1037 (+/+)	1307 (−/-)
Distance & VI	$\hat{X}_d^{j,j'}$ & $\hat{X}_{vi}^{j,j'}$	606 (−/-)	1883 (+/+)
Distance & Density	$\hat{X}_d^{j,j'}$ & $\hat{X}_p^{j,j'}$	566 (−/+)	2100 (+/-)
Taxes alone	$\Delta\widehat{\text{Tax}}^{j,j'} + \lvert\hat{X}_d^{j,j'}\rvert < \sigma_{vi}$ & $\lvert\hat{X}_{vi}^{j,j'}\rvert < \sigma_d$	119 (+)	989 (−)
Fraction of swaps		0.79	

A positive sign indicates that the chosen option leads to an increase in the variable of interest.

The first row demonstrates the trade-off between distance and tax. We observe 83 swaps leading to negative distance and positive tax changes, and 1857 swaps leading to an increase in distance and a tax decrease. Note that we observe a larger number of distance increase/tax decrease swaps because most new entries in dense areas take place late in the sample, and generate little tax changes. In contrast, we observe a large number of swaps generating trade-offs between cost and distance (second row).

The trade-offs associated with the expansion of the sortation network correspond to swaps leading to positive differences in vertical integration and tax (or cost), and negative differences in vertical integration and tax or cost. These swaps correspond to cases in which Amazon chose to increase the number of vertically integrated transactions at the expense of lower revenue or higher input prices.

The next two trade-offs capture the joint variation in distance and vertical integration, and distance and population density. Instruments associated with choices that increase both distance and vertical integration impose restrictions on the ratio of θ_d over θ_{vi}. The trade-off between distance and population density similarly restricts the ratio of θ_d and the effect of density on fixed costs (κ).

Finally, we include a pair of moments associated with variation in taxes, and minimal changes in distance and vertical integration (i.e., difference is smaller than the interquartile range in both variables). These additional moments produce positive and negative changes in gross profits, which impose restrictions on the scale of the fixed-cost parameter (κ).

B.3. *Model Fit*

We present the fit of the model's predictions of the roll-out in Figure B1, which displays information about the timing of observed first entry into a state (shade) and the corresponding predictions of our model (points).

For the low-demand state (1999), the model accurately predicts the regional distribution of active fulfillment centers, as the predicted optimal configuration includes one location each in the west (Nevada), the center (Kentucky), and the east (Massachusetts; note, however, that the observed network's location is in Delaware). As demand increases (2006 and 2012), the density and number of facilities expand rapidly. In 2006, the model predicts

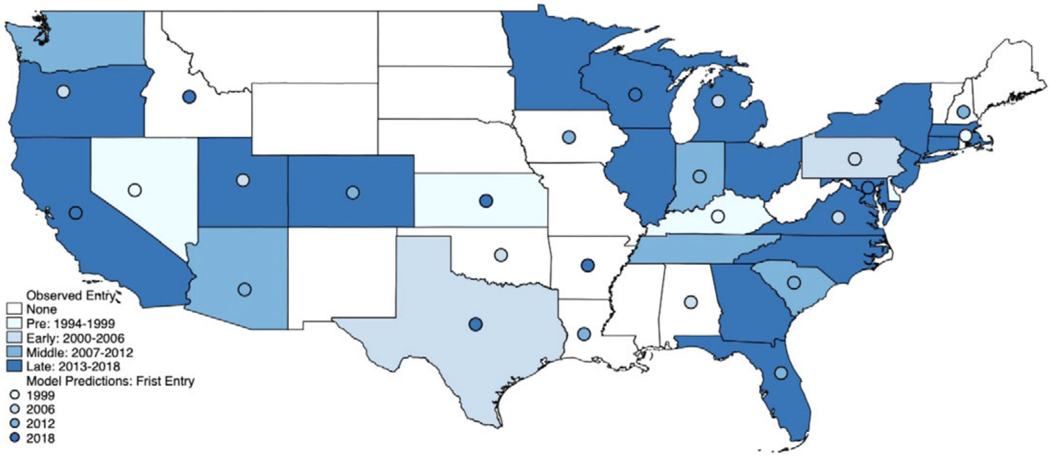

FIGURE B1.—Observed (shade) and model predicted (dot) first entry into state. Notes: The dots are places at the centroid of the state and not at a particular location.

a slightly more dense network than the actual network, with locations in additional states in the mid-west (Michigan), the southeast (Alabama), the mid-Atlantic (Virginia), and the west (Oregon), though the model accurately predicts entry on the east coast (Pennsylvania) and south (Oklahoma). However, the observed entry in the south was in Texas rather than Oklahoma.

The 2012 network is similar to the observed one, with expansions in the mid-west (Indiana), south (South Carolina), and west (Arizona). The model also predicts entry into Florida and Colorado during this time, which is slightly earlier than what we observe.

For 2018, further entry is predicted in the west (California), mid-Atlantic (Maryland), and mid-west (Wisconsin), which are all entries that are observed. However, there are some states which the model misses in the mid-west (Illinois), east (New York), and southeast (Georgia).

REFERENCES

BLUNDELL, RICHARD, LUIGI PISTAFERRI, AND IAN PRESTON (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98, 1887–1921. [12]

DE LOS SANTOS, BABUR, ALI HORTAÇSU, AND MATTHIJS R. WILDENBEEST (2012): "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 102, 2955–2980. [3]

Co-editor Liran Einav handled this manuscript.

Manuscript received 18 April, 2017; final version accepted 13 April, 2022; available online 23 June, 2022.