# Online Appendix to “The Impact of E-Commerce on Relative Prices and Consumer Welfare?” (Not for Publication)

Yoon J. Jo
Texas A&M University

Misaki Matsumura

David E. Weinstein
*Columbia University and NBER*

November, 2019

## Table of Contents



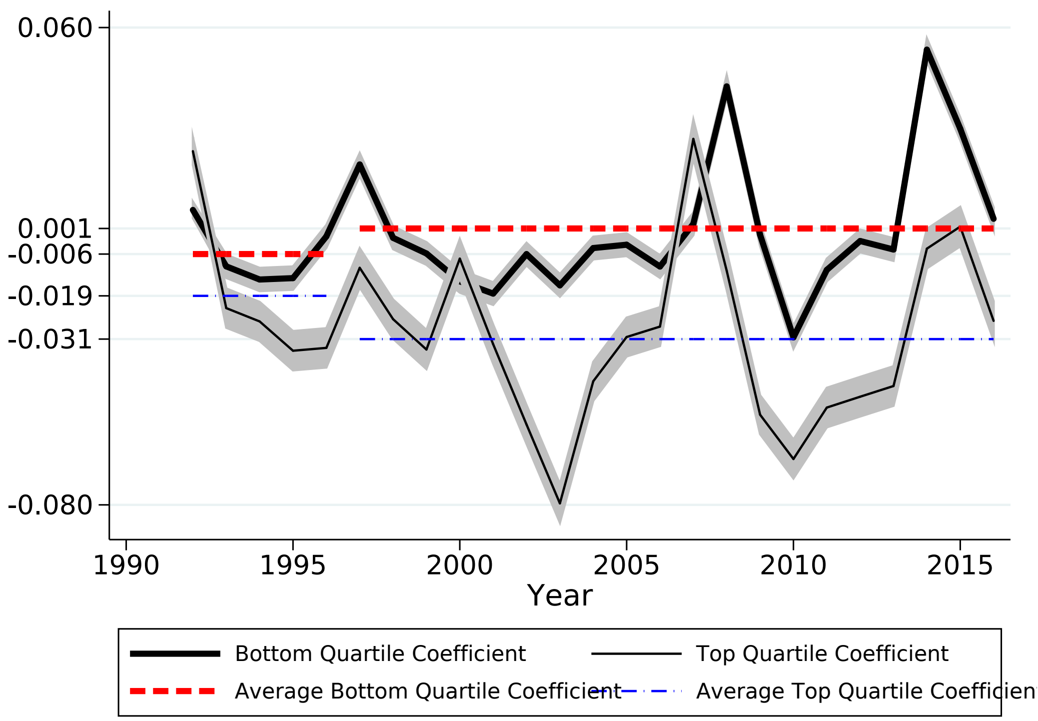
This online appendix contains supplementary empirical results. Section A.1 replicates our results on national rates of price change using different subsamples of goods and alternative estimation methods. Section A.2 presents OLS versions of our specifications, and Section A.3 replicates our results using Rakuten data instead of NSFIE data to show our results are robust to using alternative measures of e-commerce sales intensity.

## A.1 Robustness of National Price Change Differentials

One of the most volatile sectors in the Japanese CPI is fruits and vegetables. Since these products tend to have low e-commerce intensity, movements in these prices sometimes cause large swings in the prices of low e-commerce intensity goods in some years. We explore the sensitivity of our results to dropping these sectors. In Figure A.1, we replicate Figure 2 but drop fruits and vegetables from the sample. As one can see from Figure A.1, we still see the same pattern in the data in which goods with low e-commerce intensity have higher rates of price change than goods with high intensity. We also see this difference increase in the e-commerce period. Interestingly, in 1992, the first observation in the plot, we see that goods with high and low e-commerce intensity exhibit similar rates of price increase, but in Figure A.1 (which excludes fruits and vegetables), we see a negative rate of price increase in 1992. This demonstrates that the low rate of price change in that year is driven by the volatile fruits and vegetable sector.

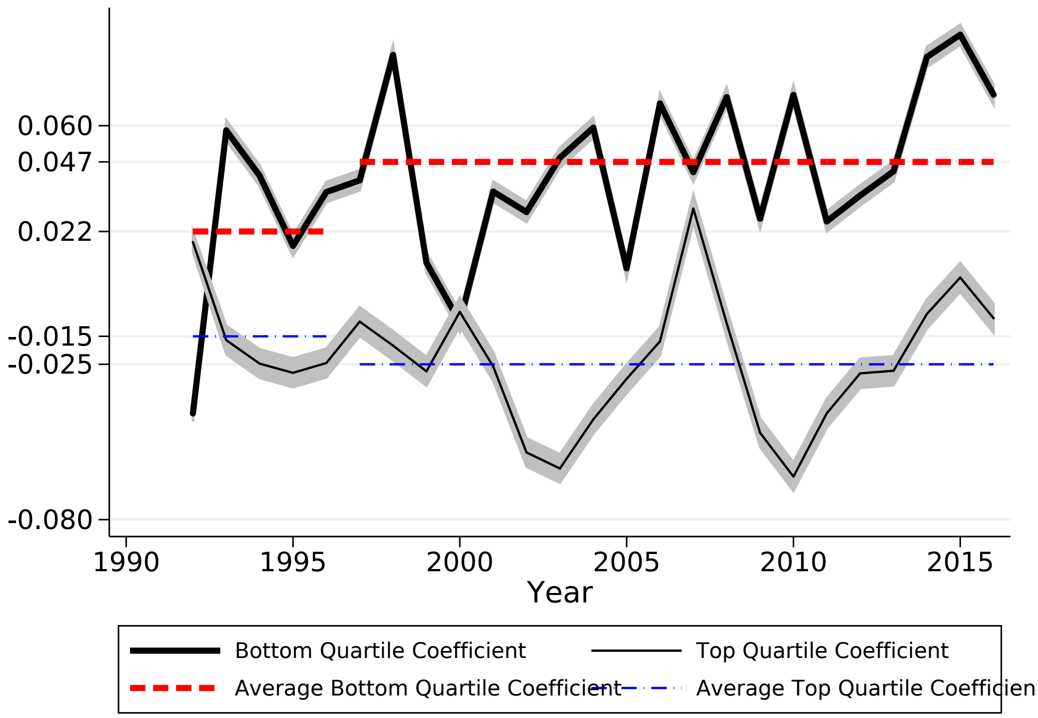
Figure 2 in the paper is based on a weighted regression of the price changes in which the weights equal the expenditures. In Figure A.2, we replicate the figure using unweighted price changes as a robustness check. As one can see, using an unweighted regression does not alter the qualitative result showing that goods with low e-commerce intensity have higher rates of price change than goods with high intensity and that this gap widened in the e-commerce period.

Figure A.1: Price Growth of Goods With High and Low E-Commerce Intensity excluding Fruits and Vegetables




Note: The black lines show the evolution of the time fixed effect $\hat{\theta}_t$ from equation (12), which corresponds to the average rate of expenditure-weighted price increase for the goods excluding the sector "Fruits and vegetables" in two groups: products with bottom quartile e-commerce sales intensity (thick black line) and products with top quartile e-commerce intensity (thin black line). The red dashed line shows the average rate of price increase before and after the entry of e-commerce firms for goods with bottom quartile e-commerce sales intensity, and the blue dashed line shows the one for goods with top quartile e-commerce sales intensity.

Figure A.2: Price Growth of Goods With High and Low E-Commerce Intensity




Note: The black lines show the evolution of the time fixed effect $\hat{\theta}_t$ from equation (12), which corresponds to the average rate of unweighted price increase for the goods in two groups: products with bottom quartile e-commerce sales intensity (thick black line) and products with top quartile e-commerce intensity (thin black line). The red dashed line shows the average rate of price increase before and after the entry of e-commerce firms for goods with bottom quartile e-commerce sales intensity, and the blue dashed line shows the one for goods with top quartile e-commerce sales intensity.

## A.2 OLS Regression Results

In this section, we replicate our main results on regional price convergence in Tables 5 and B2 without instrumenting for e-commerce intensity and just using an OLS estimator. As one can see from Tables B1 and B2, the results are similar in that we observe the same basic pattern of convergence rates for e-commerce intensive goods being higher after the entry of e-commerce merchants. The coefficient on the e-commerce intensity terms is not significantly different from that of the IV specifications.

Table B1: Price Convergence in Pre and Early Post E-Commerce Period (OLS regression)

| | (1)<br>$\Delta\ p_{ict}$ | (2)<br>$\Delta\ p_{ict}$ | (3)<br>$\Delta\ p_{ict}$ | (4)<br>$\Delta\ p_{ict}$ |
|---|---|---|---|---|
| Lagged Price | -0.293 | -0.326 | -0.295 | -0.120 |
| | (0.026) | (0.027) | (0.026) | (0.012) |
| E-Commerce Intensity | -0.166 | | -0.162 | 0.162 |
| $\times$ Lagged Price | (0.267) | | (0.274) | (0.121) |
| $D_t$ $\times$ | | | -0.030 | -0.002 |
| Lagged Price | | | (0.023) | (0.009) |
| $D_t$ $\times$ | | -1.119 | -0.965 | -0.402 |
| E-Commerce Intensity $\times$ Lagged Price | | (0.321) | (0.392) | (0.117) |
| $t$ | {1996} | {2001} | {1996,2001} | Annual<br>1991-2001 |
| Observations | 26,265 | 27,693 | 51,880 | 272,594 |
| $R^2$ | 0.52 | 0.52 | 0.52 | 0.47 |

Note: E-Commerce intensity ($x_{i09}^E$) is defined in equation (10); $D_t$ is dummy variable that is one after 1996 (i.e., the post e-commerce period). Table shows the regression results of equation (1) using OLS. The first column uses the five-year log differences in prices from 1991 - 1996, and the second column uses that from 1996 - 2001. The third column uses two five-year periods, 1991 - 1996 and 1996 - 2001. The fourth column uses the annual log differences in prices from 1992 to 2001. The IV regression results are available in Table 5. The F-statistic corresponds to the Kleibergen-Paap rk Wald statistic. Standard errors in parentheses.

Table B2: Price Convergence Over Alternative Periods (OLS regression)

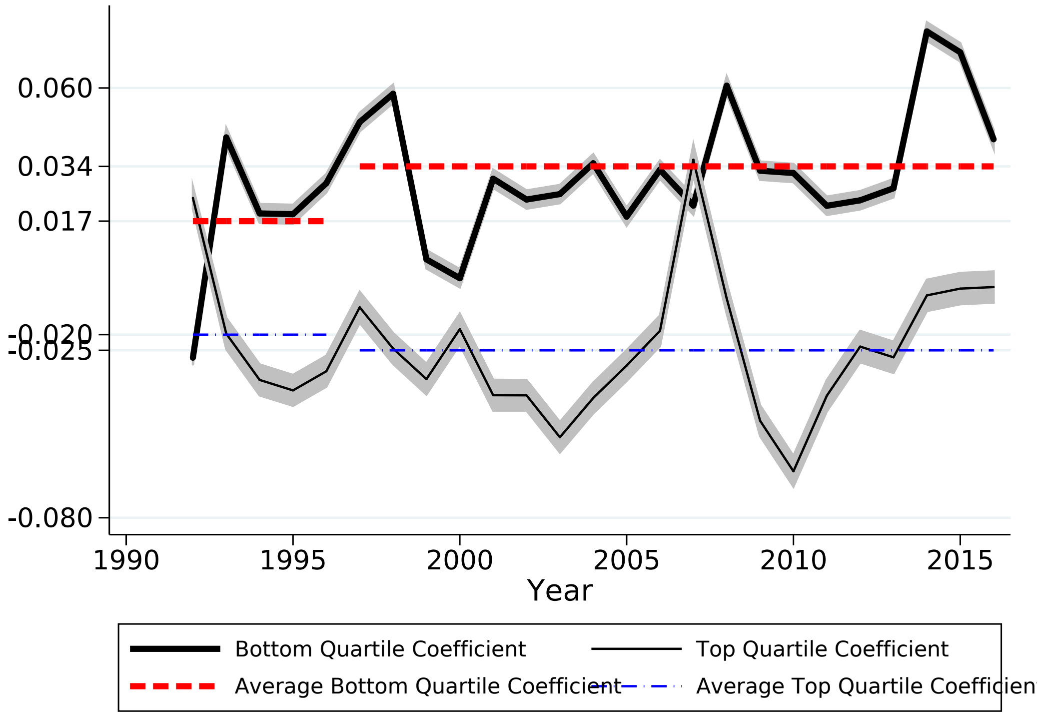
| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ |
| Lagged Price | -0.114 | -0.117 | -0.116 | -0.117 | -0.137 | -0.115 |
| | (0.010) | (0.012) | (0.013) | (0.010) | (0.012) | (0.010) |
| E-Commerce Intensity | 0.064 | 0.091 | 0.049 | | | |
| $\times$ Lagged Price | (0.105) | (0.115) | (0.087) | | | |
| $D_t$ $\times$ | -0.047 | -0.040 | -0.042 | -0.049 | -0.034 | -0.049 |
| Lagged Price | (0.009) | (0.010) | (0.010) | (0.008) | (0.011) | (0.008) |
| $D_t$ $\times$ | -0.206 | -0.237 | -0.143 | | | |
| E-Commerce Intensity $\times$ Lagged Price | (0.116) | (0.100) | (0.069) | | | |
| High E-Commerce | | | | 0.011 | 0.047 | 0.005 |
| Intensity $\times$ Lagged Price | | | | (0.006) | (0.016) | (0.004) |
| $D_t$ $\times$ | | | | -0.009 | -0.039 | -0.011 |
| High E-Commerce Intensity $\times$ Lagged Price | | | | (0.005) | (0.013) | (0.006) |
| $t$ | Annual 1991-2016 | | | | | |
| Observations | 579,157 | 581,803 | 578,704 | 579,157 | 581,803 | 578,704 |
| E-Commerce Intensity Year | 2004 | 2009 | 2014 | 2004 | 2009 | 2014 |
| $R^2$ | 0.42 | 0.42 | 0.43 | 0.42 | 0.42 | 0.43 |

Note: E-Commerce intensity ($x_{ib}^E$) is defined in equation (10); $D_t$ is dummy variable that is one after 1996 (i.e., the post e-commerce period). Table shows the regression results of equation (1) using OLS. Dependent variables are the annual log price differences from 1991 to 2016. The first three columns of the table show the regression results of equation (1) using 1999 catalog sales intensity as an instrument for e-commerce sales intensity in 2004 (column 1), 2009 (column 2), and 2014 (column 3). The last three columns of table present regression results with discrete e-commerce intensity variable that indicates whether a good has an e-commerce intensity above the median. The IV regression results are available in Table 6. Standard errors in parentheses.

## A.3 Results Using Rakuten Sales Data as a Measure of E-Commerce Intensity

In this section of the appendix, we demonstrate that all of our main results are robust to measuring e-commerce sales intensity ($x_{ib}^R$) based on Rakuten sales data. This establishes that our results are not dependent on any idiosyncrasies in how the Japanese government collects data on online expenditures.

Figure C1: Price Growth of Goods With High and Low E-Commerce Intensity




Note: This black line shows the average rate of expenditure-weighted price increase for the goods in two groups: products with bottom quartile Rakuten sales intensity (thick black line) and products with top quartile Rakuten intensity (thin black line). The red dashed line shows the average rate of price increase before and after the entry of e-commerce firms for goods with bottom quartile e-commerce sales intensity, and the blue dashed line shows that for goods with top quartile e-commerce sales intensity.

### A.3.1 National Price Results

In Figure A.2, we replicate the Figure 2 using unweighted price changes. As one can see, using Rakuten data instead of NSFIE data to measure e-commerce sales intensity does not alter the qualitative result showing that goods with low e-commerce intensity have higher rates of price change than goods with high intensity and that this gap widened in the e-commerce period.

Tables C1 and C2 replicate Tables 3 and C2 using Rakuten data to measure e-commerce intensity. The results are qualitatively similar to those pre-

Table C1: Relative Price Changes and Rakuten Intensity

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ |
| $D_t$ | -0.0010 | 0.0065 | -0.0044 | 0.0115 |
| | (0.0021) | (0.0021) | (0.0044) | (0.0038) |
| E-Commerce Intensity | -0.0080 | -0.0465 | 0.0645 | -0.1475 |
| $\times\ D_t$ | (0.0144) | (0.0141) | (0.0855) | (0.0615) |
| Fixed Effects | Product | Product | Product | Product |
| Estimation Period | 1992-2001 | 1992-2016 | 1992-2001 | 1992-2016 |
| Observations | 273,530 | 583,830 | 273,530 | 583,830 |
| $R^2$ | 0.03 | 0.03 | | |
| First-Stage F-Stat | | | 9.89 | 10.97 |
| Estimation Method | OLS | OLS | IV | IV |

Note: The dependent variable is the one-year log price change; $D_t$ is dummy variable that is one after 1996 (i.e., the post e-commerce period); and Rakuten intensity ($x^R_{i09}$) is defined in equation (10). The table shows relative price changes for goods sold online intensively relative to goods not sold online intensively before and after the entry of e-commerce firms. Column 1 and 3 are for 1992-2001, and column 2 and 4 are for 1992-2016. The first two columns show OLS estimates using e-commerce sales intensity, and the second two columns use catalog sales intensity as an instrument. The F-statistic corresponds to the Kleibergen-Paap rk Wald statistic. Standard errors in parentheses.

sented in the text.

Table C2: First Stage of Instrumental Variables Regression

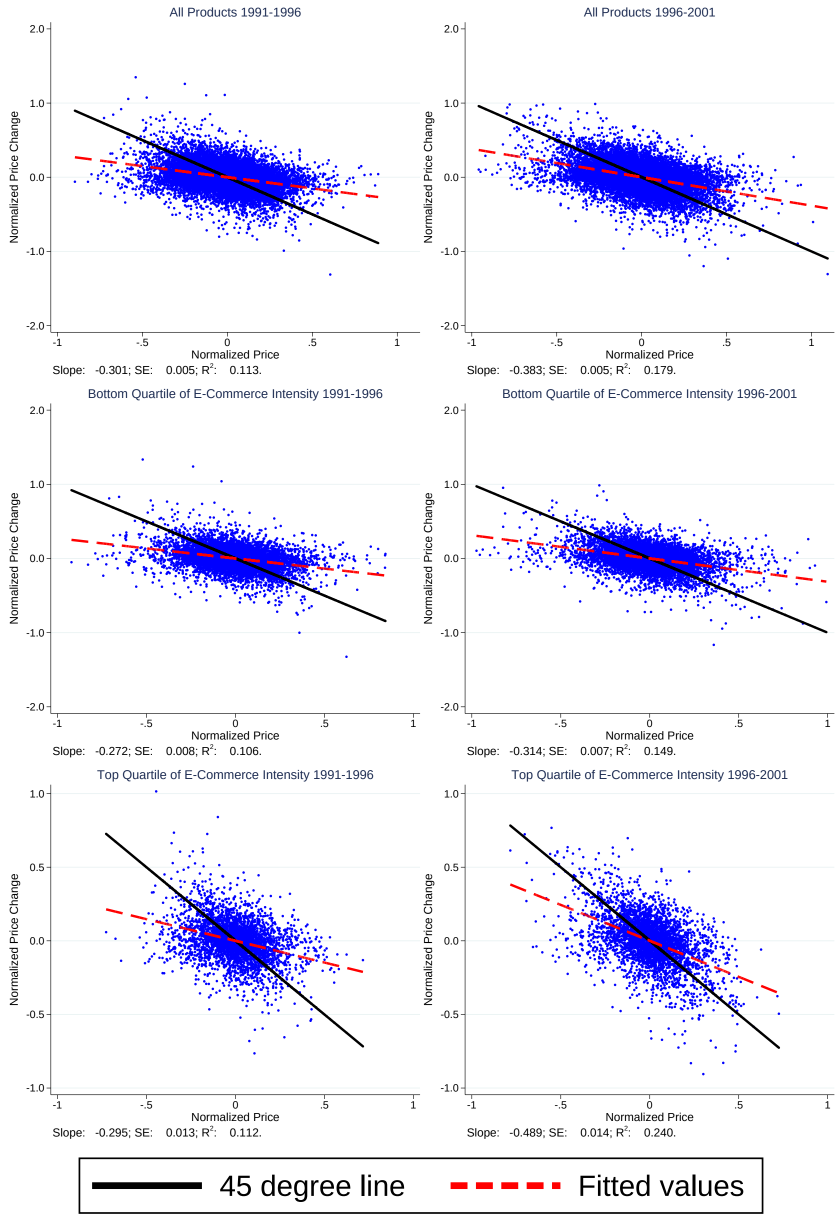
| | (1) E-Commerce Intensity $\times D_t$ | (2) E-Commerce Intensity $\times D_t$ | (3) E-Commerce Intensity |
|---|---|---|---|
| Catalog Intensity | 0.2113 | 0.2115 | |
| $\times D_t$ | (0.0672) | (0.0639) | |
| $D_t$ | 0.0321 | 0.0340 | |
| | (0.0056) | (0.0056) | |
| Catalog Intensity | | | 0.1970 |
| | | | (0.0460) |
| Constant | | | 0.0455 |
| | | | (0.0066) |
| Fixed Effects | Product | Product | None |
| Estimation Period | 1992-2001 | 1992-2016 | |
| Observations | 273,530 | 583,830 | 388 |
| $R^2$ | 0.07 | 0.07 | 0.05 |
| Estimation Method | IV-First Stage | IV-First Stage | OLS |

Note: Rakuten intensity ($x^R_{i09}$) and catalog intensity ($x^C_{i99}$) are defined in equation (10); $D_t$ is dummy variable that is one after 1996 (i.e., the post e-commerce period). The first two columns of the table present the first-stage regression results (i.e., separate observations for each city and year): column 1 shows results using data for 1992-2001, and column 2 uses data for 1992-2016. The last column presents an OLS regression using only the goods data for 2009. Standard errors in parentheses.

### A.3.2 Estimation of Price Convergence Using Rakuten Data

Figure C2 replicates Figure 3 using Rakuten data to measure e-commerce intensity instead of NSFIE data. Similarly, Tables C3 and C4 replicate Tables 5 and 6 using Rakuten data. The results are qualitatively the same as those presented in the text.

Figure C2: Normalized Price Change vs. Normalized Price




Note: This graph plots normalized price changes against normalized price levels. Normalized log price changes (log levels) equal the actual price changes (log levels) less the fixed effects from a regression of log price changes (log levels) on product and city fixed effects. The left panel shows normalized price changes before the entry of e-commerce, and the right panel shows them after the entry of e-commerce. The first panel plots for all goods, the second panel plots for goods with Rakuten intensity lower than the bottom quartile, and the third panel shows for goods with Rakuten intensity higher than the top quartile.

Table C3: Price Convergence in Pre and Early Post E-Commerce Period

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ |
| Lagged Price | -0.290 | -0.308 | -0.288 | -0.136 |
| | (0.035) | (0.042) | (0.034) | (0.017) |
| E-Commerce Intensity | -0.229 | | -0.301 | 0.528 |
| $\times$ Lagged Price | (0.513) | | (0.489) | (0.277) |
| $D_t \times$ | | | -0.019 | 0.017 |
| Lagged Price | | | (0.034) | (0.015) |
| $D_t \times$ | | -1.570 | -1.294 | -0.829 |
| E-Commerce Intensity $\times$ Lagged Price | | (0.781) | (0.653) | (0.285) |
| $t$ | {1996} | {2001} | {1996,2001} | Annual 1991-2001 |
| Observations | 26,342 | 27,796 | 52,565 | 273,530 |
| First-stage F | 9.91 | 11.64 | 5.85 | 5.54 |
| Estimation | IV | IV | IV | IV |

Note: Rakuten intensity ($x^R_{i09}$) is defined in equation (10); $D_t$ is dummy variable that is one after 1996 (i.e., the post e-commerce period). The table shows regression results of equation (1) using IV: Rakuten intensity in 2009 is instrumented using 1999 catalog sales intensity. The first column uses the five-year log differences in prices from 1991 - 1996, and the second column uses that from 1996 - 2001. The third column uses two five-year periods, 1991 - 1996 and 1996 - 2001. The fourth column uses the annual log differences in prices from 1992 to 2001. The F-statistic corresponds to the Kleibergen-Paap rk Wald statistic. Standard errors in parentheses.

Table C4: Price Convergence Over Alternative Periods

| | (1) | (2) |
|---|---|---|
| | $\Delta\ p_{ict}$ | $\Delta\ p_{ict}$ |
| Lagged Price | -0.133 | -0.125 |
| | (0.017) | (0.013) |
| E-Commerce Intensity | 0.431 | |
| $\times$ Lagged Price | (0.261) | |
| $D_t \times$ | -0.012 | -0.029 |
| Lagged Price | (0.016) | (0.015) |
| $D_t \times$ | -0.809 | |
| E-Commerce Intensity $\times$ Lagged Price | (0.238) | |
| High E-Commerce | | 0.027 |
| Intensity $\times$ Lagged Price | | (0.033) |
| $D_t \times$ | | -0.050 |
| High E-Commerce Intensity $\times$ Lagged Price | | (0.028) |
| $t$ | Annual | Annual |
| Observations | 583,830 | 583,830 |
| E-Commerce Intensity Year | 2010 | 2010 |
| First-stage F | 5.94 | 21.79 |
| Estimation | IV | IV |

Note: Rakuten and catalog intensity are defined in equation (10); $D_t$ is dummy variable that is one after 1996 (i.e., the post e-commerce period). The table shows the regression results of equation (1) using 1999 catalog sales intensity as an instrument for e-commerce sales intensityThe second column of table presents regression results with discrete e-commerce intensity variable that indicates whether a good has an e-commerce intensity above the median. The F-statistic corresponds to the Kleibergen-Paap rk Wald statistic. The fifth column uses the annual frequency of log price changes. Dependent variables are the annual log price differences from 1991 to 2016. Standard errors in parentheses.