

Appendix to “The Ongoing Evolution of US Retail: A Format Tug-of-War” by Ali Hortaçsu and Chad Syverson

This appendix contains additional information on two analyses presented in the paper: e-commerce diffusion curves and geographic patterns of warehouse club growth.

A. Diffusion Curves

The file “diffusion_curves_figure.xlsx” includes a figure that displays the predicted logistic diffusion curves (i.e., S-curves) by retail product segment (“Diffusion Curves” tab) as well as values for the predicted curves as well as the raw data upon which they are based (“Data and Predictions” tab).

The raw data on the fractions of the product segments’ retail sales that are e-commerce is constructed using the following Census data sets, also in the spreadsheet:

1. U.S. Electronic Shopping and Mail-Order Houses (NAICS 45411) - Total and E-commerce Sales by Merchandise Line: 2013-1999 (“ESMOH” tab)
2. Estimated Annual Sales of U.S. Retail and Food Services Firms by Kind of Business: 1992 Through 2013 (“Annual Sales” tab)
3. U.S. Retail Trade Sales - Total and E-commerce: 2013-1998 (“E-commerce” tab)

The template for estimating the predicted curves is based on a worksheet, “S-Curve Forecasting,” designed by Stephen R. Lawrence (template url: [leeds-faculty.colorado.edu/lawrence/Tools/SCurve/scurve.xls](https://faculty.colorado.edu/lawrence/Tools/SCurve/scurve.xls)). Each product segment’s fraction of e-commerce sales is assumed to follow a logistic diffusion curve governed by the following equation: $Y_t = \frac{L}{1 + Ae^{-Bt}}$, where Y_t is the fraction of the segment’s total retail sales classified as e-commerce at time t , L is the saturation parameter (the asymptotic e-commerce fraction, here set $L = 1$ for all product segments), and A and B are estimated parameters.

By transforming the observed shares using the function $y_t = \ln\left(\frac{L}{Y_t} - 1\right)$, we can estimate the linearized form of the diffusion curve $y_t = \alpha - Bt + \varepsilon_t$, where $\alpha = \ln A$, and ε_t is an error term. We then retransform the fitted values of this estimated linearized curve using the inverse of the function above to obtain the predicted diffusion curves. We do not report predicted values for

product-years within the sample for which we do not have data, though these can be easily constructed using the estimates.

The file “diffusion_curves_table.pdf” presents an overview of the diffusion curve estimates, including the number of observations available to estimate each product segment curve, the estimated curvature parameter B , and the R^2 of the estimated (linearized) equation by retail segment.

B. Warehouse Clubs

The file “cbp2003_2013.csv” contains the 2003 and 2013 Census County Business Patterns data for the industries used in our analysis. The “warehouse_clubs.do” file is a Stata program that runs all tests exploring the impact of warehouse club growth presented in the paper. Specifically, it imports the data from cbp2003_2013 (the specific industries in the data are described in the comments of this file), cleans the data, and runs the relevant regressions. Users can write code to output the results in their preferred format.

The file “geo_tests.pdf” presents the results from the warehouse club growth regressions in *.pdf format.