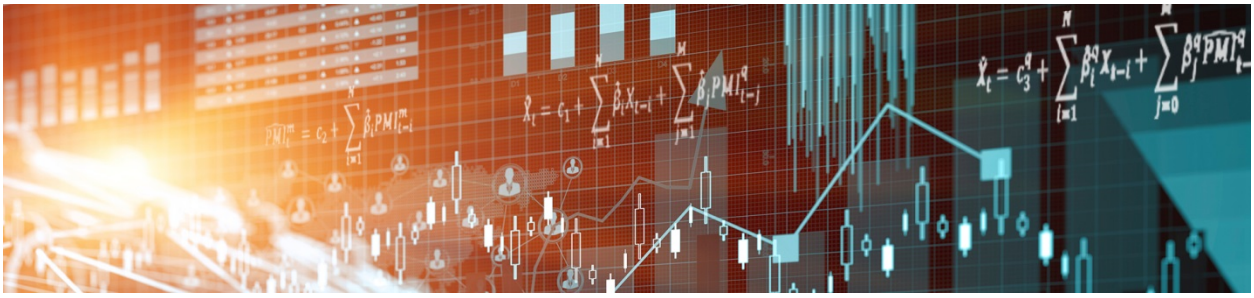
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2019-42

Amazon Effects in Canadian Online Retail Firm-Product-Level Data



by Alex W. Chernoff

Bank of Canada Staff Working Paper 2019-42

October 2019

Amazon Effects in Canadian Online Retail Firm-Product-Level Data

by

Alex W. Chernoff

International Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
achernoff@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

The views in this paper are those of the author and do not necessarily reflect those of the Bank of Canada. This paper benefited from the comments of participants at the NBER Summer Institute 2018 Workshop on IT and Digitization and the annual conferences of the European Association for Research in Industrial Economics and the Canadian Economics Association. I gratefully acknowledge helpful comments provided by Geoffrey Dunbar and Nicolas Schmitt, and research assistance provided by Amy Li. I acknowledge financial support provided by the Productivity Partnership and by Innovation, Science and Economic Development Canada. The firm-level analysis in this paper has been vetted by Statistics Canada and has passed the compliance requirements of the agency's confidentiality protection and disclosure policy.

Abstract

I use firm-product-level data for Canadian online retailers to study how product scope (the average number of product categories per firm) evolved from 1999 to 2012. During this period, product scope dropped monotonically from 59 to 5 product categories. Using a theoretical model of multi-product firms, I show that this reduction can be rationalized by increased online competition. Consistent with the model, I find that the percentage of Canadian online retailers with revenues in a product category falls when Amazon.com expands its varieties in the category. Overall, Amazon.com's expansion accounts for 37 percent of the observed reduction in product scope.

Bank topics: Firm dynamics; Service sector
JEL codes: D22, L11, L81

Résumé

J'exploite des données sur les produits vendus par les détaillants canadiens en ligne afin d'étudier comment la gamme de produits (le nombre moyen de catégories de produits offerts par entreprise) a évolué entre 1999 et 2012. Pendant cette période, la gamme a diminué de façon monotone, passant de 59 catégories à 5. En m'appuyant sur un modèle théorique de firmes multiproduits, je montre que cette réduction peut s'expliquer par l'intensification de la concurrence en ligne. Je constate que là où Amazon.com élargit son offre dans des catégories de produits, la proportion de détaillants canadiens en ligne qui affichent des revenus dans ces catégories recule, ce qui concorde avec les résultats du modèle. Dans l'ensemble, l'expansion de l'offre d'Amazon.com explique 37 % de la réduction de la gamme de produits observée.

Sujets : Dynamique des entreprises; Secteur des services
Codes JEL : D22, L11, L81

Non-Technical Summary

Central banks are interested in understanding the effects of digitization on the global economy. In this paper I use micro data for Canadian online retailers to analyze how their average revenues and product offerings evolved from 1999 to 2012. I find that the average number of products offered by these retailers decreased dramatically over this period, and I study the extent to which this specialization was the result of increased competition from Amazon.com.

The firm-product-level online retailer data used in this paper are from Statistics Canada's Annual Non-Store Retail Survey (RTNS), which is designed to collect information on all Canadian retailers that do not operate brick-and-mortar retail stores. I study the subset of online retailers in the RTNS over the period 1999–2012. Importantly for purposes of my empirical analysis, Amazon.com and Amazon.ca (Amazon's Canadian website) are not in the RTNS survey population. Online retailers' revenues are reported across 115 different product categories in the RTNS. I decompose the revenue of online retailers into the contributions of product scope (the number of product categories) and product scale (average revenue per product category). During this period of rapid online sales growth, product scope dropped dramatically from an average of 59 product categories per firm in 1999 to 5 product categories per firm in 2012.

I show that this result can be rationalized in a theoretical multi-product firm model by increased online competition. I then use data on Amazon.com to study whether the reduction in Canadian online retailers' product scope over the period 1999–2012 was partially driven by increased competition from Amazon.com. In particular, I use data from Amazon.com reviews spanning 1996–2014, including 83.68 million reviews of 9.4 million Amazon.com products. I create a correspondence that maps these products into the 115 RTNS product categories.

Using these data and guided by the theoretical model, I develop an econometric framework to study how an expansion in Amazon.com's varieties affects Canadian online retailers. I find that the percentage of Canadian online retailers with revenues in a product category falls when Amazon.com expands its varieties in the category. Furthermore, my results indicate that the effect of Amazon.com's expansion in a particular product category gets larger with Amazon.com's overall growth in product varieties across all categories. Finally, I find that Amazon.com's product expansion accounts for 37 percent of the observed reduction in product scope of Canadian online retailers during my study period.

This paper contributes to the literature on the online retail sector by estimating Amazon.com's

effect on the product offerings of other online retailers. While economists are interested in understanding the effect of Amazon's growth, a lack of data availability has limited research on this topic. I contribute to this field by using Amazon data that cover a much longer period than existing studies, and by focusing on Amazon's effect on other online firms' product offerings. In this respect, I also contribute to the literature on multi-product firms. This literature has often focused on international trade, with several papers finding that firms reduced the number of products they produce during periods of trade liberalization. My paper is the first to document a similar phenomenon in the online retail sector, notably that Canadian online firms responded to increased competition from Amazon.com by reducing their product scope.

I Introduction

Since it was founded in 1994, Amazon has been a leader in online retail sales in North America. Economists are interested in understanding the effect of Amazon's growth on issues ranging from competition to inflation (Cavallo, 2018). The paucity of data on Amazon and its online competitors, particularly during the late 1990s and early 2000s, has limited research studying how Amazon's dominance has affected competing online retailers. In this paper I use firm-product-level data on Canadian online retailers and Amazon to study this issue.

The data used in this paper are from Statistics Canada's Annual Non-Store Retail Survey (RTNS), which is designed to collect information on all Canadian retailers that do not operate brick-and-mortar retail stores. I study the subset of online retailers in the RTNS over the period 1999–2012. Importantly for my empirical analysis, Amazon.com and Amazon.ca (Amazon's Canadian website) are not in the RTNS survey population. Online retailers' revenues are reported across 115 different product categories in the RTNS. I decompose the revenue of online retailers into the contributions of product scope (the number of product categories) and product scale (average revenue per product category). During this period of rapid online sales growth, product scope dropped dramatically from an average of 59 product categories per firm in 1999 to 5 product categories per firm in 2012.

Using a version of the Mayer et al. (2014) multi-product firm model adapted to my empirical setting, I show that this result can be rationalized by increased online competition. I incorporate Amazon.com in the model as a third-party sales platform and competing firm, whose prices and product offerings are taken as given by Canadian online firms. The model predicts that Amazon.com's role as a third-party sales platform may initially increase Canadian online retailers' product scope by lowering the fixed cost of selling online. However, Canadian online retailers' product scope falls as Amazon.com expands its product offerings, and this effect is larger when Amazon.com's prices are more competitive.

I use web-scraped data to study whether the reduction in Canadian online retailers' product scope over the period 1999–2012 was partially driven by increased competition from Amazon.com. In particular, I use data from Amazon.com reviews from McAuley et al. (2015) and He and McAuley (2016), which span 1996–2014 and include 82.68 million reviews of 9.4 million Amazon.com products. I create a correspondence that maps these products into the 115 RTNS product categories. I also map the expansion of Amazon.ca into the RTNS product categories, using web-scraped

Amazon press release data and historic screen shots from *Internet Archive Wayback Machine*. My econometric analysis focuses on identifying the effect of Amazon.com's expansion, as the data on Amazon.ca are less detailed and potentially endogenous (owing to the possibility of strategic interaction between Amazon.ca and Canadian online retailers).

Using these data and guided by the theoretical model, I develop an econometric framework to study how an expansion in Amazon.com's varieties affects Canadian online retailers. I find that the percentage of Canadian online retailers with revenues in a product category falls when Amazon.com expands its varieties in the category. My results indicate that the effect of Amazon.com's expansion in a particular product category increases with Amazon.com's overall growth in product varieties across all categories. Finally, I find that Amazon.com's product expansion accounts for 37 percent of the observed reduction in product scope of Canadian online retailers during my study period.

This paper contributes to the literature on the online retail sector by estimating Amazon.com's effect on Canadian online retailers. My focus is similar to that of Cavallo (2018), whose results suggest that competition from Amazon has increased the frequency of price changes and price uniformity across stores for multi-channel retailers in the US. My work differs in that I focus on Amazon.com's interaction with other online retailers using data that cover a much longer period, dating back to the inception of e-commerce in the 1990s. Moreover, my focus on product scope differentiates my research from Cavallo (2018) and links my work to a much earlier literature on the "long-tail" theory in e-commerce. Popularized by Chris Anderson in a 2004 article in *Wired* magazine, this theory argues that products that sell at low volume collectively account for a large fraction of the overall sales of many products sold online.[1] An important distinction between the long tail literature and my analysis is the level of product aggregation. I define product scope at a relatively coarse level of product aggregation, whereas studies in the long tail literature typically use highly disaggregated product data (Brynjolfsson et al., 2010). Companies such as Netflix and Spotify typify online firms that offer extensive variety yet operate in a limited number of categories at a higher level of product aggregation. My results indicate that at this coarser level of product aggregation, Canadian online retailers responded to increased competition from Amazon.com by reducing product scope. Through the lens of my adapted version of the Mayer et al. (2014) model, I show that this response can be rationalized by firms focusing on their products of core competency in response to increased competition. While several papers in the multi-product firm literature

[1]Evidence supporting this theory is provided by Brynjolfsson et al. (2003), who note that Amazon.com carries upwards of 23 times as many book titles as the typical brick-and-mortar bookstore, while Brynjolfsson et al. (2006) report that 30–40% of Amazon.com's book sales are in titles that are not normally carried by brick-and-mortar stores.

have found evidence of similar firm-level behavior in the context of international trade,[2] to my knowledge my paper is the first to document this result for online retailers.

II Data and Descriptive Statistics on Product Scale and Scope

A Firm-Product-Level Data for Canadian Online Retailers

I use firm-product-level data on Canadian online retailers from Statistics Canada's RTNS, 1999–2012. The sample covers online retailers and other non-store retailers such as mail-order houses, telemarketing firms, and vending machine operators. I define a sample of "online firms" as the subset of RTNS firms that are classified under NAICS category 4541 (electronic shopping and mail-order houses) and have positive online sales.[3]

For 2012, the aggregate total operating revenue of firms in my sample represents 51.0% of Canadian e-commerce sales and 0.8% of total sales for the Canadian retail sector.[4] A firm must have a Canadian headquarters to be included in the RTNS, which implies that Amazon.com and Amazon.ca are excluded from the survey population.

B Scale-Scope Decomposition of Online Retail Revenues

The RTNS data include a commodity annex that details the distribution of each firm's revenue across 115 different product categories.[5] These data can be used to study how the growth in online retailers' revenues during the period 1999–2012 was driven by changes in product scope and product scale.

Drawing on the empirical literature on multi-product firms,[6] I use a log decomposition of

[2]For example, Bernard et al. (2011), Baldwin and Gu (2009), and Iacovone and Javorcik (2008) find that firms reduced the number of products they produce during periods of trade liberalization in the US, Canada, and Mexico, respectively.

[3]According to Statistics Canada's documentation, the RTNS is an enumeration of *all* Canadian non-store retailers, which are identified as enterprises that are classified by NAICS sub-sector 454 (Non-Store Retail Trade Sector). However, the average annual number of observations in my sample is only 205, which suggests that the RTNS does not cover the entire population of all online retailers in Canada. In particular, the RTNS is unlikely to capture online retailers that are too small to be captured in Statistics Canada's Business Register. For the purposes of defining online retailers, I include internet sales and sales from electronic auctions in my definition of online sales.

[4]Aggregate statistics for the Canadian retail sector (NAICS 44–45) are sourced from CANSIM table 080–0032.

[5]The taxonomy of the RTNS commodity annex was developed by Statistics Canada. A description of the commodity classification is available at: `http://www23.statcan.gc.ca/imdb-bmdi/document/2448_D1_T1_V1-eng.pdf`

[6]Similar decomposition exercises are used in Arkolakis and Muendler (2010), Goldberg et al. (2010), and Bernard et al. (2010).

product scope and scale:

$$ln\left(\bar{pq}_t\right) = ln\left(\bar{Scope}_t\right) + ln\left(\bar{pq}_t/\bar{Scope}_t\right), \tag{1}$$

where $\bar{pq}_t$ is the annual average revenue of online retailers in year t; and $\bar{Scope}_t$ is product scope, defined as the mean count of product categories with positive revenue per firm, averaged across all firms in each year. I define product scale as the average annual operating revenue per firm divided by product scope, $\bar{pq}_t/\bar{Scope}_t$.

Figure 1a and Table 1 show the scale and scope decomposition in equation (1) for my sample of Canadian online retailers. Product scope declined monotonically over this period, from 59 in 1999 to 5 in 2012. The average revenue of online retailers more than doubled over the observation period, growing from 6.5 million in 1999 to 17 million in 2012. With revenues increasing and scope declining, product scale increased dramatically from approximately 111,000 per product category in 1999 to 3.5 million per product category in 2012.

One potential explanation for the decline in product scope is a possible increase in the number of small, single-product firms that were surveyed over time. To investigate this possibility, I recalculate the scale-scope decomposition in each year separately for the sub-samples of large and small firms. For the purposes of this robustness exercise, large firms are defined as the subset of firms with revenue above the annual median value, and small firms are those with revenue equal to or less than the annual median value. The results, presented in figures 1b and 1c, show that the decline in product scope is apparent for small and large firms alike.

C Amazon Data

Existing research suggests that increased competition encourages multi-product firms to skew sales towards their products of core competency (Mayer et al., 2014). In particular, the expansion of the dominant online retailer in North America, Amazon.com, may have affected the product scope of Canadian online retailers.

To study the effect of Amazon on Canadian online retail firms, I use web-scraped data on Amazon.com and Amazon's Canadian website, Amazon.ca. For Amazon.com I use the product review data from McAuley et al. (2015) and He and McAuley (2016).[7] These data include 82.68 million reviews of 9.4 million Amazon.com products, spanning 1996–2014. Analyzing the effect of

[7]These data are available for research purposes from Julian McAuley's website: http://jmcauley.ucsd.edu/data/amazon/. For my analysis I use the "ratings only" version of these data.

Amazon.com on Canadian online retailers requires creating a many-to-one correspondence between the Amazon.com products and Statistics Canada's product classification taxonomy. The Statistics Canada product classification is from the commodity annex of the RTNS, which includes 115 different product codes. By making use of the hierarchical structure of Amazon.com's product categories, I am able to map 95% of the Amazon.com reviews to the Statistics Canada product classification. I approximate the number of Amazon.com products in each Statistics Canada product category in each year by counting the number of unique Amazon.com products in the category that were reviewed at least once during the calendar year.

For Amazon.ca, I do not have reviews or any other source that can be used to approximate the number of products by category. Nevertheless, I am able to determine the year that Amazon.ca first started selling products in each Statistics Canada product category. From Amazon's historic press releases I document the opening of every Amazon.ca "store" during my sample period. This allows me to identify the year of entry of Amazon.ca into several Statistics Canada product categories, but for many others the information in the press release is not detailed enough. I therefore complement the press release data with historic screen shots of Amazon.ca from *Internet Archive Wayback Machine*. Using these two sources of data, I document the year of Amazon.ca's entry into the product categories of the RTNS.

Figure 1d illustrates that the expansion in product scope of Amazon.com and Amazon.ca sharply contrasts with the narrowing scope of Canadian online retailers over the period 1999–2012. Competition from Amazon may have forced Canadian online retailers to narrow their product scope; however, Amazon's third-party sales platform (i.e. Amazon Marketplace) may have also encouraged product scope expansion by facilitating Canadian online retailers' access to a larger market at a lower cost.[8] In the next section I present a modified version of the Mayer et al. (2014) multi-product firm model that can be used to conceptualize how these different mechanisms may have affected Canadian online retailers' product scope.

III Amazon Effects in a Model of Multi-Product Firms

In this section I present a version of the Mayer et al. (2014) model that is adapted to study how Amazon.com's expansion may have affected the product scope of Canadian online retailers. Specifically, I incorporate three changes to customize the Mayer et al. (2014) model to address this

[8] Amazon reports that third-party selection currently accounts for more than 50% of its total unit sales (Amazon, 2019).

research question. First, I introduce an additional firm representing Amazon.com, whose prices and product scope are exogenously determined irrespective of the decisions of Canadian online retailers. This assumption captures the idea that Amazon.com focuses on the US market and does not strategically respond to the behavior of Canadian online retailers when setting prices and its product scope.[9] Second, I include a fixed cost in the online retail firm's production function. This fixed cost captures the idea that selling online entails fixed costs (e.g. website design) that will be lower if the firm uses a third-party sales platform like Amazon Marketplace.[10] Third, I consider a short-run equilibrium of the model, using Melitz and Ottaviano's (2008) concept of the short run as an equilibrium where the measure of firms and the online retailer cost distribution are fixed. I consider a short-run equilibrium since my study period is characterized by rapid growth in online retail markets, and therefore a long-run analysis is unlikely to accurately reflect firms' behavior.[11] As in Mayer et al. (2014), the model is cast in a static environment. This entails an assumption that Canadian online retailers can optimally respond on a period-by-period basis to changing market conditions, including those resulting from changes in Amazon.com's prices and product offerings.

A Consumer Preferences and Demand

There are L consumers with identical preferences for a homogeneous good (y), and a continuum of differentiated varieties indexed by $i \in \Omega$. The representative consumer has the following utility function:

$$U = y + \alpha \int_{i \in \Omega} q_i^c di - \frac{\gamma}{2} \int_{i \in \Omega} (q_i^c)^2 di - \frac{\eta}{2} \left(\int_{i \in \Omega} q_i^c di \right)^2,$$

where q_i^c is the individual consumption of variety i, and the demand parameters α, γ, and η are all positive. Thus, the household side of the model is identical to Mayer et al. (2014), and therefore demand for each variety is the same as in their model.

[9] As discussed in the next section, this assumption is also central to my econometric identification strategy. I acknowledge that it is a less plausible assumption in the case of Amazon.ca, and I interpret my empirical analysis of Amazon.ca's expansion as conditional correlations rather than causal effects.

[10] I do not model the firm's decision to sell via Amazon Marketplace, but compare the model equilibrium to a counterfactual without Amazon where fixed production costs are higher. Analyzing the decision of online retailers to sell on a third-party platform like Marketplace is not possible with my data; however, I highlight this as an interesting topic for future research.

[11] By focusing on a short-run equilibrium, I abstract from questions related to Amazon's effect on the entry and exit of other online retailers. These questions cannot be studied with the RTNS data, as firm births and deaths cannot be inferred from firms entering and exiting the survey. The inability of the RTNS to capture small firms, the transition of firms out of the survey population due to changes in their NAICS classification by Statistics Canada, and other factors all confound efforts to study entry and exit with the RTNS data. Although I am unable to study this question with my data, I highlight this as an important topic for future research.

B Online Retailers

The assumed market structure is monopolistic competition, and the production and behavior assumptions are summarized by the online retailer's profit-maximization problem:

$$\pi(\nu) = \max_{p(\nu)} p(\nu)q(\nu) - \nu q(\nu) - f, \quad subject\ to: \ q(\nu) = \frac{\alpha L}{\eta M + \gamma} - \frac{L}{\gamma}p(\nu) + \frac{\eta M}{\eta M + \gamma}\frac{L}{\gamma}\bar{p}.$$

Firms choose the price, $p(\nu)$, that maximizes profit for each variety produced, subject to demand, $q(\nu)$, taking the aggregate measure of varieties, M, and average price, $\bar{p}$, as exogenously given. The marginal cost of producing the m^{th} variety for a firm with cost parameter c is defined as $\nu = \nu(m, c) = \omega^{-m}c$, with $\omega \in (0, 1)$. The firm's core variety is defined by $m = 0$, and the firm's marginal cost increases with each incremental variety that is produced, $m \in \{1, 2, ...\}$.

Up until this point, everything in the model is identical to Mayer et al. (2014). I depart from their model by assuming that firms pay a fixed cost of production, $f > 0$. I also assume that the total measure of varieties includes varieties produced by Amazon.com, M_A, and varieties produced by other online retailers, M_o, such that $M = M_A + M_o$. M_A is assumed exogenous, and M_o is the endogenous outcome of interest in the model. The average price also incorporates Amazon.com and the other online retailers' prices, and is defined $\bar{p} = (M_A/M)\,\bar{p}_A + (M_o/M)\,\bar{p}_o$. That is, $\bar{p}$ is the weighted sum of Amazon.com's average price across all its varieties, $\bar{p}_A$, and the other online retailers' average price across all varieties, $\bar{p}_o$. The weights in this sum are equal to Amazon.com's and the other retailers' respective shares of total varieties. Amazon.com's average price is assumed to be exogenous, while other online retailers' average price is endogenously determined by the firm's profit-maximization problem.

C Short-Run Equilibrium

Consistent with Melitz and Ottaviano (2008), I consider a short-run equilibrium where the measure of firms and the online retailer cost distribution are fixed. I normalize the measure of firms to 1 so that M_o is both the aggregate and the average measure of online retailers' varieties, and the theoretical analogue to the measure of product scope in equation (1). Henceforth, I refer to M_A and M_o as the product scope of Amazon.com and (other) online retailers, respectively.

The solution to a representative firm's profit-maximization problem implies that $\pi(\nu) = (L/\gamma)(p(\nu) - \nu)^2 - f$. Define c_D as the cost cutoff such that a retailer with this cost parameter makes zero profits from producing its core variety, and firms with $c > c_D$ shut down. I assume that $\bar{p}_A < c_D$, implying

that Amazon.com's average price is lower than the marginal cost of the least efficient online retailer. The cutoff cost c_D defines the zero cutoff profit condition $\pi(\nu(0, c_D)) = 0$, which can be inverted to solve for online retailers' short-run equilibrium product scope:

$$M_o = \frac{2\gamma}{\eta} \frac{\left(\alpha - c_D - 2\sqrt{\gamma f/L}\right)}{\left(c_D - \bar{\nu}_o + 2\sqrt{\gamma f/L}\right)} - \frac{M_A}{2} \frac{\left(c_D - \bar{p}_A + 2\sqrt{\gamma f/L}\right)}{\left(c_D - \bar{\nu}_o + 2\sqrt{\gamma f/L}\right)}, \tag{2}$$

where $\bar{\nu}_o$ is the average marginal cost of all varieties produced by online retailers, which is exogenously determined by the marginal cost function and the firm cost distribution.

Define $M_o^{'}$ as the counterfactual product scope of online retailers without Amazon.com. In this counterfactual, Amazon.com's product scope is zero, $M_A^{'} = 0$, and I assume that the fixed cost of online retailing is higher, $f^{'} > f$, capturing the idea that Amazon.com's role as a platform for third-party sales reduces this cost. I also assume that the parameters of the model are such that online retailers' product scope is positive in both the equilibrium with and without Amazon.com.[12]

Figure 2 plots the online retailers' product scope in equation (2) as a function of M_A, for the equilibrium with and without Amazon.com. Note that when Amazon.com's product scope is zero, online retailers' product scope is *higher* in the equilibrium with Amazon.com, $M_o{'} < M_o$. Intuitively, if Amazon.com were to exist only as a third-party sales platform, then other online retailers would have a larger product scope as a result of the lower production cost. However, as Amazon.com's product scope grows, its competitive influence reduces the product scope of the other online retailers. Furthermore, the reduction in other online retailers' product scope is larger when Amazon.com's prices are more competitive (that is, when $c_D - \bar{p}_A$ is large). Therefore, a second prediction of the model is that growth in Amazon.com's product scope, reductions in its prices, and the interaction of these two factors all exert a negative effect on the product scope of other online retailers.

IV Econometric Specification

The theoretical model of the previous section suggests that Amazon.com's expansion may have contributed to the observed decline in product scope of Canadian online retailers, as documented

[12] In the equilibrium without Amazon.com, this assumption requires that $\alpha - c_D - 2\sqrt{\gamma f/L} > 0$. In the model with Amazon.com, the assumption that $M_0 > 0$ also requires that Amazon.com's product scope is not too large, and that its average price is not too small, to avoid having these competitive effects drive other retailers' product scope to zero.

in section II. In this section I describe how I use data from Amazon.com and the RTNS to test this prediction.

The starting point for my econometric analysis is to postulate that the probability that a Canadian online retailer has revenues in any particular product category j in year t can be written as $P(pq_{i,j,t} > 0) = f(Amazon_t, X_{j,t})$. The term $Amazon_t$ summarizes the different dimensions along which Amazon.com affects this probability, and $X_{j,t}$ captures the other relevant product and time varying factors. The theoretical model indicates that the number of varieties that Amazon.com carries in product category j is an important determinant of this probability. I therefore construct the variable $Variety_{A,j,t}$ from the Amazon.com product review data described in section II. Specifically, I define this variable as the count of unique Amazon.com products reviewed in year t belonging to RTNS product category j. Note that this is a more disaggregated measure of Amazon.com's product variety relative to the product scope measure considered in section II. The measure of Amazon.com's product scope in Figure 1d, panel d, counts the number of RTNS product categories where Amazon.com has at least one reviewed product, whereas the variable $Variety_{A,j,t}$ counts the number of reviewed products *within* each RTNS product category.

The theoretical model suggests that Amazon.com's price competitiveness is a second factor that affects the product scope of Canadian online retailers. Unfortunately, I do not have price data for Amazon.com. To address this data limitation, I draw on an insight from Bernard et al.'s (2011) multi-product firm model, which suggests that the range of products a firm sells across *all* product categories is positively correlated with the firm's ability, and negatively correlated with the firm's prices. This motivates a specification using the count of Amazon.com's reviewed products across all RTNS categories in a given year, $Variety_{A,t}$, to approximate Amazon.com's level of competitiveness. Intuitively, this measure captures the idea that company-wide efficiency improvements that led to Amazon.com's growth, such as improvement in its distribution network, make the company more competitive overall as well as within each product category.

I formalize these ideas in the following linear probability model:

$$1(pq_{i,j,t} > 0) = \beta_1 ln(Varieties_{A,j,t}) + \beta_2 ln(Varieties_{A,j,t}) \times ln(Varieties_{A,t}) + \delta_j + \alpha_t + \epsilon_{i,j,t}, \quad (3)$$

where the left-hand-side variable is an indicator that is 1 if firm i had revenues in product category j in year t. To avoid having an undefined value of $ln(Varieties_{A,j,t})$ in categories where Amazon.com has not entered, I add 1 to $Varieties_{A,j,t}$ before taking the log. I demean $ln(Varieties_{A,t})$ so that β_1

is the effect of a marginal change in the log of Amazon.com's varieties in product category j, when the variable $ln(Varieties_{A,t})$ is at its sample mean value. Finally, I include product indicators, δ_j, and year indicators, α_t, to control for other sources of time and product variation in a flexible manner. As the right-hand-side variables of interest vary only across products and time, I aggregate the model to the product-year level by averaging equation (3) across all firms, yielding

$$share_{j,t} = \beta_1 ln(Varieties_{A,j,t}) + \beta_2 ln(Varieties_{A,j,t}) \times ln(Varieties_{A,t}) + \delta_j + \alpha_t + u_{j,t}, \quad (4)$$

where the dependent variable, $share_{j,t}$, is the share of Canadian online firms with positive revenues in RTNS product category j in year t,[13] and the other variables are unchanged from equation (3). I drop the RTNS product categories for "used/second hand merchandise" and "other" goods because it is difficult to accurately determine the Amazon.com products that belong to this category. I also exclude two RTNS product groups, "services" and "automotive fuels, oils, and additives," because these include products that cannot be sold online (e.g. lunch counter items and gasoline).[14] All regressions are estimated by ordinary least squares with standard errors clustered at the product level.

Since the outcome variable in equation (4) is expressed as a share, $\beta_1 + \beta_2 ln(Varieties_{A,t})$ is the change in the percentage of Canadian online firms in a product category in response to a 1% increase in Amazon.com's product variety in the category. Because $ln(Varieties_{A,t})$ is centered, the estimate of this marginal effect is simply equal to β_1 when the log of Amazon.com's total variety is at its sample mean value.

The key identifying assumption in estimating equation (4) is that Amazon.com's product variety decisions are exogenously determined irrespective of the decisions of Canadian online retailers. Although some Canadians purchase from Amazon.com, it is reasonable to assume that Amazon.com focuses on the much larger US market and its US competitors when considering its product offerings. More generally, equation (2) in the theoretical model indicates that the product scope of Canadian online retailers is determined by the cost parameters (c_D, $\bar{\nu}_o$, and f), and the population demanding

[13]The variable $share_{j,t}$ is calculated using a custom tabulation prepared by Statistics Canada at my request. This custom tabulation includes the count of online retailers in each RTNS product category in every year over the period 1999–2012. Counts were not released for product categories with one to two firms to meet the compliance requirements of the agency's confidentiality protection and disclosure policy. In a limited number of cases, product categories with three firms were also not released in accordance with this policy. This resulted in a total of 125 observations being censored. These observations are excluded from the sample used to estimate equation (4).

[14]Including these dropped product groups has very little effect on the magnitude of my estimates, and all of my qualitative conclusions are robust to this sensitivity check. These results are available upon request.

their products (L).[15] Any variation in these parameters that is correlated with Amazon.com's product variety decisions is a potential source of omitted variable bias. The inclusion of the time and product fixed effects in regression equation (4) respectively control for any such variation that is period-specific and constant across products, or product-specific and constant over time.

I acknowledge the assumption that Amazon's product variety decisions are more likely to be endogenous in the case of Amazon.ca. This endogeneity concern, and better data availability for Amazon.com, are the main reasons I focus on the effects from Amazon.com rather than Amazon.ca in my econometric analysis. It is nevertheless interesting to consider the conditional correlation between Amazon.ca's product-level expansion and the share of Canadian retailers in each RTNS product category. For this reason I also report estimates replacing $ln(Varieties_{A,j,t})$ with an indicator variable, $d_{ca,j,t}$, that is 1 if Amazon.ca has entered product category j in year t. This variable is calculated using the data on Amazon.ca described in section II. As a basis for comparing estimates for Amazon.com and Amazon.ca, I also report a specification that replaces $ln(Varieties_{A,j,t})$ with an indicator variable, $d_{com,j,t}$, that is 1 if Amazon.com has at least one reviewed product in product category j in year t.

V Results

The results of estimating equation (4) are reported in Table 2. The baseline results reported in column 2 indicate that the percentage of Canadian retailers in a product category falls when Amazon.com expands its variety in the category, and that the magnitude of this effect is larger as Amazon.com's overall variety grows.[16] When Amazon.com's logged number of total varieties is at its sample mean, the estimate of β_1 implies that a doubling of the number of reviewed Amazon.com products in an RTNS category reduces the participation of Canadian online retailers in the category by 1.36 percentage points.

More generally, the estimated change in the percentage of Canadian online firms in a product category in response to a percentage increase in Amazon.com's product variety in the category, $\beta_1+\beta_2 ln(Varieties_{A,t})$, increases in magnitude (i.e. becomes more negative) as Amazon.com's total

[15] I assume that households' preference parameters (α, γ, and η) are constant; however, if this is not the case, then any variation in these parameters that is correlated with Amazon.com's product variety decisions is another potential source of omitted variable bias.

[16] In column (2), the estimated coefficients are statistically significant at the 1% level for the key variables of interest: $ln(Varieties_{A,j,t})$ and $ln(Varieties_{A,j,t}) \times ln(Varieties_{A,t})$. Odd-numbered columns in Table 2 report the estimates without the interaction term $\times ln(Varieties_{A,t})$ to show the difference in the estimates from excluding this variable. In all specifications, the interaction term is highly significant, and I therefore focus my discussion on the columns that include this variable.

number of varieties grows. Figure 3 plots the estimate of this marginal effect and its 95% confidence interval against my measure of Amazon.com's total varieties. Because Amazon.com's total number of varieties grows monotonically in every year of the data from 1999 to 2012, interpreting Figure 3 from left to right also shows the evolution of this marginal effect over time. In 1999, the total number of Amazon.com varieties was approximately 89,000 and the marginal effect was -0.01059. In 2012, the effect was more than 2.5 times higher at -0.02903, reflecting the growth in Amazon.com's varieties to almost 2.5 million in 2012. This is consistent with the theoretical framework in section III, as the effect on Canadian online retailers gets larger as Amazon.com becomes a more dominant competitor, as approximated by its product variety across all categories.

Column (6) of Table 2 presents estimates under the specification that uses the data on the expansion of Amazon.ca, the company's Canadian website. As previously discussed, due to endogeneity concerns, these estimates should be interpreted as conditional correlations rather than causal effects. The estimated coefficients on the Amazon.ca indicator variable and its interaction with Amazon.com's total product variety are both negative and highly significant, providing evidence that Amazon's product expansion into Canada was correlated with the exit of Canadian firms from these categories. Column (4) runs a similar regression using the Amazon.com indicator variable to provide a basis for comparing the magnitude of the estimates in the specification using the Amazon.ca data. The coefficient estimates on the Amazon.com indicator and the interaction are both higher for Amazon.com. This suggests that Amazon.com may have had a larger effect on Canadian online retailers than Amazon.ca, although caution is warranted in drawing this conclusion as the Amazon.ca estimates may be biased due to endogeneity.

The results in Table 2 can be used to estimate Amazon's contribution to the decline in Canadian online retailers' product scope that was documented in section II. For the purposes of this quantification exercise, note that product scope is the sum of the dependent variable in the estimating equation, $share_{j,t}$, aggregated across all J categories.[17] It follows that the expression

$$\sum_{j=1}^{J}\left(\hat{\beta}_1 ln(Varieties_{A,j,2012}) + \hat{\beta}_2 ln(Varieties_{A,j,2012}) \times ln(Varieties_{A,2012})\right)$$

$$-\sum_{j=1}^{J}\left(\hat{\beta}_1 ln(Varieties_{A,j,1999}) + \hat{\beta}_2 ln(Varieties_{A,j,1999}) \times ln(Varieties_{A,1999})\right)$$

[17]From the definition of product scope, $\bar{Scope}_t = \sum_{i=1}^{N_t} 1/N_t \sum_{j=1}^{J} 1(pq_{i,j,t} > 0) = \sum_{j=1}^{J}\sum_{i=1}^{N_t} 1(pq_{i,j,t} > 0)/N_t = \sum_{j=1}^{J} share_{j,t}$, where N_t is the number of firms in my sample in year t.

is an estimate of the effect of Amazon's product expansion on Canadian online retailers' change in product scope over the period 1999 to 2012. Using this approach and the baseline estimates in column (2) of Table 2, I find that Amazon's expansion reduced Canadian online retailers' scope by 20.13 product categories. That is, Amazon.com's expansion accounts for 37% of the observed reduction in product scope from 59 to 5 product categories over this period.

VI Discussion

The results of this paper show that the average product scope of Canadian online retailers declined dramatically over the period 1999–2012. My theoretical model shows how increased competition from Amazon.com can rationalize this trend, and the empirical analysis indicates that this was indeed an important factor. My results point to a number of possible directions for future theoretical and empirical research on this topic. For example, Amazon.com's rapid expansion of product scope stands in stark contrast to the reduction of product scope of Canadian online retailers over the study period. My model and empirical analysis assume that Amazon.com's range of product offerings are exogenously determined with respect to the behavior of Canadian online retailers. While I have argued that this assumption is applicable in my setting, it is likely not valid for the US, where competition between Amazon and other US online retailers is characterized by strategic interaction. It would be interesting to know whether a similar decline in product scope has occurred for US online retailers. If so, it would be interesting to analyze whether a model with an industry leader and competitive fringe could endogenously capture the divergent choices of Amazon.com and other online retailers as an equilibrium outcome. Another open question is whether online retailers have reduced product scope and at the same time expanded the varieties within their product categories. This has been a hallmark of online service providers such as Netflix and Spotify, who specialize in video and music respectively and offer consumers an abundance of choices within these product categories. Empirical research is needed to determine whether online retailers are following a similar strategy of specializing in a narrow set of product categories while offering a wide range of varieties as means to attract customers.

References

Amazon, "Start Selling Online," Technical Report 2019.

Anderson, Chris, "The Long Tail," *Wired Magazine*, 2004, *12* (10), 170–177.

Arkolakis, Costas and Marc-Andreas Muendler, "The Extensive Margin of Exporting Products: A Firm-Level Analysis," Working Paper 16641, National Bureau of Economic Research 2010.

Baldwin, John R. and Wulong Gu, "The Impact of Trade on Plant Scale, Production-Run Length and Diversification," in T. Dunne, J.B. Jensen, and M.J. Roberts, eds., *Producer Dynamics*, Chicago: University of Chicago Press, 2009.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott, "Multiple-Product Firms and Product Switching," *American Economic Review*, 2010, *100* (1), 70–97.

_ , _ , and _ , "Multiproduct Firms and Trade Liberalization," *Quarterly Journal of Economics*, 2011, *126* (3), 1271–1318.

Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Michael D. Smith, "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science*, 2003, *49* (11), 1580–1596.

_ , _ , and _ , "From Niches to Riches: Anatomy of the Long Tail," *MITSloan Management Review*, 2006, *47* (4), 67–71.

_ , _ , and _ , "Research Commentary — Long Tails vs. Superstars: The Effect of Information Technology on Product Variety and Sales Concentration Patterns," *Information Systems Research*, 2010, *21* (4), 736–747.

Cavallo, Alberto, "More Amazon Effects: Online Competition and Pricing Behaviors," Working Paper 25138, National Bureau of Economic Research 2018.

Goldberg, Pinelopi K., Amit K. Khandelwal, Nina Pavcnik, and Petia Topalova, "Multiproduct Firms and Product Turnover in the Developing World: Evidence from India," *The Review of Economics and Statistics*, 2010, *92* (4), 1042–1049.

He, Ruining and Julian McAuley, "Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering," *WWW*, 2016.

Iacovone, Leonardo and Beata S. Javorcik, "Multi-Product Exporters: Diversification and Micro-level Dynamics," Working Paper 4723, World Bank 2008.

Mayer, Thierry, Marc J. Melitz, and Gianmarco I. P. Ottaviano, "Market Size, Competition, and the Product Mix of Exporters," *American Economic Review*, 2014, *104* (2), 495–536.

McAuley, Julian, Christopher Targett, Qinfeng ('Javen') Shi, and Anton van den Hengel, "Image-Based Recommendations on Styles and Substitutes," *SIGIR*, 2015.

Melitz, Marc J. and Gianmarco I. P. Ottaviano, "Market Size, Trade, and Productivity," *The Review of Economic Studies*, 2008, *75* (1), 295–316.

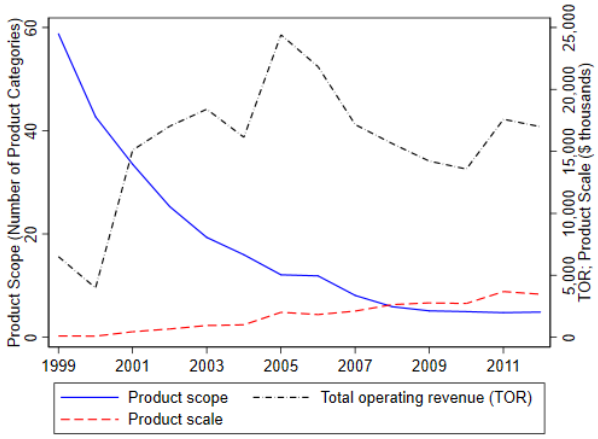
Figures and Tables

Table 1: Summary Statistics for Variables Used in Scale-Scope Decomposition in Figure 1a

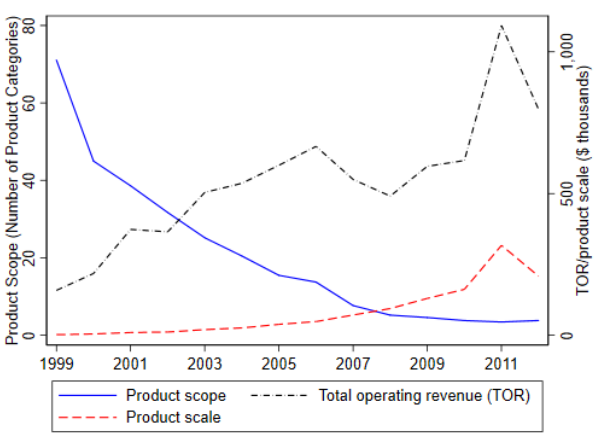
Variable	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
$\bar{pq}_t$	6,513	3,986	15,106	17,035	18,409	16,145	24,415	21,842	17,172	15,633	14,219	13,580	17,596	16,992
	(2,209)	(1,091)	(8,191)	(8,925)	(6,561)	(5,503)	(8,781)	(8,552)	(6,459)	(5,461)	(4,419)	(3,791)	(4,214)	(3,668)
$\bar{Scope}_t$	59	43	34	25	19	16	12	12	8	6	5	5	5	5
	(3.63)	(2.24)	(2.74)	(2.61)	(2.13)	(1.86)	(1.77)	(1.60)	(1.10)	(0.79)	(0.66)	(0.58)	(0.54)	(0.54)
Observations	144	252	186	175	190	219	167	180	208	218	223	243	219	248

Notes: Mean values for online firms in the Annual Non-Store Retail Survey (RTNS) are reported with standard errors beneath in parentheses. Annual average revenue per firm, $\bar{pq}_t$, is in thousands of Canadian dollars. Product scope, $\bar{Scope}_t$, is the mean count of product categories with positive revenue per firm, averaged across all firms in each year and rounded to the nearest integer. Online firms are defined as the subset of RTNS firms that are classified under NAICS category 4541 (electronic shopping and mail-order houses), and have positive online sales.

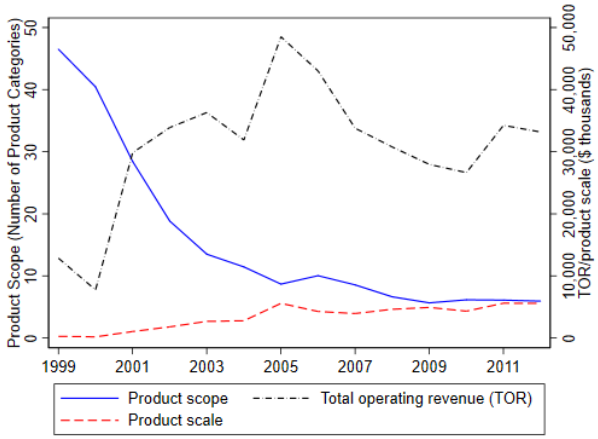
Figure 1: Decomposition of Online Retailers' Scale and Scope



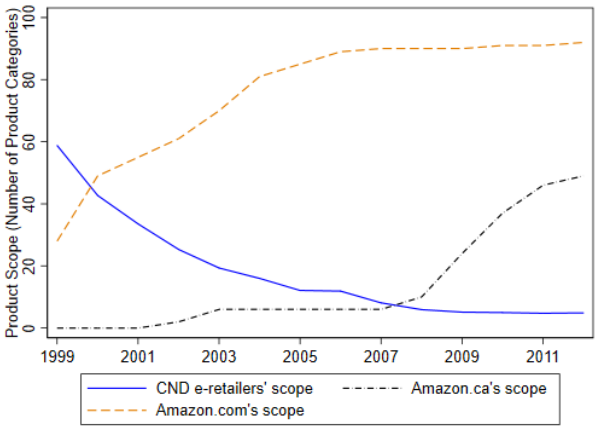
a. Scale and scope, full sample



b. Scale and scope, small firms: $TOR <= Median\ TOR$



c. Scale and scope, large firms: $TOR > Median\ TOR$



d. Product scope: Canadian online retailers and Amazon

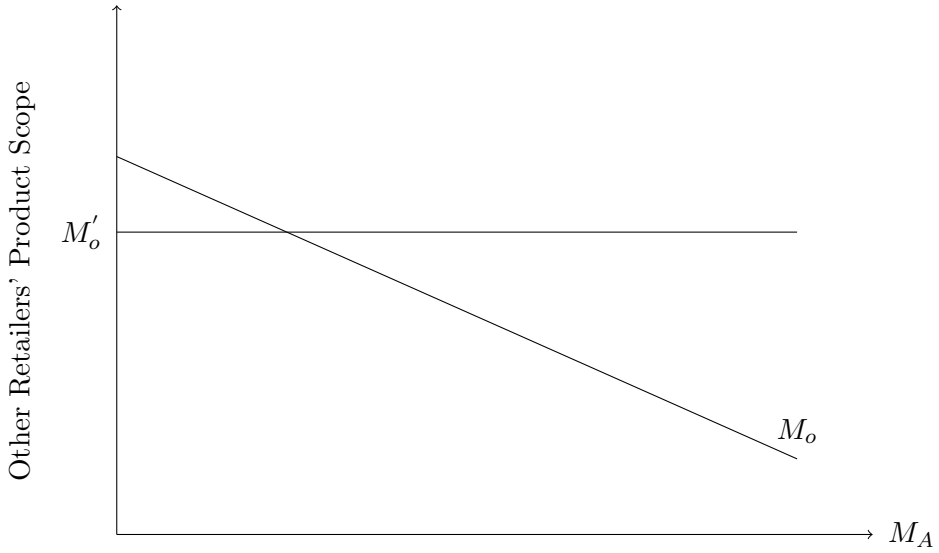
Notes: In figures 1a–1c, product scope is defined as the mean count of product categories with positive revenue per firm, averaged across all firms in each year. Product scale is the average annual operating revenue per firm divided by product scope, $\bar{pq}_t/\bar{Scope}_t$. In Figure 1d, the solid (blue) line is the product scope of RTNS firms and is identical to figure 1a. For Amazon.com, product scope is the number of RTNS product categories where Amazon.com has at least one reviewed product. For Amazon.ca, product scope is the number of RTNS categories where Amazon.ca had products for sale, as determined using Amazon.ca press releases and historic screen shots of Amazon.ca from *Internet Archive Wayback Machine*.

Table 2: Amazon Effects on Canadian Online Retailers' Extensive Margin of Product Participation

	(1) $share_{j,t}$	(2) $share_{j,t}$	(3) $share_{j,t}$	(4) $share_{j,t}$	(5) $share_{j,t}$	(6) $share_{j,t}$
$ln(Variety_{A,j,t})$	-0.0232	-0.0196				
	(0.00456)	(0.00396)				
$ln(Variety_{A,j,t}) \times ln(Variety_{A,t})$		-0.00555				
		(0.00117)				
$d_{com,j,t}$			-0.0487	-0.0980		
			(0.0164)	(0.0208)		
$d_{com,j,t} \times ln(Variety_{A,t})$				-0.0918		
				(0.0170)		
$d_{ca,j,t}$					-0.0498	-0.0261
					(0.0110)	(0.00881)
$d_{ca,j,t} \times ln(Variety_{A,t})$						-0.0244
						(0.00507)
Observations	1,394	1,394	1,394	1,394	1,394	1,394
R-squared	0.885	0.890	0.865	0.888	0.865	0.865
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Product Dummies	Yes	Yes	Yes	Yes	Yes	Yes

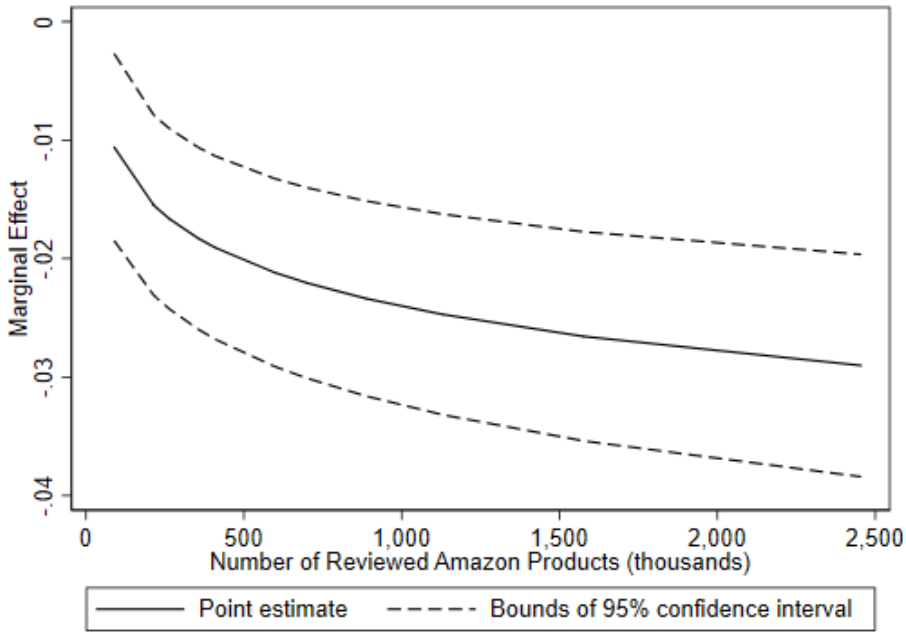
Notes: $share_{j,t}$ is the share of Canadian online firms with positive revenues in RTNS product category j in year t. $ln(Variety_{A,j,t})$ is the natural logarithm of 1 plus the count of unique Amazon.com products that are reviewed in year t belonging to RTNS product category j. $ln(Variety_{A,t})$ is the natural logarithm of the count of Amazon.com's reviewed products across all RTNS categories in year t and is normalized to have mean 0. $d_{com,j,t}$ is an indicator variable that is 1 if Amazon.com has entered product category j in year t. $d_{ca,j,t}$ is an indicator variable that is 1 if Amazon.ca has entered product category j in year t, and 0 otherwise. Standard errors (in parentheses) are clustered at the product level.

Figure 2: Online Retailers' Equilibrium Product Scope as a Function of Amazon.com's Scope



Notes: M_o is the short-run equilibrium product scope of online retailers in the model with Amazon.com, and $M_o^{'}$ is the short-run equilibrium product scope of online retailers without Amazon.com. M_A is Amazon.com's product scope.

Figure 3: Magnitude of Marginal Effect Increases with Amazon.com's Overall Product Expansion



Notes: The marginal effect is the change in the percentage of Canadian online firms in a product category in response to a 1% increase in Amazon.com's product variety in the category. This effect corresponds to $\beta_1 + \beta_2 ln(Varieties_{A,t})$, evaluated using the estimates from the baseline specification in column 2 of Table 2.