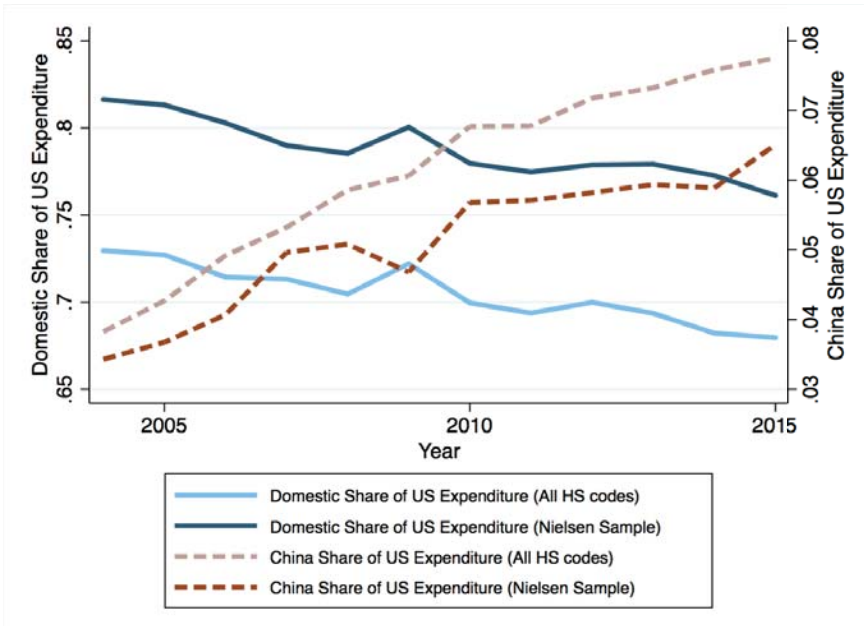
# Estimating US Consumer Gains from Chinese Imports: Online Appendix

*Liang Bai, Sebastian Stumpner*

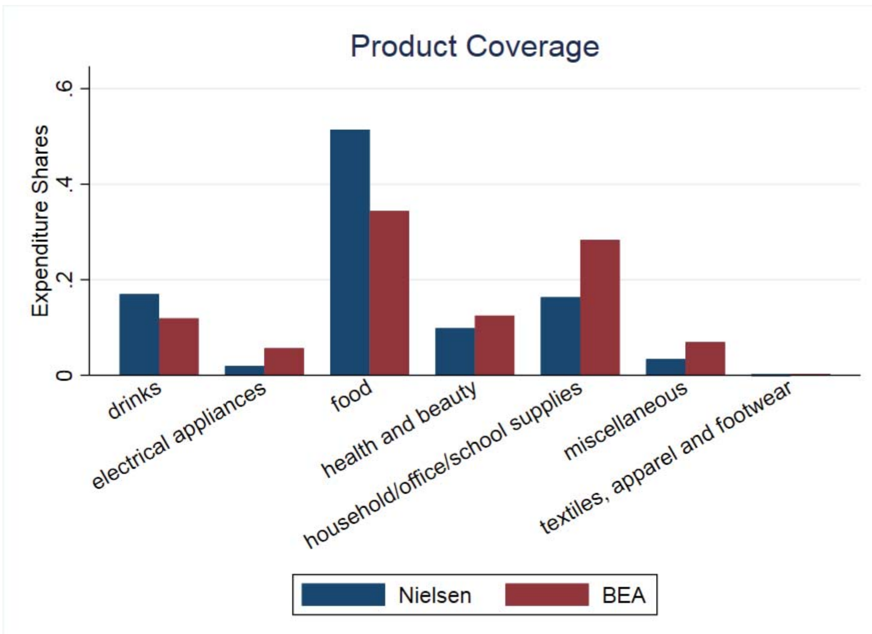
# A Additional Figures and Tables

Figure A.1: Changing Composition of US Expenditure



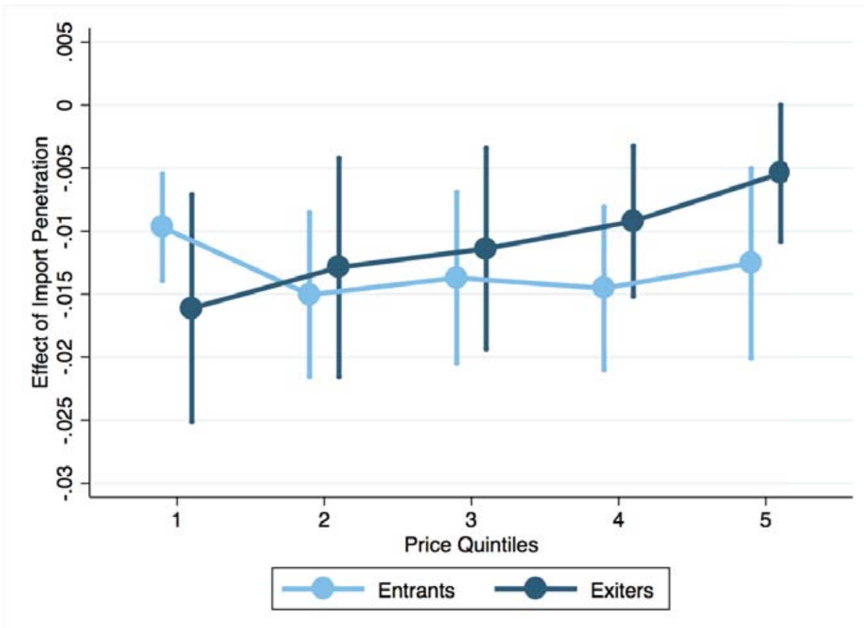

Notes: This figure shows the evolution of both the domestic and China shares of US expenditure (left and right axes respectively) during 2004-2015, comparing all HS 6-digit codes with those in the Nielsen sample. Total US expenditure is computed as production + imports - exports.

Figure A.2: Nielsen Product Coverage




Notes: This figure shows the expenditure shares by broad product groups, averaged during 2004-2015, for both projected Nielsen and imputed BEA expenditures. Imputed BEA expenditure is computed as production + imports - exports. Note that this captures only the part of imputed BEA expenditure that map into the Nielsen data, and not necessarily all BEA-implied expenditure on a particular broad product group.

Figure A.3: Effect on Entry and Exit Rates by Price Quintile




Notes: This figure shows coefficient estimates (and 95% confidence intervals) of quintile-specific entry and exit rates on $\Delta$ DSE, instrumented by Chinese import penetration in Europe. For each category and year, we first compute quintile-specific entry and exit rates. The entry rate is computed as $EntryRate^q_{c,t} = \frac{M^q_{ct}}{N_{ct}}$, where $M^q_{ct}$ denotes the number of entrants in category $c$ at year $t$ in price quintile $q$, and $N_{ct}$ is the number of all varieties observed in $c$ and $t$. We then average entry rates over time. Exit rates are computed analogously. All estimations include our main set of controls: US productivity growth and quartile dummies for average income, age and income growth among the households with purchases in each category. All regressions are weighted by 2004 projected Nielsen expenditure.

Table A.1: Nielsen Expenditures - by broad product groups

| Product Group | (1) Number of Categories | (2) Expenditure ($ bn) 2004 | (3) Expenditure ($ bn) 2015 | (4) Number of Barcodes 2004 | (5) Number of Barcodes 2015 |
|---|---|---|---|---|---|
| Drinks | 27 | 58.7 | 69.9 | 65187 | 81011 |
| Electrical Appliances | 19 | 7.6 | 7.3 | 12225 | 14989 |
| Food | 93 | 168 | 226 | 281771 | 320785 |
| Health and Beauty | 31 | 33.2 | 42.9 | 86102 | 98222 |
| Household/Office/School Supplies | 52 | 58.3 | 65.1 | 144734 | 175500 |
| Miscellaneous | 8 | 13.0 | 11.3 | 8945 | 6803 |
| Textiles, Apparel and Footwear | 2 | 0.2 | 0.2 | 986 | 918 |
| Total | 232 | 339 | 423 | 599950 | 698228 |

Notes: Based on projected Nielsen expenditures for our analytical sample of 232 product categories.

Table A.2: Nielsen - HS Concordance Merge Types

| Merge Type | (1) Number of Categories | (2) Number of Nielsen Product Modules | (3) Number of HS 6-digit Codes |
|---|---|---|---|
| 1:1 | 125 | 125 | 125 |
| 1:n | 51 | 51 | 284 |
| m:1 | 87 | 564 | 87 |
| m:n | 61 | 407 | 382 |
| Total | 324 | 1147 | 878 |

Notes: A 1:1 merge refers to a case where a single Nielsen product module is matched to a single HS 6-digit code. A 1:n merge refers to a case where a single Nielsen product module is matched to multiple HS 6-digit codes. A m:1 merge refers to a case where multiple Nielsen product modules are matched to a single HS 6-digit code. A m:n merge refers to a case where multiple Nielsen product modules are matched to multiple HS 6-digit codes.

Table A.3: Summary Statistics

| Variable | Mean | Std. Dev. | P10 | P50 | P90 |
|---|---|---|---|---|---|
| *A. Inflation (category level)* | | | | | |
| Inflation ($\sigma$ = 3) | 0.91 | 0.37 | 0.40 | 0.94 | 1.38 |
| Inflation ($\sigma$ = 5) | 1.02 | 0.32 | 0.60 | 1.04 | 1.42 |
| Inflation ($\sigma$ = Broda Weinstein Elast.) | 0.96 | 0.40 | 0.41 | 1.02 | 1.43 |
| *B. Inflation (various margins)* | | | | | |
| Intensive Margin (short stayers) | 1.17 | 0.25 | 0.86 | 1.17 | 1.48 |
| Intensive Margin (long stayers) | 1.26 | 0.22 | 1.06 | 1.24 | 1.49 |
| Extensive Margin ($\sigma$ = 5) | 0.86 | 0.13 | 0.67 | 0.89 | 0.98 |
| *C. Additional Outcomes* | | | | | |
| Log Change of Expenditure | 0.04 | 0.48 | -0.48 | 0.08 | 0.48 |
| Log Change in the No. of Barcodes | 0.02 | 0.37 | -0.34 | 0.04 | 0.45 |
| Average Entry Rate | 0.24 | 0.09 | 0.13 | 0.23 | 0.36 |
| Average Exit Rate | 0.24 | 0.09 | 0.13 | 0.22 | 0.35 |
| *D. China Trade Shock* | | | | | |
| $\Delta$ DSE | -0.18 | 0.39 | -0.57 | -0.05 | 0.02 |
| China IP (Europe) | 0.05 | 0.11 | 0.00 | 0.02 | 0.13 |

Notes: $N = 232$. All variables are computed for the time period 2004-15. P10, P50, and P90 refer to the 10th, 50th, and 90th percentile values, respectively. The short stayer intensive margin inflation measure includes all goods that are observed in any two consecutive years, while the long stay measure only includes goods that are observed in every year. For the average entry (exit) rate, we first compute the category-by-year entry (exit) rate as the number of entering (exiting) barcodes, divided by all barcodes in that category and year, and then average these rates over time. $\Delta$ DSE is the log change in the domestic share of expenditure for the US during 2004-2015. China IP is China import penetration in the five largest European economies (Germany, France, UK, Italy and Spain).

Table A.4: Barcode-level Regression Results

| Dependent Variable: | Price | Expenditure | Exit | Price | Expenditure |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $\Delta$ DSE | 0.649*** | 3.898** | -6.409*** | 0.140 | 2.198** |
| | (0.171) | (1.574) | (1.505) | (0.114) | (1.024) |
| Observations | 472,335 | 472,335 | 472,335 | 142,993 | 142,993 |
| Sample | All | All | All | Stayer | Stayer |

Notes: Category-level clustered standard errors in parentheses. *** $p<0.01$, ** $p<0.05$,* $p<0.10$. Columns 1-3 use as sample all barcodes that were present in 2004, while Columns 4-5 use as sample barcodes that are present in every year. $\Delta$ DSE is the log change in the domestic share of expenditure for the US during 2004-2015. The instrument used is China import penetration in the five largest European economies (Germany, France, UK, Italy and Spain). Unless otherwise noted, all regressions include the following controls: US productivity growth and quartile dummies for average income, age and income growth among the households with purchases in each category. All estimations are weighted with initial expenditure for a given barcode.

Table A.5: Further Robustness

| | *A. Sub-Sample Analysis* | | | *B. Inflation w/ Alternative Elasticities* | | |
|---|---|---|---|---|---|---|
| | No Food & Drinks | No Electr. Appl. | 10th-90th | | | |
| Dependent Variable: | Inflation | Inflation | Inflation | $\sigma = 3$ | $\sigma = 10$ | $\sigma = BW$ |
| | (1) | (2) | (3) | (1) | (2) | (3) |
| Δ DSE | 0.212** | 0.208** | 0.437** | 0.491*** | 0.293*** | 0.925*** |
| | (0.093) | (0.090) | (0.214) | (0.156) | (0.091) | (0.193) |
| Observations | 112 | 213 | 167 | 232 | 232 | 232 |

| | *C. Alternative Weights* | | | | *D. Gravity IV* |
|---|---|---|---|---|---|
| | First Stage | | IV | | |
| Dependent Variable: | Δ DSE | Δ DSE | Inflation | Inflation | Inflation |
| | (1) | (2) | (3) | (4) | (1) |
| China IP (Europe) | -2.498*** | -2.050*** | | | |
| | (0.351) | (0.436) | | | |
| Δ DSE | | | 0.668*** | 0.745** | 0.387*** |
| | | | (0.159) | (0.294) | (0.146) |
| Weights | No | BEA | No | BEA | Nielsen |

Notes: $N = 232$ (unless otherwise indicated). Robust standard errors in parentheses. *** p<0.01, ** p<0.05,* p<0.10. Δ DSE is the log change in the domestic share of expenditure for the US during 2004-2015. The instrument used is China import penetration in the five largest European economies (Germany, France, UK, Italy and Spain). Panel A reports inflation results when we separately exclude food and drinks (Column 1), electrical appliances (Column 2), and categories outside of the 10th and 90th percentile values in inflation and $\Delta DSE$ (Column 3). Columns 1 and 2 of Panel B show results when the inflation rate is computed using an elasticity of substitution equal to 3 and 10 respectively, while Column 3 shows results for category-specific elasticities, estimated using the methodology by Broda and Weinstein (2006), and obtained from Soderbery (2015). Panel D uses a gravity-model based instrument, constructed as the difference between Chinese and US import growth in the EU, weighted by the initial China share in EU expenditure. Unless otherwise noted, all regressions include the following controls: US productivity growth and quartile dummies for average income, age and income growth among the households with purchases in each category. The weights used in Panels A, B. and D are total projected Nielsen expenditure in a given category in 2004. BEA weights in Panel C are total US expenditure in a given category in 2004, computed as production - exports + imports.

# B Data Appendix

## B.1 Description of AC Nielsen Homescan Data

The Homescan data come from a private vendor, AC Nielsen, and are made available through the Kilts Center at the University of Chicago. It is a panel of roughly 60,000 households that provide information about their product purchases (price and expenditure) at the level of individual barcodes.[20] For our study period, we observe over 1.5 million barcodes, grouped by Nielsen into 1,147 product modules. Some examples of product modules include olive oil, pasta-spaghetti, dental accessories, cameras, batteries, and printers.

Nielsen recruits households by mail or online, and provides incentives to join and to remain active in reporting transactions. Examples of incentives include monthly prize drawings and gift points. Households that do not regularly report their transactions are removed from the sample and new households are introduced. In all of our results, we use the household projection weights provided by Nielsen to make the sample demographics representative of national demographics.[21] The panelists were provided with in-home scanners to record all of their purchases at the universal product code (UPC) level. These are the prices effectively paid by households, and include discounts (e.g., getting the second item at 50% off). Prices of products are collected from one of two sources. If the store in which the product was bought also reports to Nielsen's store-level survey Scantrack, then the price reported from the store is taken directly. If not, then the household's reported price is used. Einav et al. (2010) test the accuracy of the price data by using a sample of transactions for which they observe both the retailer's price and the household's recorded price. They find that, even though mistakes in price entry do occur, the correlation between the two is reasonably high (88%).

Household income in the Nielsen data is reported in 16 brackets. To compute the share of expenditure in income, we impute income using the average value within each bracket. For the two extreme brackets ($>$ \$100,000 and $<$ \$5,000), we impute income as \$100,000 and \$5,000, respectively. Household age is computed as the average age between the male and female household head.

We use quartile dummies for average income, growth in average income and average age of consumers at the product-category level as control variables. We compute these variables from the Nielsen data, as expenditure-weighted average of the variable among all households with purchases in a partcular product category.

[20] From 2004-2006, sample size is limited to roughly 40,000 households.

[21] We also ran all regressions without using household projection weights to compute category-level inflation rates, and find very similar results.

## B.2 Description of Production Data

European production data are obtained from UNIDO at the 4-digit ISIC industry level. Data on U.S. production are obtained at the 6-digit industry level from the BEA. In order to map these production data to 6-digit HS codes, we use a concordance between ISIC and HS obtained from WITS (http://wits.worldbank.org) and a concordance between BEA-industries and HS codes made available by Thibault Fally. To map production data into HS codes, we make a proportionality assumption whereby the export-to-output ratio is assumed to be constant within each ISIC code. We then compute output for each product category using our concordance between HS codes and Nielsen product modules. A similar procedure is used for the US.

## B.3 Protocol for Merging Nielsen and Comtrade Datasets

Given the differential classifications used in Nielsen's Consumer Panel and the COMTRADE database, we need to construct a concordance between the two datasets. This was done using a complete list of Nielsen's 1,147 product modules and 5,226 HS 6-digit commodities. A majority of HS codes correspond to intermediate goods and non-barcoded consumer goods, and therefore do not match to Nielsen product modules. The merge was carried out using online tools such as the US Census Bureau's Schedule B Search Engine[22] and the Canadian Importers Database[23], which can identify relevant HS codes for a given product. We aimed to produce the largest number of merged categories possible, while ensuring all relevant Nielsen modules and HS commodities are included within each category. The resulting concordance contains 324 distinct categories, spanning 1,147 Nielsen product modules and 878 HS 6-digit commodities. Table A2 lists the number of categories by merge type (i.e. 1:1, 1:n, m:1, m:n). Our main analytical sample is a subset of 232 categories, due to missing values for Chinese import penetration in Europe (32 categories, almost all fresh foods), missing values for inflation (40 categories that had zero expenditure in one or more years), and finally excluding those with extreme values in our dependent and explanatory variables (20 categories).[24] To illustrate the protocol used for carrying out this exercise, we hereby discuss some examples for each type of merge.

### B.3.1 1:1 Merges

This is the simplest type, where a Nielsen product module fits exactly into an HS code, and there is no other product module that fits into the same HS code. For example, to find the HS codes corresponding to the product module "Fresh Apples" (4010), we type in "apples" in the Canadian Importers Database Search Engine, which reveals four items: "080430 - Pineapples

[22] https://uscensus.prod.3ceonline.com/

[23] https://www.ic.gc.ca/app/scr/ic/sbms/cid/searchProduct.html?lang=eng

[24] This is defined as either less than the 1st, or greater than the 99th, percentile values.

- Fresh Or Dried"; "080810 - Apples - Fresh"; "081330 - Apples - Dried"; and "200820 - Pineapples Nes - Prepared, Whether Or Not Sugared, Sweetened or Spirited". This yields a 1:1 merge with the HS 6-digit commodity "080810 - Apples - Fresh".

Similar merges are found for a number of other fresh fruits and vegetables, such as oranges, strawberries, carrots, potatoes, and so on. Furthermore, there are also some 1:1 merges for products that are neither food nor drink. For instance, the product module "shelf paper and wall coverings" (7325) and the HS code "wallpaper and similar wall coverings" (481490) results in the category "wallpaper"; while the product module "toaster and toaster oven appliance" (7756) and the HS code "toasters, electrical" (851672) yields the category "toasters". Similarly, the product module "vacuum and carpet cleaner appliance" (7772) and the HS code "vacuum cleaners, including dry and wet vacuum cleaners, with self-contained electric motor" (850910) merge into a single category "vacuum cleaners".

### B.3.2 1:n Merges

This is the case where a Nielsen product module has more than one HS 6-digit counterpart. We follow a similar procedure to the one above. For instance, to find the HS codes corresponding to the product module "batteries" (7870), we search through the entire set of HS 6-digit commodities, using both Excel and the Search Engines mentioned above. This reveals the relevant HS codes to be as follows: "Primary cells and primary batteries, manganese dioxide (850610)"; "Primary cells and primary batteries, mercuric oxide (850630)"; "Primary cells and primary batteries, silver oxide (850640)"; "Primary cells and primary batteries, lithium (850650)"; "Primary cells and primary batteries, air-zinc (850660)"; "Primary cells and primary batteries, n.e.s. in 85.06 (850680)". Other similar cases of this type include "printers", "heating appliances", and "fridges and freezers".

### B.3.3 m:1 Merges

This is the case where multiple Nielsen product modules map into a single HS 6-digit commodity. It is relatively rare, compared to the other types. One such example is "food processors", where five Nielsen product modules ("RBC food processor and grinder appliance" (6063), "blender appliance" (7757), "mixer appliance" (7758), "juicer appliance" (7760) and "food processor and grinder appliance" (7763)) map into the HS commodity "Food grinders and mixers; fruit/veg. juice extractors, dom., with self-contained elec. motor" (850940). Another example is "creams and cosmetics", where a number of Nielsen product modules ("hand cream"; "hand and body lotions"; "baby care products - lotions", etc.) map into the HS commodity "beauty/make-up preps and preps for the care of the skin, including sunscreen/sun tan preps" (330499).

### B.3.4 m:n Merges

These are the remaining cases where certain products are classified along different dimensions by Nielsen and COMTRADE. For instance, "tea" is classified into "tea - herbal - instant", " tea - herbal bags", "tea - packaged", "tea - bags", "tea - mixes", "tea - instant" and "tea - herbal packaged" in the Nielsen data, while it is classified into "tea, green (not fermented), whether or not flavoured, in immediate packings of a content not >3kg", "tea, green (not fermented), whether or not flavoured, in immediate packings of a content >3kg", "tea, black (fermented), whether or not flavoured, in immediate packings of a content not >3kg", "tea, black (fermented), whether or not flavoured, in immediate packings of a content >3kg" in the trade data. Similarly, "coffee" is classified into "ground and whole bean coffee", "coffee - soluble flavored", "coffee - soluble" in the Nielsen data, while it is classified into "coffee - not roasted, not decaffeinated", "coffee - not roasted, decaffeinated", "coffee - roasted, not decaffeinated", and "coffee - roasted, decaffeinated" in the trade data.

# C Aggregation

This section gives more details on the aggregation performed in the main part of the paper. We start from the predicted change in the aggregate price index induced by the China trade shock:

$$\widehat{\Delta\log(P)}^{Chn} = \sum_i \omega_i \widehat{\Delta\log(P_i)}^{Chn}$$

We next assume that a foreign trade shock only affects prices if there is a change in the domestic share of expenditure. In other words, the level effect in the second stage is zero:

$$\widehat{\Delta\log(P_i)}^{Chn} = \hat{\beta}_{IV}\Delta\widehat{\log(DSE_i)}^{Chn}$$

where $\hat{\beta}_{IV}$ is the second-stage coefficient estimate, and $\Delta\widehat{\log(DSE_i)}^{Chn}$ is the change in log domestic share caused by the China trade shock.

From the first-stage equation, the predicted decline in the DSE takes the form $\Delta\widehat{\log(DSE_i)} = \hat{\mu}+\hat{\delta}IP_i+\hat{\rho}Z_i$. While the second component ($\hat{\delta}\times IP_i$) captures the degree of a differential change in the DSE across categories due to the China trade shock, the estimated constant $\hat{\mu}$ captures the common change in the DSE across categories, which may or may not be related to supply shocks in China. We thus express the decline in the DSE of category $i$ that is a result of the China trade shock as

$$\Delta\widehat{\log(DSE_i)}^{Chn} = c+\hat{\delta}\times IP_i,$$

where the constant $c$ is unknown. This is assuming that the controls $Z_i$ (U.S. productivity growth, average income and age of U.S. consumers) do not change as a response to Chinese supply changes.

To solve for $c$, we assume that in the aggregate time series, the China shock is responsible for the same share of the decline in the domestic share of expenditure as in the cross-section among product categories. That is,

$$\widehat{\Delta\log(DSE)}^{Chn} = \kappa\Delta\log(DSE),$$

where $\kappa$ is the share of the cross-sectional variance in $\Delta\log(DSE_i)$ that is due to supply shocks in China.[25] In our sample, $\kappa$ takes on a value of 0.535. The decline of the aggregate domestic share of expenditure is $\Delta\log(DSE) = -0.057$. This implies that $\widehat{\Delta\log(DSE)}^{Chn} = -0.030$. That is, roughly half of the aggregate decline in the US domestic share of expenditure is attributed to supply shocks in China.

We then approximate the aggregate decline in the domestic share of expenditure due to the China shock as follows:[26]

$$\widehat{\Delta\log(DSE)}^{Chn} = \sum_i \frac{\omega_i DSE_i}{DSE}\widehat{\Delta\log(DSE_i)}^{Chn} = c + \hat{\delta}\sum_i \frac{\omega_i DSE_i}{DSE} IP_i$$

The expression $\sum_i \frac{\omega_i DSE_i}{DSE} IP_i$ can be computed directly from the data and equals 0.025. Using the previous result $\widehat{\Delta\log(DSE)}^{Chn} = -0.030$ and the estimate $\hat{\delta} = -2.677$, we then obtain a value for the constant term of $c = 0.031$. This allows us to compute the aggregate price decline due to the China shock as

$$\widehat{\Delta\log(P)}^{Chn} = \hat{\beta}_{IV}\sum_i \omega_i \widehat{\Delta\log(DSE_i)}^{Chn}$$

Doing so, we find an aggregate decline in the ideal price index of 2.1 percentage points for the period 2004-2015, or 0.19 percentage points per year.

[25] That is, $\kappa = \frac{Var(\widehat{\Delta\log(DSE_i)}^{Chn})}{Var(\Delta(\log(DSE_i)))} = \frac{Var(\hat{\delta}IP_i)}{Var(\Delta(\log(DSE_i)))}$.

[26] The aggregate domestic share of expenditure can be written as $DSE = \sum_i \omega_i DSE_i$. Taking a first-order approximation gives $\Delta\log(DSE) = \sum_i \frac{\omega_i DSE_i}{DSE}\Delta\log(DSE_i)$