Gains from Trade and the Food Engel Curve*

Farid Farrokhi
Purdue University

David Jinkins
Copenhagen Business School

Chong Xiang
Purdue University

September 7, 2022

Abstract

This paper examines the extent to which gains-from-trade predictions from commonly-used trade theories are consistent with observed household consumption decisions. Our approach is based on inference from household-level estimation of food Engel curves in the US and in a few other countries. For a given price index as the deflator of income, deviations from food Engel curves indicate how biased that price index is relative to the correct deflator. We construct open-economy price indices based on trade theory and data, evaluate their biases according to our approach, and compare them with the bias of official CPI statistics. We find that theory-consistent open-economy price indices that account for industry-level heterogeneity and input-output linkages tend to eliminate a large fraction of the bias of CPI.

Keywords: Food Engel Curves, Price Indices, Household-level Consumption, Gains from Trade

1 Introduction

Studying the relationship between international trade and the welfare of nations is a venerable tradition in economics. A large rigorous literature has predicted how much consumers gain from international trade.[1] The literature has also shown that the gains-from-trade predictions from quantitative general-equilibrium models depend on whether these models allow for certain elements such as input-output linkages (e.g. Costinot and Rodríguez-Clare (2014), or CRC). To what extent are observed consumption decisions of households consistent with the predictions of the gains-from-trade literature? Which model specification produces predictions that track observed household consumption decisions more closely? The goal of this paper is to answer these two interrelated questions. Our results provide a validity check for gains-from-trade predictions, and may aid governments in evaluating the usefulness of quantitative trade models for policy analyses.

*This research was supported by the Danish Independent Research Fund grant no. 8019-00031B. We thank Sunham Kim and Danyang Zhang for their excellent research assistance. We are grateful to Hamid Firooz as well as participants in various seminars for their helpful feedback. All errors are our own. Email: ffarrokh@purdue.edu, dj.eco@cbs.dk, and cxiang@purdue.edu.

[1] These studies exploit detailed import data such as Feenstra (1994), and Broda and Weinstein (2006) or calibrated quantitative general-equilibrium models such as Eaton and Kortum (2002); Arkolakis et al. (2012); Caliendo and Parro (2015).

Because many studies in this literature generate their predictions using general equilibrium models calibrated to the gravity equation of trade, we need other empirical regularities to validate them, preferably regularities as well-established as the gravity equation. We use the food Engel curve, also known as Engel's law (Engel, 1895). The food Engel curve is a strong empirical regularity in micro household survey data,[2] indicating that increases in income lead to a lower share of expenditure on food. A large literature makes inferences based on the food Engel curve, for example, to recover the bias in the consumer price index (CPI) in the United States (Costa, 2001; Hamilton, 2001) and the bias in measured purchasing power parity deflators across countries (Almås, 2012), or to evaluate the accuracy of official statistics on inflation and growth (Nakamura et al., 2016). Using the tools from this empirical literature, we first describe how to recover the average bias of a given price index, as its overall deviation from the "true price index" that correctly deflates income when we estimate a stable Engle curve with household-level consumption data. Next, we construct price indices based on standard models from the gains-from-trade literature, which we call open-economy price indices (OPIs). We then use OPIs as income deflators in the estimation of the food Engel curve. With this procedure, we are able to rank and compare the average biases of the OPIs against the benchmark bias of official CPI.

In constructing the OPIs, we distinguish between the following two approaches in the literature. The first uses sufficient statistics (e.g. Arkolakis et al. 2012, or ACR). We refer to the resulting price indices as OPI-T, to stress that they rely relatively more on theory. The second approach uses disaggregated data on the values and quantities of imported products (e.g. Feenstra 1994; Broda and Weinstein 2006). We refer to the price indices based on this approach as OPI-M, to emphasize their substantive requirement for imports data. Although these two approaches stem from a common theoretical foundation, the implementations of the OPI-T and OPI-M indices require different data. As a result, the OPI-T and OPI-M indices provide complementary ways to evaluate predictions of the gains-from-trade literature.

The OPI-T indices are tightly linked to general equilibrium trade models. We examine OPI-T indices from three model specifications that are commonly used in this literature.[3] The first has a single sector for non-food merchandise. We refer to this specification as the single-sector model,[4] and its price index as OPI-T1. The second specification, the multi-sector model, has multiple non-food-merchandise sectors, and its price index is OPI-T2. The last specification augments the multi-sector model with sectoral input-output linkages, and its price index is OPI-T3. By construction, the differences between the values of the OPI-T1 and -T2 indices reflect the different gains-from-trade predictions of having multiple sectors in the model vs. having a single sector, and those between OPI-T2 and -T3 capture the contributions of having sectoral input-output linkages in the model.

[2] A large literature (e.g. Almås (2012); Deaton and Muellbauer (1980, 1986); Banks et al. (1997); Hamilton (2001); Fajgelbaum and Khandelwal (2016); Nakamura et al. (2016)) has shown that the food Engel curve is an approporiate characterization of household consumption for low-, middle- and high-income countries.

[3] e.g. Eaton and Kortum (2002) has a single-sector model with input-output linkages, and Caliendo and Parro (2015) has a multi-sector model with input-output linkages. Costinot and Rodríguez-Clare (2014) provide a survey.

[4] We also have a separate sector for food, because we estimate the food Engel curve. A single non-food-merchandise sector is as close as we can get to a specification with a single sector for all merchandise.

By estimating the average biases of OPI-T1 through -T3 using the food Engel curve, we are able to study whether, and by how much, specific model elements move the model predictions about real consumption to better track observed household consumption decisions.

Turning to OPI-M, our first specification follows the methodology of Feenstra (1994) and Broda and Weinstein (2006), which is an extension of the Sato-Vartia index (Vartia, 1976). We refer to the resulting index as OPI-M1. Our second specification, OPI-M2, follows Redding and Weinstein (2020)'s (RW) corrections of its precedent methodology. The Feenstra-Sato-Vartia methodology assumes away movements of demand residuals when assigning weights to the price of each product, whereas RW recovers demand residuals and uses them in the calculation of price indices. By construction, the differences in the values of the OPI-M1 and -M2 indices reflect the contributions of RW's corrections. By estimating the average biases of OPI-M1 and -M2 from the food Engel curve, we study whether, and by how much, RW's corrections move the overall price index closer to the true price index that consumers use in making consumption decisions.

The data required for constructing the OPI indices are from standard sources, such as the BACI-CEPII data for bilateral trade, STAN data for sectoral production, and WIOD data for input-output linkages. In addition, we obtain micro data of nationally-representative household consumption surveys for estimating food Engel curves. We have gathered such micro data for several countries, consisting of the United States, South Korea, Canada, the United Kingdom, and Peru.

Our first result is that both OPI-T and OPI-M indices track U.S. household consumption decisions better than the official CPI. For example, the use of OPI-T3 reduces the magnitude of official CPI's average bias by 48% for the U.S., and OPI-M2 reduces the magnitude by 28%. Our results provide evidence that household consumption decisions incorporate not only the average consumer price changes over time as measured by CPI, but also core elements from the gains-from-trade literature, such as changes in the number of available varieties.

A large literature has shown that the quality and new-goods biases are important reasons why the U.S. CPI overstates the true inflation (e.g. Moulton 1996; Hausman 2003). These biases likely affect other advanced countries' CPI's, too, because they are constructed in relatively similar ways to the U.S. CPI (e.g. Handbury et al. 2013). While these biases are well recognized, they have proved difficult to tackle. The sufficient-statistics approach described above has the potential of mitigating these biases, because the gains-from-trade predictions may capture all changes in quality and new goods that originate in foreign countries, under a set of stylized assumptions. By incorporating these predictions about real consumption into over-time changes in price indices, OPI-T operationalizes the complementarity of the gains-from-trade studies for the literature that focuses on CPI measures, and so might be useful for national statistical agencies as well.

Our second result is that the magnitude of the average bias in the U.S. decreases from OPI-T1 through OPI-T3. For example, while the use of OPI-T1 increases the magnitude of official CPI's average bias by 3.7% for the U.S., OPI-T2 decreases it by 18.3%, and OPI-T3 shrinks it by 48.0%. More specifically, having multiple sectors in the model moves the overall OPI-T index closer to the true price index by 22.0%, relative to having a single sector, and having sectoral input-output linkages moves it even closer, by an additional 29.7%. Moreover, the average bias of

OPI-T3 is not statistically significant for the U.S. These results not only show that the specifications of multiple sectors and input-output linkages tend to help the gains-from-trade predictions of general-equilibrium trade models better track household consumption decisions, but also quantify the contributions of these model elements.

Third, the magnitude of the average bias decreases from OPI-M1 to OPI-M2 for the U.S. OPI-M1 decreases the magnitude of the official CPI's average bias by 5.9%, while OPI-M2 shrinks it by 28.0%. These results show that the corrections to the Feenstra-Sato-Vartia index by RW move the overall price index closer to true price index by 22.1%. This finding complements results in RW which show that the correction for time-varying demand shocks makes a substantial difference in the implied price index.

While our main focus is on the United States, we have also calculated OPI-T for other countries in our sample. Averaging across our results for other countries, we find that OPI-T1 performs somewhat worse that CPI on average, but OPI-T2 and OPI-T3 are generally closer than CPI to the true price index. The OPI-M's are more difficult to construct in the smaller non-US countries in our sample because domestic currency shocks can lead to large fluctuations in import prices which are not immediately reflected in prices paid by consumers. Since our data is based on dollar prices at the dock, we must make pass-through assumptions in order to make our data consistent with consumer prices. This is less of a concern in the United States because many international transactions are themselves indexed in dollars. We will return to this issue in Sections 6-7.

We have also conducted several validation and robustness exercises. We briefly discuss one using the Nielsen home scanner data here, and refer the readers to Sections 6 and 7 for other robustness checks. The Nielsen home scanner data record the values and quantities of barcode-level products purchased by a nationally representative panel of consumers.[5] As a result, even though this data set has limited coverage for non-food merchandise, it is widely regarded as the best available measure of actual prices faced by households. We pair the Nielsen data with the methodology of RW and use the resulting price index as the benchmark to rank the OPI-T indices and CPI, and then the OPI-M indices and CPI, for the U.S. food sector for 2004-2015. We find that this approach delivers similar rankings between OPIs and CPI, as compared with the use of the average bias from the food Engel curve.

The estimation of gains from trade is the focus of a large literature.[6] Feenstra (1994) started the import-price index approach, which uses product-level trade data, and ACR shows the usefulness of sufficient-statistic approach in general equilibrium trade models.[7] We show that households behave as if they take the real-consumptions-gains predictions from this literature into account when they

[5]These results are based on the authors' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are the authors', and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

[6]In addition to the papers cited above, see also Arkolakis et al. (2008); Balistreri et al. (2011); Ossa (2015); Hsieh and Ossa (2016); Levchenko and Zhang (2016), and Giri et al. (2021).

[7]Recent studies have also examined the pro-competitive effects of trade, through markups; e.g. Feenstra and Weinstein (2017); Jaravel and Sager (2019). We do not explore this channel in our study.

make consumption decisions, as revealed by the food Engel curve.[8] We also clarify how specific model elements help the overall gains-from-trade predictions track household consumption decisions. While the model elements in OPI-T1 through -T3 by no means cover every model variation analyzed in the literature, they comprise the most common and well-researched specifications, and the range of their predictions spans a substantial portion of the range reported in the literature.[9]

An emerging literature provides reduced-form evidence on the effects of Chinese imports on U.S. consumer prices, using the core approaches of the gains-from-trade literature as the starting point. For example, Amiti et al. (2020) use the import-price-index approach, and Bai and Stumpner (2019) use the sufficient-statistics approach. We provide external validations of these core approaches. Our results may also be useful for the recent studies that extend the applications of sufficient statistics to a variety of topics.[10]

Kehoe et al. (2015) and Kehoe et al. (2017) compare the observed changes in trade flows with those predicted by the studies that use applied-general-equilibrium (AGE) models. Their work and ours share the same motivation of providing external validations to quantitative trade models. On the other hand, the AGE literature is distinct from the gains-from-trade literature that we discuss above.[11]

A literature uses the food Engel curve to estimate the biases in a single price index, typically CPI.[12] This literature tends to focus on measurement, rather than the underlying theoretical mechanisms of the estimated biases. We use the food Engel curve to compare across price indices, and explicitly derive the OPI indices using trade theory. Our results suggest that the estimated CPI biases could be related to mismeasured gains from international trade in addition to the more traditional price measurement issues.In the literature that examines the biases of the U.S. CPI, quality and new goods are widely regarded as the key issues. Many proposed solutions are for specific products or product categories.[13] We show that the sufficient-statistics approach of the trade literature, which accounts for all foreign-based quality changes and new goods, can be implemented as the OPI-T indices. OPI-T draws on readily available data and generally outperforms CPI in tracking

[8] Since we examine annual data, our inference is more relevant for predictions that come from small changes in model parameters, as opposed to extreme counterfactuals like moving to autarky.

[9] For example, industry-level scale economies tend to have small effects on gains-from-trade predictions (e.g. CRC, Kucheryavyy et al. 2016; Farrokhi and Soderbery 2022). Mixed CES, which allows trade elasticity to differ across country pairs, also has limited effects (e.g. Costinot and Rodríguez-Clare 2018). Trade in commodities can bring about larger gains from trade (e.g. Farrokhi 2020; Fally and Sayre 2018). Meanwhile, firm entry and exit have small effects in some settings (e.g. CRC) but larger effects in others (e.g. Melitz and Redding 2015).

[10] Examples include infrastructure (e.g. Donaldson, 2018), contract frictions (e.g. Chor and Ma, 2020), economic growth (e.g. Hsu et al., 2019; Anderson and Yotov, 2020), multi-national production (e.g. Arkolakis et al., 2018; Du and Wang, 2021), correlated productivity shocks (e.g. Lind and Ramondo, 2018), costly labor mobility (e.g. Caliendo et al., 2019), misallocation (e.g. Bai et al., 2019), global value chains (e.g. Antràs and De Gortari, 2020), carbon emissions (e.g. (Shapiro, 2016)), technology adoption (e.g. Farrokhi and Pellegrina, 2022) and search frictions (e.g. Krolikowski and McCallum, 2021; Eaton et al., 2021, 2022).

[11] Examples of the AGE literature include Brown et al. (1989); Markusen et al. (1995), and Hertel (2013).

[12] In addition to the papers cited above, see also Beatty and Larsen (2005); Larsen (2007); Barrett and Brzozowski (2010); Chung et al. (2010); Clerc and Coudin (2011); de Carvalho Filho and Chamon (2012); Sacerdote (2017), and Dabalen et al. (2020).

[13] In addition to the papers cited above, see also Shapiro and Wilcox (1996); Boskin et al. (1997); Moulton et al. (1997); Berndt (2006); Gordon (2006); Greenlees and McClelland (2011); Melser and Syed (2016).

household consumption for multiple countries and sample periods.

A recent literature estimates the exact cost-of-living index, or welfare, under non-homothetic preferences. Baqaee and Burstein (2021) show that it is difficult to tackle non-homothetic preferences and mis-measured prices simultaneously, and so this literature typically assumes that well-measured prices are available.[14] Atkin et al. (2020), for example, use relative Engel curves and well-measured prices of a subset of goods.[15] In comparison, we start from the premise that well-measured prices, adjusted for quality and variety, are difficult to obtain for most merchandise goods, and so our goal and our approach are different from this literature. Still, our results and those of Atkin et al. (2020) both show that the price indices that correct for quality and variety show lower increases over time, relative to official CPI.

The remainder of this paper is organized as follows. Section 2 shows how we use the estimation of the food Engel curve to obtain the average bias, and how this metric allows us to rank and compare measured price indices. Section 3 sketches our basic idea for calculating OPIs using a simple example. Section 4 presents our full model and our procedure of calculating OPIs in detail, and discusses threats to identification. Section 5 describes our data, presents noteworthy data patterns, and conducts some preliminary data analysis. Sections 6 and 7 report our results for the U.S. and for non-U.S. countries, respectively, and discusses the implications of our results. Section 8 concludes.

2 Comparing Price Indices Using Household Consumption Data

In this section, we first present the basics of the food Engel curve, and how it is used to recover the true price index, as in Costa (2001) and Hamilton (2001). We then show how we adapt the Costa-Hamilton approach to rank and compare across measured price indices.

2.1 Households' Consumption and Food Engel Curve

Consider households indexed by h within a country. Household h's utility in year t is given by the indirect utility function, $U_h^t = U(E_h^t, \{\tilde{P}_F^t, \tilde{P}_N^t\})$, where $\tilde{P}_F^t$ and $\tilde{P}_N^t$ denote the price indices of food and non-food that are common across households. Household h's expenditure function is $E_h^t = E(U_h^t, \{\tilde{P}_F^t, \tilde{P}_N^t\})$, and its functional form depends on that of the household utility, $U(.)$. Rather than specifying the exact preferences, we note that all non-homothetic preferences share the following second-order expenditure function approximation (Deaton and Muellbauer, 1980),

$$\ln E(\tilde{P}_F^t, \tilde{P}_N^t, U_h^t) = \alpha_0 + (\alpha_F \ln \tilde{P}_F^t + \alpha_N \ln \tilde{P}_N^t) + \frac{\gamma}{2}(\ln \tilde{P}_F^t - \ln \tilde{P}_N^t)^2 + U_h^t \beta_0 (\tilde{P}_F^t / \tilde{P}_N^t)^{\beta_F},$$

[14] See also Fajgelbaum and Khandelwal (2016); Borusyak and Jaravel (2018); Almås et al. (2018); Argente and Lee (2021); Auer et al. (2022).

[15] In theory, computation remains valid without well-measured prices under an orthogonality condition, but well-measured prices are needed to test this condition. In the data, Atkin et al. (2020) show that this orthogonality condition is a good approximation for rural Indian households.

where $\{\alpha_0, \alpha_F, \alpha_N, \gamma, \beta_0, \beta_F\}$ are time-invariant parameters and $\alpha_N = 1 - \alpha_F$. Let $Q^t_{h,F}$ denote the quantity of food consumption by household h at t. Then the above equation implies the following expenditure share of food, $\lambda^t_{h,F} \equiv \tilde{P}^t_F Q^t_{h,F}/E^t_h$,

$$\lambda^t_{h,F} = \alpha_F + \beta_F(\ln E^t_h - \ln \tilde{P}^t) + \gamma(\ln \tilde{P}^t_F - \ln \tilde{P}^t_N), \tag{1}$$

where

$$\ln \tilde{P}^t = \alpha_0 + (\alpha_F \ln \tilde{P}^t_F + \alpha_N \ln \tilde{P}^t_N) + \frac{\gamma}{2}(\ln \tilde{P}^t_F - \ln \tilde{P}^t_N)^2. \tag{2}$$

Equation (1) is the theoretical motivation for a large body of work that estimates food Engel curves. The sign and magnitude of the income coefficient, β_F, indicate whether food is a luxury ($\beta_F > 0$), a necessity ($-\overline{\lambda}_F < \beta_F < 0$, where $\overline{\lambda}_F$ is the mean food expenditure share), or an inferior good ($\beta_F < -\overline{\lambda}_F$). The relative-price coefficient, γ, can be either positive or negative, depending on the demand elasticity of food.[16] Finally, the price index, $\tilde{P}^t$, outlined in equation (2), is a collection of first- and second-order terms of prices that make equation (1) look intuitive. The tilde notation emphasizes that it represents the *true* values of prices, as used by households to deflate income in their consumption decisions, as opposed to prices measured by researchers. Deaton and Muellbauer (1980) show that one can get a reasonable approximation of $\tilde{P}^t$ by using first-order terms only, i.e. $\ln \tilde{P}^t \approx \alpha_0 + (\alpha_F \ln \tilde{P}^t_F + \alpha_N \ln \tilde{P}^t_N)$, and the use of this approximation is common in the literature (e.g. Nakamura et al. 2016).

The variable $\tilde{P}^t$ is not the cost-of-living index, for two reasons. One, we have not specified the exact system of preferences. Two, even if the demand system were AIDS so that equations (1)-(2) were exact, the cost-of-living index would be necessarily household specific (since demand is non-homothetic), whereas $\tilde{P}^t$ is common to all households.[17] Nonetheless, $\tilde{P}^t$ captures the way households deflate their nominal expenditure, and it is the focal point of the Costa-Hamilton approach, outlined below. We follow this literature and refer to $\tilde{P}^t$ as the true price index.

2.2 Evaluating Biases in Measured Prices Using Engel Curves

In order to estimate the food Engel equation (1), we use data on observed households' total expenditure, E^t_h, and food expenditure share, $\lambda^t_{h,F}$, and follow the common practice in the literature to introduce the following additional elements. First, we assume that the true price index, $\tilde{P}^t$, is observed with a potential bias:

$$\ln \tilde{P}^t = \ln P^t + \ln B^t \tag{3}$$

[16] Equation (1) implies that (i) the income elasticity of food for the average household is $\frac{\partial \ln \overline{Q}_F}{\partial \ln E} = 1 + \frac{\beta_F}{\overline{\lambda}_F}$ where $\overline{Q}_F$ is the mean food consumption quantity, and (ii) the own price elasticity of food for the average household is $\frac{\partial \ln \overline{Q}_F}{\partial \ln \tilde{P}_F} = \frac{1}{\overline{\lambda}_F}(\gamma - \beta_F(\alpha_F + \ln \frac{\tilde{P}_F}{\tilde{P}_N})) - 1 \approx \frac{1}{\overline{\lambda}_F}(\gamma - \beta_F \alpha_F) - 1$,where the approximation is based on the first-order approximation of $\ln \tilde{P}$.

[17] In a related context, Almås et al. (2018) and Atkin et al. (2020) develop methods to recover a cost-of-living index that can be used for welfare analysis. We can connect our examination to welfare analysis under the assumption that the demand system is AIDS. In that case, the change to the utility of household h in time t, is given by $d \ln U_h = d\left(\frac{\ln(E_h/\tilde{P})}{(\tilde{P}_F/\tilde{P}_N)^{\beta_F}}\right)$ which can be calculated with knowledge of $E_h, \tilde{P}$, $(\tilde{P}_F/\tilde{P}_N)$, and β_F.

where P^t is some *measured* price index, and B^t is its associated bias. We incorporate the biases B^t in the estimation by adding year dummies, D^t. Second, we add household characteristics Z_h^t as controls. Finally, let r index regions within a country. We assume that, once we include regional data on the relative price of food, $\ln(P_{r,F}^t/P_{r,N}^t)$, and control for region fixed effects, α_r, there is no systematic bias for the term $\ln \tilde{P}_F^t - \ln \tilde{P}_N^t$ in equation (1). Adding these elements, and an error term ε_{hr}^t, into equation (1), we obtain:

$$\lambda_{hr,F}^t = \alpha_F + \beta_F(\ln Y_{hr}^t - \ln P^t) + \theta Z_{hr}^t + \alpha_r + \gamma \ln(P_{r,F}^t/P_{r,N}^t) + \sum d^t D^t + \varepsilon_h^t, \quad \text{where} \quad d^t = -\beta_F \ln B^t. \tag{4}$$

Equation (4) shows that the year dummy estimates, d^t —also, sometimes called "drifts"— are informative about the bias B^t in equation (3). Note that d^t and $\ln B^t$ are both normalized to 0 for the initial year in the sample, $t = 0$, with the aid of the constant, α_F, in the estimation of (4).

2.3 Comparing Across Measured Price Indices

We have now reached the fork where our framework diverges from the Costa-Hamilton approach. In their approach, one uses official CPI to estimate equation (4), compute the bias, B^t, and recover the true price index, $\tilde{P}^t$, using equation (3). While the recovered $\tilde{P}^t$ is likely to have useful policy implications, the method is silent on which theoretical model can deliver it.

In comparison, we are interested in comparing across measured price indices, such as the official CPI, and those implied by the gains-from-trade literature, where the true price index, $\tilde{P}^t$, serves as the benchmark. To be specific, let the subscript "m" denote measured price indices. For each m, we estimate the food Engle curve, equation (4), and obtain m-specific coefficient estimates, $\beta_{F,m}$, θ_m, γ_m and d_m^t. Below, we spell out the useful properties of these estimates that facilitate making comparisons across measured price indices.

First,

$$\beta_{F,m}, \theta_m, \gamma_m \text{ are invariant across } m. \tag{5}$$

Intuitively, expression (5) holds because for all m, the measured price index itself, P_m^t, varies by year only, while β_F and θ are identified by variation across households, and γ by variation across regions. Expression (5) enables us to focus on the differences in the year-dummy estimates, d_m^t, across measured price indices.

Second, the mean value of d_m^t quantifies how well measured price index m tracks the true price index, $\tilde{P}^t$. To see this point, equations (3), (4) and (5) imply that

$$\underset{\text{explained by } m}{\ln P_m^t} + \underset{\text{unexplained}}{d_t^m/(-\beta_F)} = \underset{\text{needs to be explained}}{\ln \tilde{P}^t} \quad \text{for all } m$$

This expression is another way to show equation (3), and clarifies that, the more closely index m, P_m^t, tracks the true price index, $\tilde{P}^t$, the smaller is the magnitude of the year-dummy estimate, d_t^m. In this sense, the year-dummy estimates are the residuals that are unexplained by measured price indices, relative to the true price index of $\tilde{P}^t$. Let T denote the number of time periods in the estimation of (4).[18] Summing across periods and dividing by $(T-1)$, we obtain

$$\underbrace{\overline{\ln P_m^t}}_{\text{explained by } m} + \underbrace{AB_m}_{\text{unexplained}} = \underbrace{\overline{\ln \tilde{P}^t}}_{\text{need to be explained}}, AB_m = \frac{\sum_t d_m^t}{(T-1)(-\beta_F)} = \overline{\ln B_m^t}. \quad (6)$$

Equation (6) says that if average bias $AB_m < 0$, P_m^t is lower than $\tilde{P}^t$ on average; i.e. the measured index m overstates the true prices and understates consumers' real income, and vice versa if $AB_m > 0$. Because d_m^t are coefficient estimates from regression (4), their standard errors allow us to compute the standard error of the average bias, and then test whether index m is statistically distinguishable from the true price index, $\tilde{P}^t$ (i.e. $AB_m = 0$).

Equation (6) also implies that:

$$\left|\frac{AB_m}{AB_1}\right| = \left|\frac{\overline{\ln \tilde{P}^t} - \overline{\ln P_m^t}}{\overline{\ln \tilde{P}^t} - \overline{\ln P_1^t}}\right| \quad (7)$$

Equation (7) says that if the relative average bias, $|AB_m/AB_1|$, is small in magnitude, the mean deviation of index m from the true price index, relative to the benchmark index 1, is also small in magnitude; i.e. index m tracks household consumption decisions well, relative to the benchmark index 1. Equation (7) thus allows us to rank measured price indices. Note that the relative average bias is in absolute value, and so it is not affected by the signs of the average biases, AB_m and AB_1.

Third, the difference in the average bias provides an exact decomposition of the additional explanatory power of any measured price index m relative to a benchmark index, 1. To see this point, equation (6) implies that, for indices m and 1,

$$\underbrace{\overline{\ln P_m^t}}_{\text{explained by } m} + \underbrace{AB_m}_{\text{unexplained}} = \underbrace{\overline{\ln P_1^t}}_{\text{explained by } 1} + \underbrace{AB_1}_{\text{unexplained}}$$

or,

$$\underbrace{\frac{\overline{\ln P_m^t} - \overline{\ln P_1^t}}{AB_1}}_{\text{explained}} + \underbrace{\frac{AB_m}{AB_1}}_{\text{unexplained}} = 1 \quad (8)$$

To put equation (8) into perspective, suppose, hypothetically, that we lived in a world where index 1 were the only available measure of prices. We would then be puzzled by our inability to fully capture consumers' real income, as quantified by the gap AB_1. Now, suppose we invent index m. Relative to index 1, the benchmark, index m explains some portion of AB_1, but does not explain all of it. The portion explained by m is the first term on the left-hand side of equation (8), and the portion left unexplained is the second term. If index m consists of multiple components, then the decomposition (8) applies to all the components of m; e.g. if $P_m^t = P_m^{t,1} P_m^{t,2}$, then equation (8) says that

$$\frac{\overline{\ln P_m^{t,1}} - \overline{\ln P_1^t}}{AB_1} + \frac{\overline{\ln P_m^{t,2}}}{AB_1} + \frac{AB_m}{AB_1} = 1. \quad (9)$$

[18] The average bias excludes $t = 0$ because $\ln B^0 = 0$ by normalization.

Equation (8) allows us to quantify the extent to which price index m improves on the benchmark index, and equation (9) allows us to clarify what is driving this improvement by examining the individual components of index m. Note that the portion explained by m or its components, relative to the benchmark index, could be negative, in which case m or its components deviate further from the true household prices of $\tilde{P}^t$ than the benchmark index.

In summary, equations (6) through (9) are able to quantify how well gains-from-trade predictions about real consumption track observed household consumption decisions relative to the true price index and relative to CPI, if we can translate these predictions into over-time changes in price indices.

3 Sketching the Idea: Open-Economy Price Indices (OPI)

In this section, we use a simple example to illustrate how we construct the price indices that are implied by the gains-from-trade literature. We leave out the time superscript to keep our exposition compact. We set up our full model in the next section.

Suppose that the world economy consists of multiple countries, and a single sector. Goods are differentiated by country of production, indexed by $j = 1, ..., N$. Preferences are CES with substitution elasticity σ and demand shifters b_j. Consider a country called Home. Then, the CES price index in Home is

$$P = \left[\sum_{j=1}^{N} b_j \left(p_j\right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} . \tag{10}$$

A large body of work in the gains-from-trade literature goes after the same objective, which is, in its simplest form, the change to the CES price index specified in equation (10). We denote this change between two time periods by $\hat{P}$, and refer to the resulting price index as the open-economy price index, or OPI. We now outline two approaches to compute OPI.

3.1 OPI-T

Our first approach draws on the literature that uses sufficient statistics to compute general equilibrium changes to prices. Specifically, the change to the price index, $\hat{P}$, can be expressed as:

$$\hat{P} = \hat{p}_D \left(\hat{\pi}_D\right)^{\frac{1}{\sigma-1}} , \tag{11}$$

where $\hat{p}_D$ is the change to the price of the domestic variety, adjusted by the trade channel. This trade adjustment is fully characterized by the change in the domestic expenditure share, $\hat{\pi}_D$, and the trade elasticity, $(\sigma - 1)$. Intuitively, if we observe a decrease in the domestic expenditure share, the price of tradable goods must have fallen relative to domestic inflation, since Home's consumers now spend more on imported goods. Equation (11) implies that if we have data on $\hat{p}_D$ and $\hat{\pi}_D$ between any two time periods, we can compute the corresponding change in the price index, $\hat{P}$.

A special case of Equation (11) delivers the gains from trade, defined as the loss in real income from moving the economy to autarky. This case, referred to in the literature as the ACR formula, corresponds to $\hat{\pi}_D = 1/\pi_D$ where π_D is the baseline value of the domestic expenditure share, and $\hat{p}_D = \hat{w}$ where $\hat{w}$ is the change to the nominal wage in Home. In that case, the change to real income, $\hat{w}/\hat{P}$, is given by $(\pi_D)^{-\frac{1}{\sigma-1}}$.

We refer to the OPI based on equation (11) as OPI-T, where "T" refers to theory, because OPI-T is derived from a generic general-equilibrium trade model.

3.2 OPI-M

Our second approach draws on the literature that uses detailed import data to examine the gains from trade. In this approach, we use data from foreign variety-level prices, p_j, and demand shifters, b_j, to compute the OPI in equation (10). We refer to the resulting indices as OPI-M, where "M" stands for imports.

To be specific, varieties for Home can be partitioned into domestic and imported varieties, whose price indices are, respectively, p_D and $p_M \equiv \left[\sum_{j \neq D} b_j \left(p_j\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$. This allows us to re-write equation (10) as

$$P = \left[(p_D)^{1-\sigma} + (p_M)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}. \tag{12}$$

From here, we can express the change to OPI-M as:

$$\hat{P} = \hat{p}_M \left(\hat{\pi}_M\right)^{\frac{1}{\sigma-1}}, \tag{13}$$

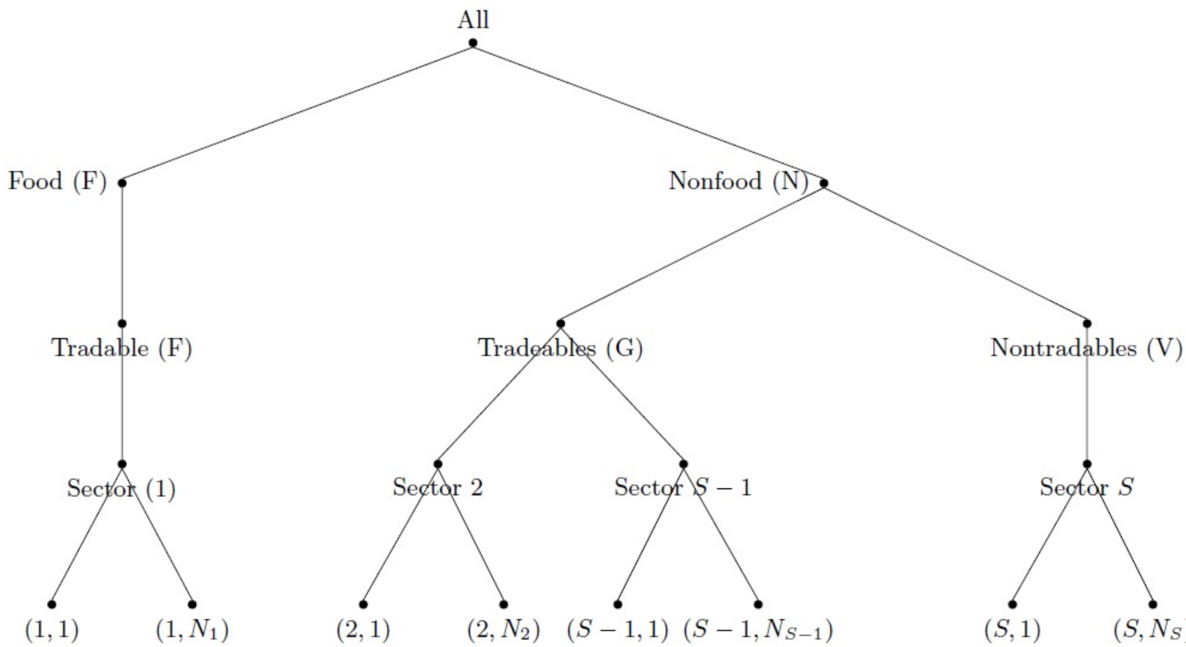
where $\pi_M = 1 - \pi_D$ is the imported share of expenditure. To see the intuition behind equation (13), suppose that import share, π^M, decreases; i.e. households have substituted away from imports towards domestic varieties for consumption. This implies that domestic prices fell relative to import prices, and so we adjusted the change in import prices downward using the drop in import share, to get the correct change of the overall index, OPI-M.

Equation (13) is the mirror image of Equation (11). This result is intuitive: OPI-T and OPI-M are theoretically two sides of the same coin. Although both calculate the change in the CES price index in Equation (10), they use different data: OPI-M uses import price data but OPI-T does not. As a result, in practice they will likely generate different values of OPI, and so provide two complementary approaches to taking the gains-from-trade predictions to household consumption data.

Finally, we have maintained the assumption of homothetic CES preferences following many studies in the trade literature.[19] In the next section, we derive OPIs using a framework that combines non-homothetic demand between aggregate bundles of food and non-food with CES demand across disaggregated sectors.

[19] Note that in this setting the price index given by equations (11) and (13) can be used to measure both real consumption and welfare. For the distinction between these two measures, see Baqaee and Burstein (2021).

Figure 1: Product Space



Notes: This figure shows the space of products and the nesting structure used in our analysis.

4 The Framework

Our full model is designed to bring the simple model of Section 3 to the data. To this end, we begin with specifying the space of products. This specification is motivated by (i) the availability of micro data that can be used to construct the OPIs, and (ii) the manner in which non-homothetic preferences, which motivate the food Engel curve, can be combined with homothetic preferences, which are the mainstay of the trade literature.

We illustrate the space of products in Figure 1. The consumption of each household at time t consists of two broad groups, food (F), and non-food (N). This non-homothetic upper-tier demand gives rise to the food Engel curve as discussed in Section 2.1. The non-food group includes two categories, tradeable merchandise (G) and nontradeable services (V), and the food group is its own category. We model services as non-tradable because the data for import prices are for merchandise only. Within categories $\{F,G,V\}$, there are multiple sectors indexed by s, with the food category F consisting of sector $s=\{1\}$, non-food tradeable category consisting of $s=\{2,...,S-1\}$, and nontradeable non-food V consisting of $s=\{S\}$. In this section, we first explain how we follow the trade literature to construct different specifications of the OPI-T and OPI-M indices for the homothetic categories F, G and V. We then show how we aggregate across categories.

We assume that the sub-utility for sectors $s=\{1,...,S-1\}$ is CES with substitution elasticity σ_s. We can then obtain the price index of a single sector s by adding the sector subscript s and time superscript t to equation (10). Let $c=\{F,G,V\}$ index categories.

4.1 OPI-T

Let the production side of the economy be the class of perfect-competition models considered in CRC, including Armington (1969) and Eaton and Kortum (2002). Following common practice in the trade literature, we assume that the non-tradable category, V, produces a homogeneous product. In this sub-section, we spell out three specifications of the OPI-T index. Our specifications and model settings closely follow CRC.

First, assume that the tradable merchandise category, G, consists of a single sector. Labor is the only factor of production, and there is no intermediate input. Then equation (11) applies to the price index of category G, with the modification that the trade elasticity specific to the Armington model, $\sigma_G - 1$, is replaced by the more general notation for the trade elasticity, θ_G. We can obtain the price index of the food category, F, in a similar way, because F consists of a single sector. Meanwhile, with perfect competition, the price change of the homogeneous product of category V is given by the domestic price index. Summarizing,

$$\hat{P}^t_c = \hat{p}^t_{D,c} \left(\hat{\pi}^t_{D,c}\right)^{\frac{1}{\theta_c}} \text{ for } c = \{F, G\}, \;\; \hat{P}^t_V = \hat{p}^t_{D,V}. \tag{14}$$

We refer to the price indices given in equation (14) as OPI-T1.This specification is as close as we can get, to the single-sector model in CRC.

Second, assume that category G consists of sectors $s = \{2, ..., S-1\}$, and that the sub-utility across these sectors, within G, is Cobb-Douglas, with consumption expenditure shares β_s. On the other hand, the other elements of the specification are the same as in OPI-T1. Then we can apply equation (11) to the price index of $s = \{2, ..., S-1\}$, replacing $\sigma_s - 1$ with the more general notation of θ_s. Taking the weighted geometric mean of these sectoral price indices, with weights β_s, we obtain the price index of category G. The price indices of categories F and V are the same as in equation (14). Summarizing,

$$\hat{P}^t_F = \hat{p}^t_{D,F} \left(\hat{\pi}^t_{D,F}\right)^{\frac{1}{\theta_F}}, \;\; \hat{P}^t_V = \hat{p}^t_{D,V}, \;\; \hat{P}^t_G = \prod_{s=2}^{S-1} \left(\hat{P}^t_s\right)^{\beta_s}, \;\; \text{where } \hat{P}^t_s = \hat{p}^t_{D,s} \left(\hat{\pi}^t_{D,s}\right)^{\frac{1}{\theta_s}}. \tag{15}$$

We refer to the price indices given in equation (15) as OPI-T2, and this specification corresponds to the multi-sector model in CRC (2014).

Finally, we add the following element into the specification of OPI-T2: sector $s = \{1, ..., S\}$ uses the outputs of all the sectors in the economy, including itself, as intermediate inputs. Following common practice in the trade literature, we assume that the production cost of sector $s = \{1, ..., S\}$ is a Cobb-Douglas aggregate of domestic prices, with share γ_s, and intermediate-inputs price, with share $1 - \gamma_s$. The latter is, in turn, a Cobb-Douglas aggregate of the price indices of all the sectors in the economy with shares α_{ks}, $k = \{1, ..., S\}$ and $\sum_{k=1}^{S} \alpha_{ks} = 1$. The input-output table of the economy is then the $S \times S$ matrix $\mathbf{A}$, whose element on row $k-$column s is $(1 - \gamma_s)\alpha_{ks}$. For sector $s = \{1, ..., S\}$:

$$\hat{P}_s^t = \hat{p}_{D,s}^t \left[\prod_{k=1}^{S} \left(\hat{\pi}_{D,k}^t\right)^{\frac{\tilde{\alpha}_{ks}}{\theta_k}} \right]. \tag{16}$$

In equation (16), $\tilde{\alpha}_{ks}$ represents the elements of the Leontief-inverse matrix of the economy; i.e. it is the row $k-$column s element of $(\mathbf{I} - \mathbf{A^T})^{-1}$, where $\mathbf{I}$ is the identity matrix and $\mathbf{A^T}$ is the transpose of the input-output matrix $\mathbf{A}$. The parameters $\tilde{\alpha}_{ks}$ reflect the importance of sector k as an intermediate input for the production of sector s. To see the intuition of $\tilde{\alpha}_{ks}$, suppose that import prices fall in sector k, such that $\hat{\pi}_{D,k}^t < 1$. Without input-output linkages, this change lowers sector k's own price index, P_k^t, by $\left(\hat{\pi}_{D,k}^t\right)^{\frac{1}{\theta_k}}$, as given in equation (15), and has no impact on the other sectors' price indices. With input-output linkages, however, the other sectors $s \neq k$ experience price decreases of $\left(\hat{\pi}_{D,k}^t\right)^{\frac{\tilde{\alpha}_{ks}}{\theta_k}}$, where $\tilde{\alpha}_{ks} > 0$, because they all use sector k's outputs as inputs. In addition, the effect on sector k's own price index is amplified to $\left(\hat{\pi}_{D,k}^t\right)^{\frac{\tilde{\alpha}_{kk}}{\theta_k}}$, where $\tilde{\alpha}_{kk} > 1$ captures the extent of the amplification, because sector k uses all the other sectors' outputs as inputs.

We now use the sectoral prices of equation (16) to construct the price indices by category. Food, F, and non-tradable services, V, are one-sector categories, and so their prices are given by (16). Tradable merchandise, G, continues to be a Cobb-Douglas aggregate of sectors $s = \{2, ..., S-1\}$, and so we can plug the sectoral prices in equation (16) into equation (15) to compute its price. Summarizing our results,

$$\hat{P}_F^t = \hat{P}_s^t \quad \text{for } s = 1, \quad \hat{P}_G^t = \prod_{s=2}^{S-1} \left(\hat{P}_s^t\right)^{\beta_s}, \quad \hat{P}_V^t = \hat{P}_s^t \quad \text{for } s = S, \quad \text{where} \quad \hat{P}_s^t \text{ is given by (16).} \tag{17}$$

We refer to the price indices given in equation (17) as OPI-T3, and this specification corresponds to the multi-sector model with intermediate inputs in CRC.

A common feature of equations (14) through (17) is that they do not require data on import transactions to account for changes in import prices. Instead, they allow us to take advantage of sectoral data on domestic expenditure shares, $\hat{\pi}_{D,s}^t$, as well as available estimates of trade elasticities θ_s.

4.2 OPI-M

As we discussed in Sub-section 3.2, the centerpiece of the OPI-M approach is the sectoral import price indices, $P_s^{M,t}$ (we have added the sector and time scripts). Once we have the import prices, we can use equation (13) to compute the overall sectoral price indices, which include both imports and domestic products, and then apply the same aggregation procedure as in Section 4.1, to obtain category price indices. Therefore, we start with sectoral import price indices.[20]

To compute $P_s^{M,t}$, we add a CES layer within each tradable sector. With this additional layer, $P_s^{M,t}$ captures the effects of variety entry and exit more accurately. To be specific, for the sectors

[20]For more details on our formulas and derivations in this subsection, see Appendix B.1.

in the tradable categories of food, F, and other merchandise, G, $s = \{1, ..., S-1\}$, the overall price index, P_s^t, is given by equation (12), with sector and time scripts added on the right-hand side of the equation. The substitution elasticity between the import bundle and domestic bundle is σ_s. The import bundle, in turn, has the price of $P_s^{M,t}$, which is a CES aggregate across imported goods, indexed by g, whose price indices are $P_{gs}^{M,t}$. The imported goods are shown as the bottom-most nodes in Figure 1, and the substitution elasticity among them, within sector s, is σ_s^M. This implies that we can construct the sector-level price, $P_s^{M,t}$, from goods-level prices, $P_{gs}^{M,t}$, using the standard Sato-Vartia aggregation:

$$\hat{P}_s^{M,t} = \prod_{g \in \Omega_s} (\hat{P}_{gs}^{M,t})^{d_{gs}^t}, \qquad \text{where} \qquad d_{gs}^t = \frac{(\lambda_{gs}^t - \lambda_{gs}^0)/(\ln \lambda_{gs}^t - \ln \lambda_{gs}^0)}{\sum_{g \in \Omega_s} (\lambda_{gs}^t - \lambda_{gs}^0)/(\ln \lambda_{gs}^t - \ln \lambda_{gs}^0)} \tag{18}$$

In equation (18), Ω_s is the set of goods within sector s, and it remains unchanged over time. λ_{gs}^t is the share of good g in import value within sector s in time t, and d_{gs}^t is the standard Sato-Vartia weights by good, computed as the log mean of λ_{gs}^t and λ_{gs}^0, the goods' import-value shares in time t and in the base period of 0. Equation (18) is common to both the FBW and RW procedures.

Each good, in turn, is differentiated into varieties, indexed by j, whose prices are $p_{j,gs}^t$. The substitution elasticity between the varieties within good g is σ_g, and their taste parameters are $b_{j,gs}^t$. The set of varieties within g, $\Omega_{gs}^{M,t}$, may change over time. Let Ω_{gs}^M denote the common set, or the set of varieties that are present in both time t and the base period 0. We can summarize FBW and RW procedures using the same expression:

$$\hat{P}_{gs}^{M,t} = (\widehat{\lambda}_{gs}^{*t})^{\frac{1}{\sigma_g - 1}} \prod_{j \in \Omega_{gs}^M} (\widehat{p}_{j,gs}^t)^{d_{j,gs}^t}, \qquad \text{where} \qquad \lambda_{gs}^{*t} = \frac{\sum_{j \in \Omega_{gs}^M} p_{j,gs}^t q_{j,gs}^t}{\sum_{j \in \Omega_{gs}^{M,t}} p_{j,gs}^t q_{j,gs}^t}. \tag{19}$$

In Equation (19), $p_{j,gs}^t q_{j,gs}^t$ is import value of variety j at time t, and λ_{gs}^{*t} is the import-value share of the common set in t. The first factor on the right-hand side of Equation (19) captures the contribution to the good-level price index from variety entry and exit. Intuitively, a rise in import values on new varieties at t implies a lower λ_{gs}^{*t} relative to λ_{gs}^{*0} , and leads to a fall in the overall index, $\hat{P}_{gs}^{M,t}$. This factor is the same for both the FBW and RW procedures. The second factor in (19) captures the price changes that come from continuing varieties, and it is different for the FBW and RW procedures.

The FBW procedure applies the standard Sato-Vartia weights of log means to the variety level:

$$d_{j,gs}^t = \frac{(\lambda_{j,gs}^t - \lambda_{j,gs}^0)/(\ln \lambda_{j,gs}^t - \ln \lambda_{j,gs}^0)}{\sum_{j \in \Omega_{gs}^M} (\lambda_{j,gs}^t - \lambda_{j,gs}^0)/(\ln \lambda_{j,gs}^t - \ln \lambda_{j,gs}^0)}, \tag{20}$$

where $\lambda_{j,gs}^t$ is the import-value share of variety j within the common set, Ω_{gs}^M. Within this common set, the FBW procedure assumes, in addition, that the taste parameters of every variety remain unchanged over time; i.e. $\hat{b}_{j,gs}^t = 1$ for all $j \in \Omega_{gs}^M$. Redding and Weinstein (2020) point out that the use of the weights (20) are incompatible with the assumption of constant taste parameters, and show, instead, that their changes over time can be recovered from the residual of import-value shares conditional on observed prices, under the weaker assumption that taste changes are zero, on geometric average, within the common set. To show this point explicitly, let $\widetilde{x}$ denote the simple geometric mean across varieties in the common set for variable x. Then the assumption of zero taste changes on average can be expressed as $\widetilde{b}^t_{gs} = \widetilde{b}^0_{gs}$ for all t, and the over-time changes to variety-level taste parameters can be recovered as follows:

$$\left(\ln b^t_{j,gs} - \ln b^0_{j,gs}\right) = \ln\left[\left(\frac{\lambda^t_{j,gs}}{\widetilde{\lambda}^t_{gs}}\right)\Big/\left(\frac{\lambda^0_{j,gs}}{\widetilde{\lambda}^0_{gs}}\right)\right] - (1-\sigma_g)\ln\left[\left(\frac{p^t_{j,gs}}{\widetilde{p}^t_{gs}}\right)\Big/\left(\frac{p^0_{j,gs}}{\widetilde{p}^0_{gs}}\right)\right]$$

This expression implies the following modified Sato-Vartia weights for the RW procedure:

$$d^t_{j,gs} = \frac{\left(\lambda^t_{j,gs} - \lambda^0_{j,gs}\right)\Big/\left((\ln\lambda^t_{j,gs} - \ln\lambda^0_{j,gs}) - (\ln b^t_{j,gs} - \ln b^0_{j,gs})\right)}{\sum_{j\in I^M_g}\left(\lambda^t_{j,gs} - \lambda^0_{j,gs}\right)\Big/\left((\ln\lambda^t_{j,gs} - \ln\lambda^0_{j,gs}) - (\ln b^t_{j,gs} - \ln b^0_{j,gs})\right)}. \quad (21)$$

The price index calculated based on these modified weights, by equation (19), precisely reproduces the price index of equation (8) in Redding and Weinstein (2020) (see Appendix B for derivations).

With the sectoral import prices, (18) - (21), in hand, we can apply equation (13) to compute the sectoral prices, $\hat{P}^t_s$, and then the aggregation procedure in equations (15) and (17) to compute the prices of the tradable categories of food, F, and other merchandise, G. Because variety-level price data do not exist for services, we assume that for the non-tradable category of V, price changes are given by the service component of CPI. Summarizing our results, the OPI-M index given by the FBW procedure is

$$\hat{P}^t_F = \hat{P}^t_1, \quad \hat{P}^t_G = \prod_{s=2}^{S-1}\left(\hat{P}^t_s\right)^{\beta_s}, \quad \hat{P}^t_V = \hat{P}^t_{V,CPI}, \text{ where } \hat{P}^t_s = \left(\hat{\pi}^{M,t}_s\right)^{\frac{1}{\sigma_s-1}}\hat{P}^{M,t}_s, \\ \text{and } \hat{P}^{M,t}_{gs} \text{ is given by equations (18), (19), and (20).} \quad (22)$$

We refer to the price indices given by equation (22) as OPI-M1. The OPI-M index given by the RW procedure is

$$\hat{P}^t_F = \hat{P}^t_1, \quad \hat{P}^t_G = \prod_{s=2}^{S-1}\left(\hat{P}^t_s\right)^{\beta_s}, \quad \hat{P}^t_V = \hat{P}^t_{V,CPI}, \text{ where } \hat{P}^t_s = \left(\hat{\pi}^{M,t}_s\right)^{\frac{1}{\sigma_s-1}}\hat{P}^{M,t}_s, \\ \text{and } \hat{P}^{M,t}_{gs} \text{ is given by equations (18), (19), and (21).} \quad (23)$$

We refer to the price indices given by equation (23) as OPI-M2.

Equations (22) and (23) show that the OPI-M indices have three components: the intensive margin, or the price changes within the common set; the extensive margin, or variety entry and exit; and the import-share adjustment, $\left(\hat{\pi}^{M,t}_s\right)^{\frac{1}{\sigma_s-1}}$, which corrects for changes in the relative price of imports.[21]

Comparing OPI-M with OPI-T, we see that the OPI-M indices take advantage of detailed data on

[21] In the literature, Feenstra (1994) and Broda and Weinstein (2006) focus on import price indices, and do not include import-share adjustments. Redding and Weinstein (2018), however, include this margin in their decomposition of the US aggregate price index.

variety-level unit values and quantities of imports. However, calculation of OPI-M relies on a larger set of elasticity parameters. While OPI-M and OPI-T are identical in theory, the differences in data, parameter estimates, and assumptions imply that OPI-M provides an alternative for computing gains from trade.

So far we have derived the expressions for the category-level price indices, $\hat{P}^t_F$, $\hat{P}^t_G$, and $\hat{P}^t_V$, according to multiple specifications in commonly-used trade models. Turning to empirics, however, we note that inference based on the food Engel curve provides us with the true price index at the level of the aggregate economy, and not for individual categories. This means that we need to aggregate the category-level price indices into an economy-wide index in a manner that allows us to compare the OPIs with the conventional CPI. We turn to this task below.

4.3 Aggregation Across Categories

Figure 1 shows the structure of preferences for the product space. Preferences are non-homothetic between the groups of food, F, and non-food, N, as specified in Section 2.1, where we used the non-homotheticity to derive the food Engel curve. Within group N, expenditure shares between the categories of tradable merchandise, G, and non-tradable services, V, may vary over time. Given that we observe those expenditure shares, we can leave their exact formulation unspecified. Within the tradable categories of F and G, the sub-utility function is homothetic, as specified in Sub-sections 3.1 and 3.2, where we used the CES structure to derive the OPI-T and OPI-M indices.

Let $\beta^t_c,\ c = \{F, G, V\}$, denote the weights by category. The aggregate price index of the economy can be approximated as the weighted geometric mean across categories, with weights β^t_c; i.e. $\hat{P}^t = (\hat{P}^t_F)^{\beta^t_F}(\hat{P}^t_G)^{\beta^t_G}(\hat{P}^t_V)^{\beta^t_V}$ (e.g. see our discussion in sub-section 2.1).[22]

Specifically, we use the weights from the official CPI statistics. In other words, we aggregate the category-level prices of OPI, obtained in Sub-sections 3.1 and 3.2, according to

$$\hat{P}^t = (\hat{P}^t_F)^{\beta^t_{F,CPI}}(\hat{P}^t_G)^{\beta^t_{G,CPI}}(\hat{P}^t_V)^{\beta^t_{V,CPI}}. \tag{24}$$

The use of CPI weights in equation (24) is useful for our analyses for several reasons. First, it allows us to use CPI as the benchmark for the open-economy price indices. By using the same category weights, calculations of average bias for OPI indexes and CPI become more comparable.

Second, we will use components of CPI in the construction of OPI, to measure general inflation of

[22] Two related comments come in order. First, our specification of the price index is not meant to capture a cost-of-living index. We are interested in a nation-wide price index such as CPI that can be used as a deflator of income, whereas the cost-of-living index, under non-homothetic preferences between food and nonfood, varies across households. Second, while the theory-mandated categorical indices, $\hat{P}^t_F$, $\hat{P}^t_G$, and $\hat{P}^t_V$, measure both real consumption and welfare within categories (due to homothetic preferences within category), one needs theory-mandated consumption weights to measure welfare because of non-homothetic preferences across categories. For example, if β^t_F, β^t_G, and β^t_V are the Hicksian expenditure shares at final-time utility, $\hat{P}^t$ measures compensating variation (Baqaee and Burstein, 2021). In practice, these theory-mandated weights are typically not observable, and it is difficult to make general characterizations about them without additional assumptions of preferences. We thus choose to measure real consumption, not welfare, by using observed consumption weights for aggregation. In other words, while the theoretical predictions of the gains-from-trade literature, such as the ACR formula, apply to both welfare and real consumption, what we can take to the data is the predictions about real consumption.

domestic varieties in the OPI-T indices, and to measure the price change of the non-tradable category V in the OPI-M indices. The use of CPI weights allows us to cleanly separate the contribution of elements in OPI missing in CPI to differences in average bias.

4.4 Threats to Identification

In this sub-section, we outline potential concerns for our framework, and the steps that we have taken to address them.

Domestic Prices. A challenge we face in bringing our formulas to data is that it is very rare for researchers to directly observe the prices of domestic varieties.[23] Since we will not observe changes in domestic prices at the sector level, we will assume in our exercises below that $\hat{p}^t_{D,s}$ is constant across sectors. One candidate for measuring overall changes in domestic prices is CPI itself. If we use CPI to measure changes in domestic prices, however, we may capture movements in import prices as well. For instance, suppose prices of imported varieties decrease with no change in their quality. If these imported varieties are included in the CPI sample, then their price decreases would show up twice in our OPI-T indices, once in CPI itself, and again in the domestic-share corrections.

The way we address this issue is motivated by our focus on comparing OPIs with *statistical* measures of price indices. In this regard, we want a measure of domestic price inflation that can be thought of as the domestic component of the official CPI inflation. Specifically, we adjust the official CPI using the official import price index, $\hat{P}^t_{IMP}$, according to:

$$\hat{P}^t_{CPI,D} = \left[\left(\hat{P}^t_{CPI}\right) \Big/ \left(\hat{P}^t_{IMP}\right)^{\beta^t_M}\right]^{\frac{1}{1-\beta^t_M}}, \tag{25}$$

where β^t_M represents the economy-wide import share in year t. Our idea for equation (25) is that, if official import price indices, $\hat{P}^t_{IMP}$, measure inflation due to imports, then removing $\hat{P}^t_{IMP}$ from CPI will give us a measure of domestic inflation, $\hat{P}^t_{CPI,D}$, which we will call domestic CPI. Whether CPI or domestic CPI provides a better measure for domestic price inflation, $\hat{p}^t_D$ in equations (14) - (17), depends on whether the data of $\hat{P}^t_{IMP}$ are compatible with official CPI data, as we discuss in Section 7 below.

Final consumption vs intermediate use. Both CPI and the food Engel curves draw on household consumption data, but many goods in the trade data, which are used to compute the OPI indices, are not directly purchased by consumers (e.g. iron ore, crude oil). We address this concern in two ways. One, the index of OPI-T3 explicitly accounts for input-output linkages in its construction. Two, we use the consumption shares from WIOD (World Input Output Database) to compute the sectoral weights, β_s, when we aggregate across the non-food merchandise sectors in the category

[23] An exception is Auer et al., 2022, who identify source countries from the product labels for a subsample of the universe of products in the Swiss market.

G (e.g. equations 15 and 23). This leads to low β_s values for the sectors with a lot of intermediate goods, as we show in the next section.

Asymmetry between food and non-food products. Our framework does not treat food and the other merchandise sectors symmetrically; e.g. food is its own category and its own group (e.g. Figure 1). We do so for two reasons. One, in the household consumption data, there is a tight relationship between the share of food and log household income, known as the Engel's law, which we will illustrate using our data in the next section. Two, to the best of our knowledge; the empirical relationship between log household income and the shares of sub-categories of food, or that between log income and shares of non-food products such as clothing, is not as stable across countries and time(e.g. Banks et al., 1997, Atkin et al., 2020).

Nonlinear food Engel curves. We have so far assumed that the food Engel curve is linear in log household income, conditional on control variables, in equation (4). We will show evidence that the linear food Engel curve provides a reasonable approximation in our data in Section 5. In addition, we will also extend our framework to allow for the quadratic food Engel curve in Sections 6 and 7 below.

5 Data and Preliminary Analyses

In this section, we first briefly outline our data and parameter values, relegating the full details to the Data Appendix. We then show the salient features of our data, and conduct preliminary data analyses.

5.1 Data Sources and Parameter Values: Outline

We draw on standard publicly available data for variety-level international trade (BACI-CEPII), sector-level domestic and import shares (e.g. OECD STAN), and category-level CPI weights and prices (national statistical agencies). We have taken micro data of household consumption surveys for the United States from PSID for 1995-2015. In addition, we have obtained household-level data for four other countries in our sample, consisting of the UK, South Korea, Canada, and Peru, as listed in Table 1 below. All these data are publicly available. Table 2 lists the brief descriptions and ISIC-revision-4 codes of our 12 tradeable sectors, one for the category of food, F, and 11 for non-food merchandise, G. For each of these 11 sectors, we first obtain its consumption share within G using WIOD (the World Input-Output Database)[24]. We then multiply this relative consumption share by the CPI weight of category G to obtain β_s in (22) and (23).[25]

We draw on the literature for the values of our parameters: the goods-level substitution elasticity, σ_g in equations (19) - (21), the sectoral substitution elasticity between imported and domestic

[24]Since Peru is not covered in WIOD, we instead use for it the input-output data from GTAP.

[25]It is not straightforward to obtain β_s directly from CPI data, because different countries use different CPI sub-categories, and it is often challenging to map these sub-categories into the ISIC rev4 classification.

bundles, σ_s in equations (22) and (23), and the trade elasticity, θ_s in equations (14) - (16). For σ_g, we take the estimates from Broda and Weinstein (2006) at the level of 3-digit SITC (Standard International Trade Classification) codes. We have slightly over 200 SITC goods per country; e.g. footwear, metal containers for storage or transport, and food-processing machines.[26] We set σ_s as the mean values of $\{\sigma_g\}$ across goods g in sector s, following Imbs and Mejean (2015).[27] Finally, we set $\theta_s = \sigma_s - 1$. These values are in the range of estimates produced in the trade literature (see Data Appendix).

Table 1: Summary of Household Data for non-US Countries

Country and Year	HH Data Source	inc coef	R-sq
U.S., 1995-2015	PSID	-0.078***	0.40
U.K., 1996-2008	BHPS	-0.128***	0.50
S. Korea, 2006-2015	KLIPS	-0.094***	0.55
Canada, 2000-2009	SHS	-0.062***	0.46
Peru, 2001-2010	NHS	-0.139***	0.25

Notes: This table reports, for every country, the following information: sample years; sources of micro household consumption data; the income coefficients β_F and R^2 of the food Engel curve estimation; and mean values of CPI and the three categories of OPI-T. BHPS is the British Household Panel Survey, KLIPS is the Korean Labor and Income Panel Study, SHS is Survey of Household Spending, and NHS is National Household Survey. All β_F estimates are significant at the 1 percent level or more, and note that for a given country, all food Engel curve estimations yield the same β_F estimates and R^2, regardless of which measured price index is used to deflate household income.

Table 2: Tradeable Sectors

	ISIC Code	Description
1	01-03 & 10-12	Food & Agriculture
2	13-15	Textile and Apparel
3	16-18	Wood and Paper
4	19	Refined Petroleum
5	20-21	Chemicals
6	22	Plastics
7	23	Minerals
8	24-25	Metals
9	26-28	Machinery and Electronics
10	29-30	Transport Equipment
11	31-32	Furniture & Other Mfg.
12	05-08	Mining

Notes: This table shows our food and non-food tradable sectors and their corresponding ISIC rev. 4 codes.

[26] One may be concerned about the large number of parameters involved. In the Data Appendix, we show that we obtain similar results for OPI-M if we apply σ_s to all the goods within sector s. Note that the OPI-T indices do not require the parameters σ_g.

[27] The mean values produce the substitution elasticity among imported goods, which we denoted by σ_s^M in Section 4. We do not require the σ_s^M values in our computation, as shown in equation (18), and have set $\sigma_s^M = \sigma_s$. (Feenstra et al., 2018) show that σ_s is statistically indistinguishable from σ_s^M for two thirds of the goods they examine, and that $\sigma_s^M > \sigma_s$ for the remaining one third.

5.2 Noteworthy Data Patterns

In this sub-section, we present two noteworthy patterns in our data. We start by illustrating how the WIOD-based consumption shares of the non-food merchandise sectors, β_s, within the category G, help address the concern that many imported goods are not purchased directly by consumers. To do so, we use STAN data to compute the following alternative consumption shares. We first obtain the sectoral relative consumption shares within category G using total absorption (gross output minus exports plus imports) in STAN, and then multiply these relative shares by the CPI weight of category G. While the WIOD- and STAN-based consumption shares differ for individual non-food merchandise sectors, they have the same cross-sector aggregate within G, by construction.

Figure 2 plots β_s, based on WIOD data, against the STAN-based consumption shares by year by sector, for the U.S. (the results for the other countries are very similar). Mining has sizable apparent consumption and its STAN-based consumption shares range from roughly 2% to 4%. However, most of its goods are intermediate goods, and so its WIOD-based consumption shares are close to 0. We see a similar pattern for the sectors of Wood and Paper and Metals. This implies that the sectors for which intermediate goods are less important, such as Textile and Apparel and Transport Equipment, have larger WIOD-based consumption shares than STAN-based ones. Overall, Figure 2 illustrates that WIOD data assign low consumption shares to the sectors where intermediate goods are important, and so the use of WIOD-based consumption shares in our framework helps alleviate the concern that many imported goods are not directly purchased by consumers.

We now illustrate the tight empirical relationship between log household income and the share of food in household consumption in our data. We pick one year of household consumption data for each country, often times the middle year of the sample period, and divide the households into decile bins of income. We then compute the mean values of food share and log income within each decile bin, netting out household characteristics. Figure 3 plots the mean food share against the mean log income by country. We see that food share decreases with log income for every country, and that the relationship between food share and log income is close to a linear relationship. Figure 3 confirms Engel's law in our data.

5.3 Preliminary Analyses

Having shown that the food Engel curve is a strong empirical regularity in the previous sub-section, we perform the following two validation exercises for our approach in using food Engel curves to compare price indices. In the first validation exercise, we use the constant price index in all years in the food Engel equation (4). Because most countries have positive inflation in most years, we expect that by using the constant price index as the deflator, we will overstate household real income. This implies that our year-dummy estimates under constant price index are likely to be mostly positive.

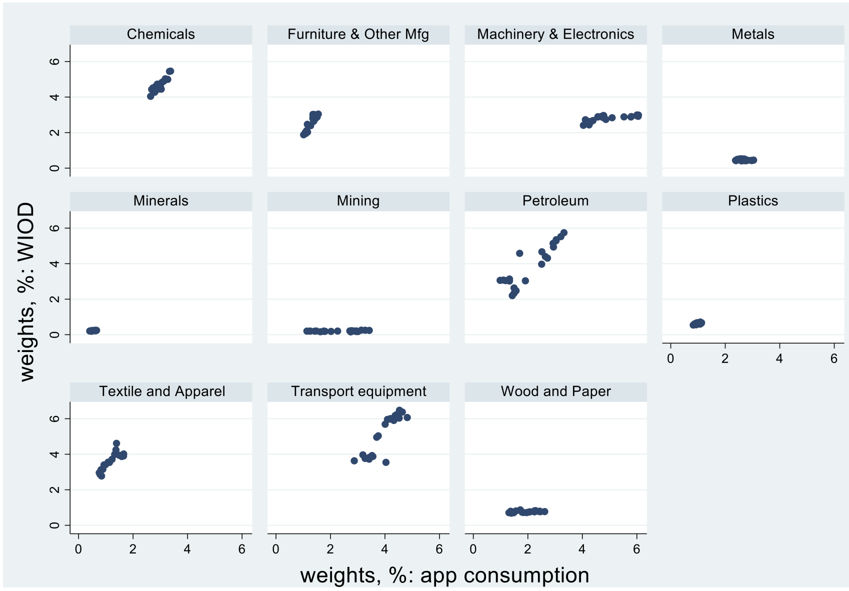
To implement this exercise, we use household nominal income in the estimation of equation (4), and include the full set of control variables. Column (1) of Table 3 reports the results for the U.S. We see that the log-income coefficient, β_F in equation (4), is negative and significant, and that the coefficients of household head's age and number of children are positive and significant.

Table 3: Estimates of US Food Engel Curve

VARIABLES	(1) Con. Prices	(2) CPI	(3) Con. Prices	(4) CPI
ln income	-0.0783***	-0.0783***	-0.604***	-0.604***
	(0.00125)	(0.00125)	(0.0241)	(0.0241)
$(\ln \text{income})^2$			0.0279***	0.0279***
			(0.00124)	(0.00124)
age head	0.000398***	0.000398***	0.000388***	0.000388***
	(9.50e-05)	(9.50e-05)	(9.24e-05)	(9.24e-05)
age spouse	-0.000188*	-0.000188*	-0.000320***	-0.000320***
	(9.62e-05)	(9.62e-05)	(9.42e-05)	(9.42e-05)
number of children	0.0145***	0.0145***	0.0140***	0.0140***
	(0.000405)	(0.000405)	(0.000391)	(0.000391)
hours head	0.00155***	0.00155***	0.00181***	0.00181***
	(0.000353)	(0.000353)	(0.000336)	(0.000336)
hours spouse	-0.00295***	-0.00295***	-0.00173***	-0.00173***
	(0.000319)	(0.000319)	(0.000311)	(0.000311)
edu head	-1.84e-05	-1.84e-05	0.000218	0.000218
	(0.000168)	(0.000168)	(0.000163)	(0.000163)
edu spouse	8.10e-05	8.10e-05	3.69e-05	3.69e-05
	(0.000142)	(0.000142)	(0.000138)	(0.000138)
$\ln(P_f/P_n)$	0.0124	0.0124	0.0459	0.0459
	(0.0482)	(0.0482)	(0.0402)	(0.0402)
1996	-6.68e-06	-0.00227	-0.000258	-0.00252
	(0.00192)	(0.00192)	(0.00206)	(0.00206)
1997	1.97e-05	-0.00404**	0.000281	-0.00378*
	(0.00199)	(0.00198)	(0.00215)	(0.00215)
1999	0.000316	-0.00665***	-1.08e-05	-0.00697***
	(0.00203)	(0.00202)	(0.00217)	(0.00217)
2001	0.00185	-0.00987***	0.000623	-0.0111***
	(0.00186)	(0.00184)	(0.00213)	(0.00213)
2003	0.00771***	-0.00715***	0.00910***	-0.00576***
	(0.00203)	(0.00200)	(0.00217)	(0.00217)
2005	0.0151***	-0.00424**	0.0155***	-0.00387*
	(0.00204)	(0.00200)	(0.00218)	(0.00218)
2007	0.0164***	-0.00751***	0.0178***	-0.00615***
	(0.00210)	(0.00202)	(0.00235)	(0.00235)
2009	0.0185***	-0.00847***	0.0180***	-0.00895***
	(0.00326)	(0.00319)	(0.00321)	(0.00321)
2011	0.0258***	-0.00491*	0.0257***	-0.00499*
	(0.00284)	(0.00277)	(0.00268)	(0.00268)
2013	0.0288***	-0.00475*	0.0292***	-0.00436
	(0.00290)	(0.00282)	(0.00271)	(0.00271)
2015	0.0307***	-0.00485	0.0310***	-0.00457
	(0.00381)	(0.00372)	(0.00354)	(0.00354)
Region FE	Yes	Yes	Yes	Yes
Observations	23,921	23,921	23,921	23,921
R-squared	0.403	0.403	0.444	0.444

Notes: This table shows regression results, based on constant price index and CPI for food Engel curves in the sample of US data. We omit the constant term and region dummies to save space. 1995 is the benchmark year. In columns (1) and (3), we deflate household income using constant prices (i.e., we use nominal income), and in columns (2) and (4), we deflate it using U.S. CPI. Columns (1) and (2) report results for linear food Engel equation (4), and columns (3) and (4) do so for quadratic food Engel equation (26).

Figure 2: Sectoral Consumption Shares, WIOD Data vs. STAN Data, U.S.



Notes: This figure shows the weights of each of the eleven non-food tradeable sectors.

These patterns are consistent with the literature. We have also run the estimation for the other countries in our data, and gather the estimates of β_F and the R^2 of the regressions in Table 1. The magnitudes of the β_F estimates are all negative—in the range of -0.06 and -0.14—and statistically significant, reflecting that food is a necessity. The statistical significance of the β_F estimates and the high R^2 show, again, that food Engel curve is an important empirical regularity in household survey data.[28]

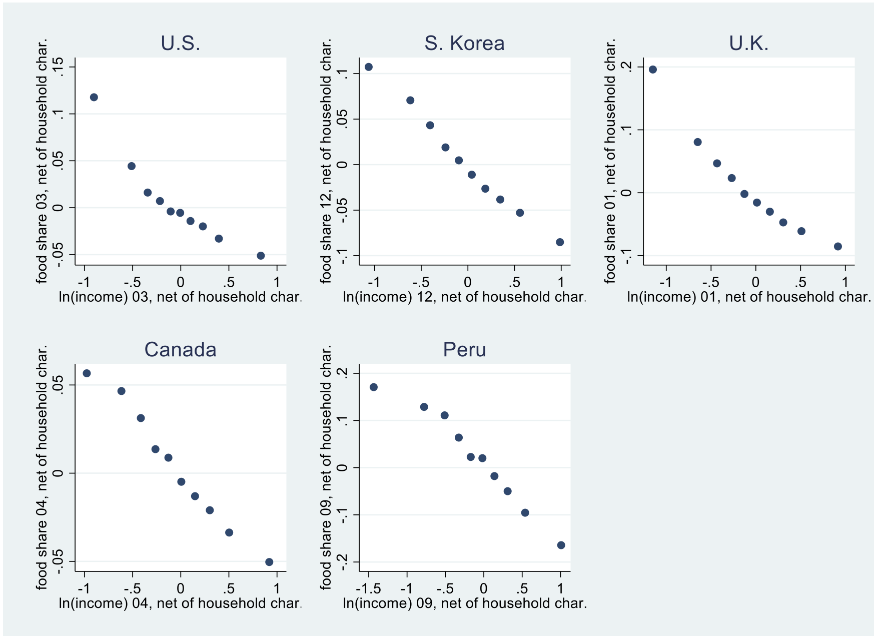
Meanwhile, we also see, from Table 3, that the estimates of the year dummies are all positive in sign for the U.S., and they are statistically significant starting 2003. We plot the year dummy estimates and their 95 percent confidence intervals, by country, in Figure 4. Although these estimates come from multiple countries, different time periods, and are based on different data sources, they are all largely positive, and often times statistically significant. This provides evidence that the food Engel curve correctly identifies the downward bias of the naive index of constant prices.

For our second validation exercise, we examine the estimates of the bias of the official U.S. CPI that our framework produces. Ever since the Boskin report (Boskin et al., 1997), there have been extensive studies of the U.S. CPI, and this literature has reached the consensus that the U.S. CPI is upward-biased. Here, we are interested to compare the bias estimates of U.S. CPI from our analysis that is based on inference from food Engel curves, with the estimates from this literature that are based on different approaches.

To carry out this exercise, we go back to the U.S. data, replace the index of constant prices with official CPI in regression (4), and report the results in column (2) of Table 3. As compared with the results under constant prices, in column (1), the only difference is the year-dummy estimates. The

[28]Previous studies listed in Footnote 2 that have estimated the food Engel curve for different time periods and different countries report their β_F estimates to be between -0.2 and -0.05.

Figure 3: Food Expenditure Share against Log Income



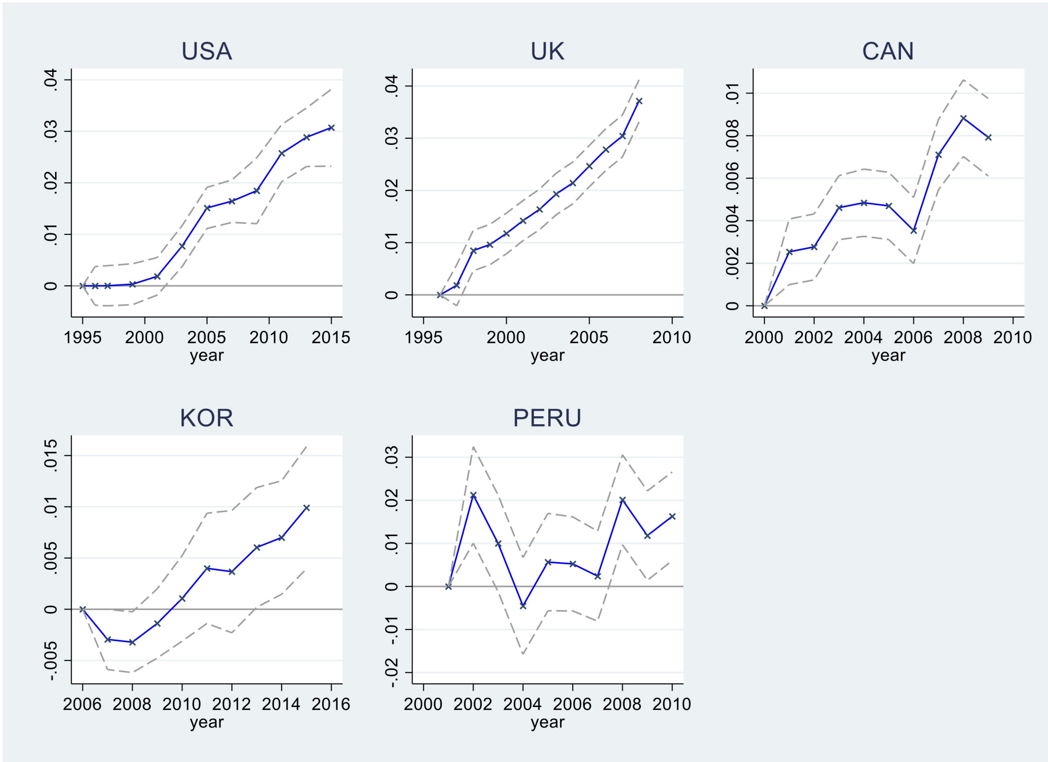
Notes: This figure shows, for each country in a given year, the food expenditure share against log of income along the ten bins of income distribution. The graph of a given country is based on a single year of data, which is described in the text of the x-axis label; e.g. it is 2003 for the U.S. and 2012 for S. Korea.

other coefficient estimates, and their standard errors, and the R^2 of the regression, are all identical, consistent with equation (5). On the other hand, for the U.S., the year-dummy estimates under CPI are negative, and often times statistically significant, and we obtain an average bias of -0.0751, using equation (6). In other words, our estimates indicate that the U.S. CPI produces an upward bias of 0.75 log points on average per year. In comparison, Shapiro and Wilcox (1996) show that there is an 80% probability that the CPI bias lies between 0.6% and 1.5% per year, Gordon (2006) reports an annual bias of 0.8%, and Berndt (2006) reports 0.73%-0.9%. The similarity between our results and the literature provides additional evidence that the food Engel curve is a useful tool for estimating the biases of measured price indices.

6 Results for the U.S.

In this section, we report our results for the United States. We start our analysis with the U.S. for several reasons. First, the availability of data for the U.S. is better than other countries. e.g. the U.S. household consumption data, from PSID, are available for a significantly longer time span on a consistent basis. Second, a large literature has examined the U.S. CPI in detail, to which we can relate and benchmark our results. Finally, it is easier to compare the OPI-M indices with the official CPI if the same currency is used for both import invoicing and domestic transactions, as is often the case for the U.S. We present our results for the other countries in the next section.

Figure 4: Estimates of Year Dummy Coefficients under Constant Price Index, by Country



Notes: This figure shows the estimates of year dummy coefficients from the food Engel estimation under constant price index for each of the countries in our sample.

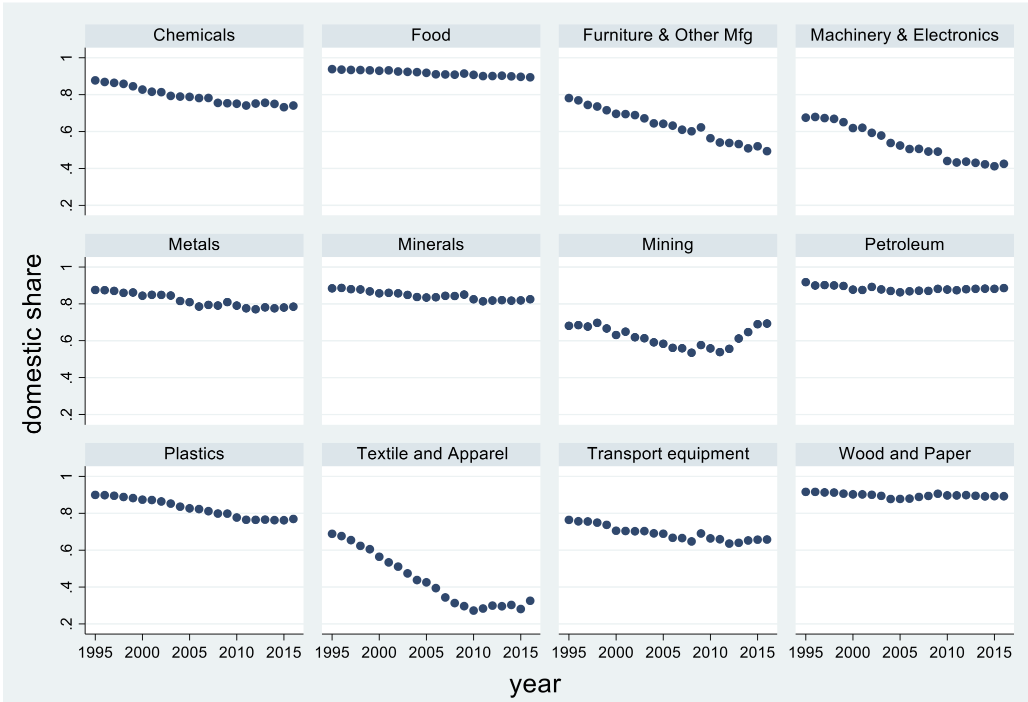
6.1 Results for OPI-T

We start by reporting the patterns of sector-level domestic expenditure shares, which play an important role in the construction of the OPI indices. Figure 5 plots the U.S. domestic expenditure share by sector by year, where the sectors are in the tradable categories of food and non-food merchandise. A salient feature is that US domestic expenditure shares over the period of 1995-2015 tended to decrease or stay unchanged, and they rarely increased. The decrease in domestic expenditure share has been particularly notable in Textile and Apparel, Chemicals, Plastics, Machinery and Electronics, and Furniture and Other Manufacturing. These decreases are likely driven by the global wave of trade liberalizations in the 1990's and early 2000's, such as China's rise starting in the early 1990s (e.g. Autor et al. 2013) and its subsequent WTO accession in 2001, the signing of NAFTA in 1994, and the expiration of the Agreement on Textiles and Clothing quotas in 2005, among other considerations.[29]

Figure 6 plots the U.S. import price index and the U.S. CPI. The comparison of these two indices is reasonable, because their construction is similar; e.g. both indices collect data through surveys of

[29] Additional examples of trade liberalizations involving the U.S. are as follows. The tariff cuts under the U.S.-Canada Free Trade Agreement were not complete until the end of 1998 (e.g. Lileeva and Trefler 2010), and the U.S. substantially lowered tariffs on Vietnamese products after the U.S.-Vietnam Bilateral Trade Agreement in 2001 (e.g. McCaig 2011).

Figure 5: Domestic Share by Sector, U.S.



Notes: This figure shows the US domestic expenditure share for each industry between 1995 and 2015. The domestic share is the ratio of gross output minus exports to total absorbtion (which is gross output minus exports plus imports), the data of which come from STAN.

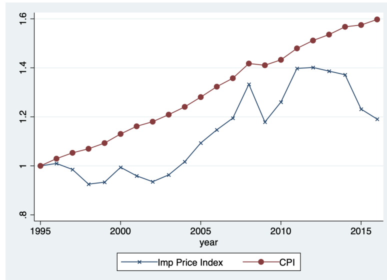
establishments, and both are Laspeyres.[30] This figure shows that the U.S. import price index has smaller increases over time than CPI during our sample period. This result is intuitive, given the overall increase in U.S. trade openness that we saw in Figure 5. This result also implies that, when we construct the U.S. domestic CPI, $\hat{P}^t_{CPID}$, using equation (25), we obtain an index that registers larger increases over time than CPI.

We next construct the OPI-T indices for the U.S. using domestic CPI. Relative to CPI, the OPI-T indices tend to have lower mean values in our sample.[31] This result suggests that, relative to CPI, the core logic of the gains from trade literature implies lower over-time changes in the prices that consumers face, because the overall decline in domestic shares over time indicates better access to international markets. In addition, OPI-T3, with input-output linkages and multiple sectors, has a lower mean value than OPI-T2 which incorporates only multiple sectors. Both OPI-T3 and OPI-T2 have lower mean values than OPI-T1, which has a single sector for non-food merchandise.

[30] See, e.g., the Handbook of Methods by the U.S. BLS (Bureau of Labor Statistics). The official import price index measures the prices paid by importing firms (i.e. border prices) but CPI measures those paid by consumers (i.e. retail prices). We implicitly assume that the border-to-retail-price pass through is 1 in the long-run (more than one year), as in many studies in the gains-from-trade literature. Using the Nielsen home scanner data for Switzerland, Auer et al. (2022) estimate this pass through, for identical goods, to be 0.57 at 2 quarters and 0.95 at 3 quarters (their IV estimates).

[31] The mean values of CPI and OPI-T1 through -T3 are, respectively, 1.26, 1.26, 1.24 and 1.21. The mean of domestic CPI is 1.27. All these indices are normalized to 1 for the base year of 1995.

Figure 6: The Import Price Index and CPI of the United States, 1995-2015



Notes: This figure shows the import price index based on national statistics as well as CPI for the United States between 1995 and 2015.

These results are consistent with the typical findings in the gains from trade literature (e.g. CRC), based on comparative statics exercises, that both the model elements of input-output linkages and multiple sectors tend to produce larger gains-from-trade predictions.[32]

6.1.1 OPI-T vs. CPI

We now use the three OPI-T indices to deflate household income in the food Engel curve, equation (4), and obtain their average biases using equation (6). We plot these average biases and their 95 percent confidence intervals in the upper left graph of Figure 7, where we have included the average biases of CPI for comparison. Figure 7 shows that the average bias is negative for all the OPI-T indices and CPI; i.e. they all tend to overstate the price increases that households experience, as revealed by their consumption decisions. The average bias is statistically significant for CPI and OPI-T1 and -T2, but insignificant for OPI-T3.

As shown in Section 4, each of the three OPI-T indices consists of two components: domestic CPI, which measures domestic inflation, and the correction implied by the gains-from-trade literature. The correction terms for all the three OPI-T indices involve over-time changes in domestic shares, as shown in Equations (14), (15), and (16).

Applying the decomposition of Equation (9), we see that the changes in average bias from CPI to an OPI-T index arises for two reasons. First, the index of domestic CPI used in OPI-T measures domestic inflation, while CPI measures overall inflation, due to both domestic goods and imports. This shows up as the first term on the left-hand-side of equation (9), and we report it in column (2) of Table 4-Panel (a). The use of domestic CPI mechanically increases the average bias relative to CPI, because by construction, domestic CPI leaves out the inflation due to imports. We have taken this step in order not to double count the contribution of the changes in import prices (holding variety and quality fixed), as we discussed in sub-section 4.4 above. Because the same domestic CPI is used for all three OPI-T indices, the value of column (2) is also the same for them.

On the other hand, OPI-T indices incorporate the domestic-share corrections as predicted by the gains-from-trade literature, which are designed to capture the new variety and quality changes

[32] Giri et al. (2021) show, however, that the theoretical predictions of gains from trade could be also larger for single sector models than for multi-sector models.

Table 4: Relative Average Bias for OPI-T Indices, U.S.

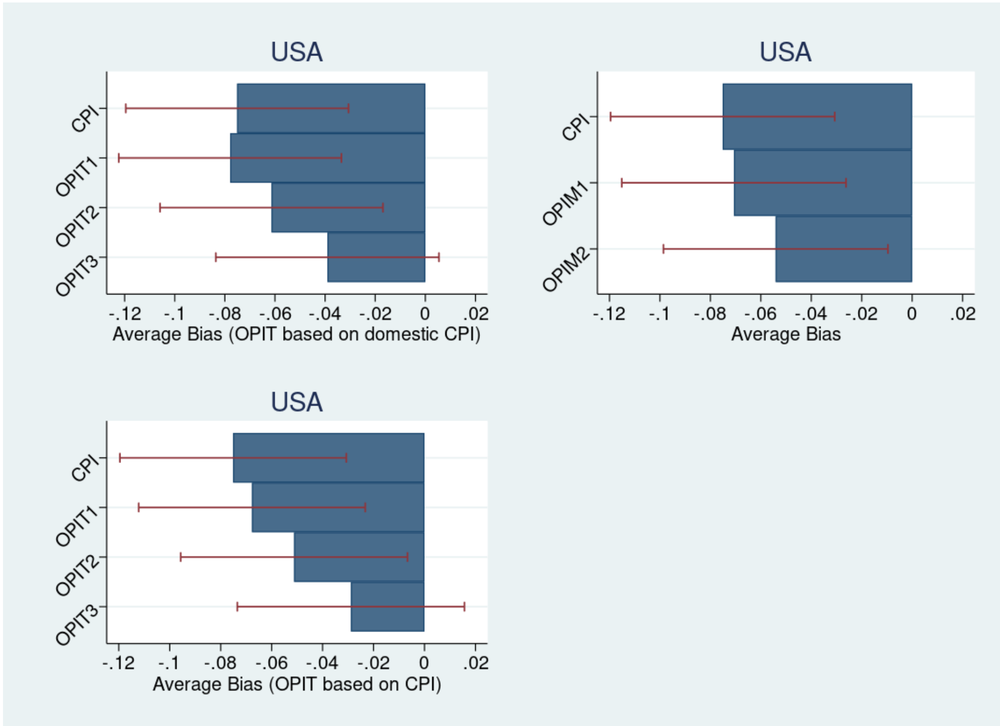
Panel a. OPI-T with domestic CPI	(1) Relative Average Bias	(2) Domestic CPI vs. CPI, % CPI Ave. Bias	(3) Gains-from-Trade Correction, % CPI Ave. Bias
OPI-T1	103.7%	-13.6%	9.9%
OPI-T2	81.7%	-13.6%	31.9%
OPI-T3	52.0%	-13.6%	61.6%
Average	79.2%	-13.6%	34.4%
Panel b. OPI-T with CPI	Relative Average Bias	% CPI Bias Explained by OPI-T	
OPI-T1	90.1%	9.9%	
OPI-T2	68.1%	31.9%	
OPI-T3	38.5%	61.5%	
Average	65.6%	34.4%	
Panel c. OPI-M	Relative Average Bias		
OPI-M1	94.1%		
OPI-M2	72.0%		
Average	83.1%		

Notes: This table reports the average bias for each OPI index relative to CPI. Panel (a) reports the result for OPI-T indices that are calculated using domestic CPI. Panel (b) is the same as Panel (a) except that in its collocation the over CPI is used. Panel (c) repots the results for OPI-M indices. Across all the panels, Column (1) shows the relative average bias of OPI indices, as computed using equation (7).

stemming from imports as the U.S. becomes more open to trade. In comparison, CPI does not fully adjust for variety and quality (more on this below). This is the second term on the left-hand-side of equation (9), and we report it in column (3) of Table 4-Panel (a). Intuitively, the average bias of CPI is like a residual, and quantifies the over-time changes in true household prices that CPI is unable to explain. We see that, for example, relative to the average bias of the U.S. CPI, the contribution of the domestic-share correction in U.S. OPI-T2 is 31.9%; i.e. the gains-from-trade prediction embedded in U.S. OPI-T2 can explain 31.9% of the movements in true household prices left unexplained by official U.S. CPI. Table 4 shows that in most cases, the domestic share correction, exactly as specified in the gains-from-trade literature, accounts for a sizable portion of the over-time changes in true household prices that are unexplained by official CPI statistics.

Putting these two components together, we report the relative average bias, computed using equation (7), in column (1) of Table 4-Panel (a). Because the relative average bias is equal to the third term on the left-hand side of equation (9) for all three OPI-T indices, columns (1)-(3) sum up to 100%. As we move from CPI to OPI-T1, the effect of removing import price changes is roughly offset by the addition of gains-from-trade correction. Overall, the average bias increases in magnitude slightly. On the other hand, for OPI-T2 and -T3, the effect of adding the gains-from-trade corrections dominates, and so the average bias becomes smaller in magnitude relative to CPI (see also Figure 7). For example, the relative average bias of OPI-T2 is 81.7%. This implies that the

Figure 7: Average Bias of US CPI and OPI Indices



Notes: This figure shows average bias of OPI indices and of CPI: Top-left panel for the three OPI-T indices based on domestic CPI, bottom-left panel for the three OPI-T indices based on overall CPI, and the top-right panel for the two OPI-M indices. The red lines are 95 percent confidence intervals.

use of the OPI-T2 index reduces the magnitude of CPI's average bias by 18.3%. Averaging across the three OPI-T indices, we obtain a mean relative average bias of 79.2%, implying that the use of OPI-T reduces CPI's average bias by 20.8%.

An alternative approach is to use CPI as the measure of domestic inflation to construct OPI-T. The potential caveat is that the price changes of imports (holding variety and quality fixed) might be double counted. On the other hand, this approach gives us a cleaner decomposition of the difference between CPI and OPI-T. Applying the decomposition of Equation (9), we see that the first term on the left-hand side of Equation (9) is exactly zero, because the first component of OPI-T is CPI. It follows that the change in average bias, from CPI to an OPI-T index, is completely driven by the domestic-share correction. Our results for OPI-T with overall CPI are reported in Table 4-Panel (b). The domestic-share correction in the U.S. closes the gap between CPI and the true price index by 9.9% for OPI-T1, by 31.9% for OPI-T2, and by 61.5% for OPI-T3. We plot the average biases and their 95 percent confidence intervals for CPI and OPI-T in the bottom left graph of Figure 7. While the average bias is statistically significant for CPI and OPI-T1 and -T2, it is statistically insignificant for OPI-T3. These results are very similar to those of our previous approach of using domestic CPI.

6.1.2 Relating OPI-T to the CPI Literature

What drives these differences between CPI and OPI-T? To answer this question, it is useful to discuss how CPI is constructed in the United States. The U.S. constructs its CPI using a two-tier structure. The upper tier consists of about 270 ELI's (Entry Level Items), and the lower tier consists of individual items within ELI. The prices of individual items are collected by Bureau of Labor Statistics employees at retail outlets, and quantity data are available at the ELI-level, but not at the individual-item-level (e.g. Klenow and Kryvtsov, 2008). Therefore, item-level prices are aggregated into ELI-level via simple geometric mean, and then across ELI's via weighted geometric mean (e.g. Nakamura and Steinsson, 2008).

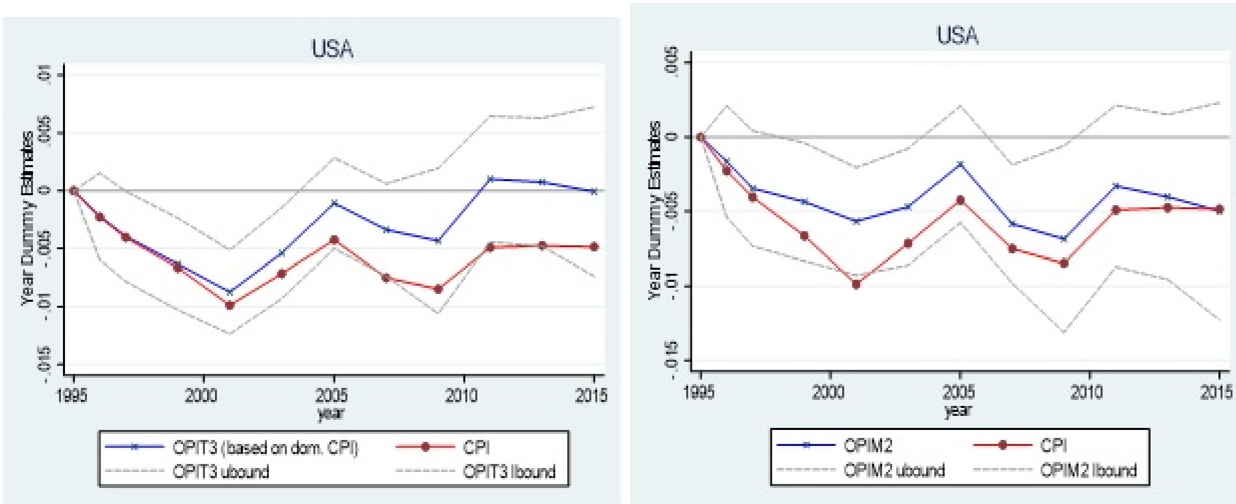
The resulting CPI index may deviate from the true prices that households use in their consumption decisions for the following reasons. First, an individual item may be replaced by a higher-quality item. This is not always easy to spot, and even if spotted, could be difficult to correctly adjust (e.g. Moulton, 1996). Second, new goods and new varieties may fail to show up in the CPI sample, and even when they do, the CPI procedure may not fully capture their effects on true household prices (e.g. Hausman, 2003). These two issues are widely recognized in the studies of U.S. CPI, commonly referred to as the quality bias and new-goods bias in that literature.

Using intuition from our OPI sketch in 3.1 shows how the gains-from-trade predictions embedded in the OPI-T indices could capture the effects of new varieties[33] and quality improvements if they originate in foreign countries. Consider one of the tradable sectors in the categories of food, F, or tradable merchandise, G. Products in this sector are differentiated into a number of varieties, somesourced from a foreign country.

Suppose, first, that some foreign country experiences a trade liberalization, and starts to export new varieties, and that the prices of existing varieties remain unchanged. In this scenario, consumers gain from trade, and the true price index they face decreases. The ability to choose among a wider set of products increases consumer welfare, but the welfare gains may not be reflected in CPI. While CPI may fail to capture all new foreign varieties, they are naturally captured by OPI-T, because domestic share decreases as consumption shifts towards new foreign varieties.Suppose, alternatively, that some foreign country increases the quality of its varieties, but the prices remain unchanged per physical unit. The prices of existing varieties, again, remain unchanged. In this scenario, consumers again gain from trade and experience lower prices, adjusted for quality. While CPI may fail to capture this margin of gains, OPI-T captures it, because domestic share decreases as consumption shifts towards higher-quality foreign varieties. Together, the fact that OPI-T is designed to capture both new goods and quality improvements originating from abroad – potentially a large source of U.S. consumer welfare gains in recent years – explains why OPI-T better tracks U.S. households' true price index than CPI.

[33] Our terminology of "new variety" is in the context of trade models, and corresponds to both "new good" and "new variety" in the CPI literature.

Figure 8: Estimates of Year Dummy Coefficients for US OPI Indices



(a) OPI-T3 (based on domestic CPI)

(b) OPI-M2

Notes: This figure shows estimates of year dummy coefficients coming from food Engel curves: the left panel for OPI-T3 based on domestic component of CPI, and the right panel for OPI-M2.

6.1.3 Comparison Among OPI-T Indices

Having compared OPI-T with CPI, we now compare the three OPI-T indices. Because these indices share the same measure of domestic inflation, our comparison yields the same ranking whether the measure is domestic CPI or CPI.

Column (3) of Table 4-Panel (a) (and column (2) of Panel (b)) shows that the gains-from-trade prediction embedded in OPI-T1 moves the index 9.9% closer to the true household prices than CPI. While this movement is in the right direction, it is small in magnitude. One reason is that foreign shocks could be heterogeneous across the sectors in the category of tradable merchandise, G. For example, we see, from Figure 5, that the movements in domestic shares are much larger for Textile and Apparel, than for Food. The underlying model of OPI-T1 treats the entire category of G as a single entity, and so does not recognize this heterogeneity. In addition, even if all sectors were experiencing the same decline in their corresponding domestic expenditure shares, the trade elasticity, θ_s, could be heterogenous across sectors. An average trade elasticity may create an aggregation bias compared to the case where one allows for sector-specific trade elasticities (Ossa, 2015).

The underlying model of OPI-T2 accommodates these sources of sectoral heterogeneity. Intuitively, heterogeneity matters because disaggregated sector-specific foreign shocks may have a sizable impact on the sector-level prices. This impact is then passed on to the category level via sectoral weights.[34] We see that the overall index of OPI-T2 closes the gap between CPI and true household prices by 18.3%. In other words, the addition of the multi-sector specification into standard trade

[34]This same intuition also explains some of the difference between multi-sector and single-sector trade models in comparative-static counter-factual exercises (e.g. CRC).

models leads to more realistic gains-from-trade predictions, because it moves the OPI-T2 index closer to the true household prices used in household consumption decisions.

Figure 5 also shows that foreign shocks are likely important for sectors that are intensive in the use of imported intermediate products, such as Machinery and Electronics. While input-output linkages are absent from the underlying model of OPI-T2, they are a central feature of OPI-T3. As we discussed in Sub-section 4.1, input-output linkages can amplify the direct effect of trade shocks in a sector, lowering the price there, because all sectors in the economy use the outputs of that sector as inputs. Table 4 reports that, the addition of input-output linkages leads to even more realistic predictions for gains from trade, because the index of OPI-T3 closes the gap between CPI and true household prices by an additional 29.7% (48.0% − 18.3%).

Summarizing our results in this sub-section, the average bias is smaller in magnitude under the OPI-T2 and OPI-T3 indices than under CPI. Because the only difference between CPI and OPI-T is the method by which foreign goods are incorporated, our results provide evidence that when households deflate their income for making consumption decisions, they behave as if they take into account gains from trade, as hypothesized and formulated in the trade literature. In addition, the average bias becomes smaller in magnitude as we progress from OPI-T1 through OPI-T3, suggesting that the additional model elements of multiple sectors and input-output linkages deliver gains-from-trade predictions that better track the true price index that households use in their consumption choices.

The left sub-graph of Figure 8 plots the year-dummy estimates under OPI-T3 and CPI, as well as the 95 percent confidence intervals of the OPI-T3 year dummies. We see that the estimates are not statistically different from zero for many years under OPI-T3 (recall that the average bias itself is also statistically insignificant under OPI-T3, as we discussed previously).

6.2 Results for OPI-M

The OPI-M indices incorporate gains from trade into over-time price changes by taking advantage of rich product-level trade data. In practice, the variety-level prices computed from such data are typically noisy (e.g. General Accounting Office (1995)), and our data is no exception. For example, our U.S. imports data cover 5,354 unique HS6 products exported by 218 countries, and over 89,000 varieties (export country by HS6) are present in both 1996 and the base year of 1995. Among these varieties, the absolute values of log price changes, over 1995-1996, have the mean of 0.76 and the standard deviation of 1.05. Following common practice in previous studies (e.g. Redding and Weinstein (2020)), we exclude the varieties with large price changes from the computation of the average price changes within the common set in the intensive margin, and treat them as variety entry and exit in the extensive margin instead. Our idea is that, if the exclusion removes the varieties with noisy price changes, then the remaining varieties should show similar aggregate moments in their price changes as compared with high-quality price data. Using one such data set, the raw data used in the construction of U.S. CPI, Klenow and Kryvtsov (2008) show that the mean of 8-month changes in log prices in absolute value is 0.11. Through experimentation, we find that if

Table 5: Percentage of US CPI Bias Explained by OPI-M Indices and their Components

	Full Index		No Variety Entry/Exit		No Imp. Share Adj.	
	Explained (1)	Unexplained (2)	Explained (3)	Unexplained (4)	Explained (5)	Unexplained (6)
OPI-M1	5.9%	94.1%	-62.1%	162.13%	84.8%	15.2%
OPI-M2	28.0%	72.0%	-40.1%	140.09%	106.8%	-6.8%

Notes: This table reports the percentage of CPI bias that is explained by each OPI-M index (column 1), and that if the following components were missing from OPI-M: variety entry/exit (column 3), or import-share adjustment (column 5). Columns 2, 4 and 6 are 100% minus, respectively, columns 1, 3 and 5.

the exclusion cutoff is set to the median changes in log prices in absolute value, relative to the base year, then the mean value of one-year log price changes (absolute values) in the common set in our data is 0.13, close to the value of 0.11.[35]

OPIM1 has a mean value of 1.29 and OPIM2 has a mean value of 1.26, both lower than CPI, though OPIM1 only marginally. We plot the average biases of OPI-M and their 95 percent confidence intervals in the upper right graph of Figure 7, and have included the average bias of CPI for comparison. The average biases of the OPI-M indices are negative and statistically significant, like that of CPI, but they have smaller magnitudes than CPI. We compute the relative average biases by OPI-M index and report them in Table 4-Panel (c). When we average across the two OPI-M indices, we obtain that the use of OPI-M reduces CPI's average bias by 16.9%. These results show that the over-time price changes computed from product-level trade data, OPI-M, tend to track true household prices better than official CPI statistics. Because OPI-M incorporates gains from trade, our results provide additional evidence that households behave as if they take such gains into account when they make consumption decisions, corroborating our earlier results for OPI-T.

While smaller than CPI, the average biases of both OPI-M indices are statistically significant. We believe that this result ought to be interpreted in the context that the CPI is built upon much higher-quality raw data than the OPI-M indices; e.g. the individual items in the CPI raw data are often times at the barcodes level. Indeed, Broda and Weinstein (2010) show that for the category of food, the US CPI is very similar to the Sato-Vartia index (which corresponds to the intensive margin of OPI-M1) based on the Nielsen home scanner data. Viewed from this angle, it is quite remarkable that the OPI-M indices deliver smaller average biases (in magnitude) than CPI.

We now clarify what is driving the differences between the average biases of OPI-M and CPI. We start by implementing the decomposition in Equation (8). Column (1) of Table 5 shows the first term of the decomposition, the mean of the log difference between OPI-M and CPI divided by the average bias of the CPI. Column (2) reports the second term of the decomposition, the ratio of the average bias of OPI-M to that of CPI. By construction, the numbers in columns (1) and (2) sum up to 100%. We see that, for example, OPI-M2 is able to account for 28.0% of the over-time changes in true household prices that official CPI is unable to explain, leaving a remainder of 72.0%.

[35] This statistic is also close to 0.11 for the other countries in our data; e.g. it is 0.12 for S. Korea, and 0.11 for Canada. We have experimented with other exclusion cutoffs and obtained larger mean values. e.g. it is 0.23 for the U.S. if we use the 75th percentile as the cutoff, and 0.34 if we use the 90th percentile.

The OPI-M indices consist of three components, as shown in Sub-section 4.2. The intensive margin computes the average changes in variety-level import prices within the common set, the extensive margin captures the entry of new varieties and exit of old ones, and these two components form the index of import prices. The third component, import-share adjustment, then translates the import-price index into the full index, by correcting the potential changes in the relative price of imports.

We first show the contribution of the extensive-margin component, by dropping it from OPI-M, and then performing the decomposition, (8), for this hypothetical index without variety entry and exit. We report the first and second terms of the decomposition in columns (3) and (4) of Table 5. Comparing columns (1) and (3), we see that the removal of variety entry and exit from OPI-M takes the overall index substantially farther away from the true price index; e.g. the movement is 68.1% $(5.9 - (-62.1))$ of the CPI's average bias for OPI-M1. These findings are consistent with the previous studies that emphasize new foreign varieties as an important contributor to gains from trade (e.g. Feenstra (1994); Broda and Weinstein (2006)).

We next examine the contribution of the import-share-adjustment component by dropping it from OPI-M. The resulting index is for import prices only, and we report its decomposition relative to CPI, based on equation (8), in columns (5) and (6) of Table 5. Column (5) is substantially larger than Column (1); e.g. without the import-share adjustment, OPI-M1 would explain 84.8% of the CPI's average bias for the U.S. during 1995-2015. These results suggest that the import price index is overly optimistic about gains from trade. To see the intuition, recall that domestic expenditure shares for the US sample generally decrease over time (Figure 5), and so, imported expenditure shares generally increase. This implies that the relative price of imports decreases over time, and so the import price index is lower than the full index of OPI-M. Our results thus provide empirical evidence that looking at the import price index alone may exaggerate gains from trade.

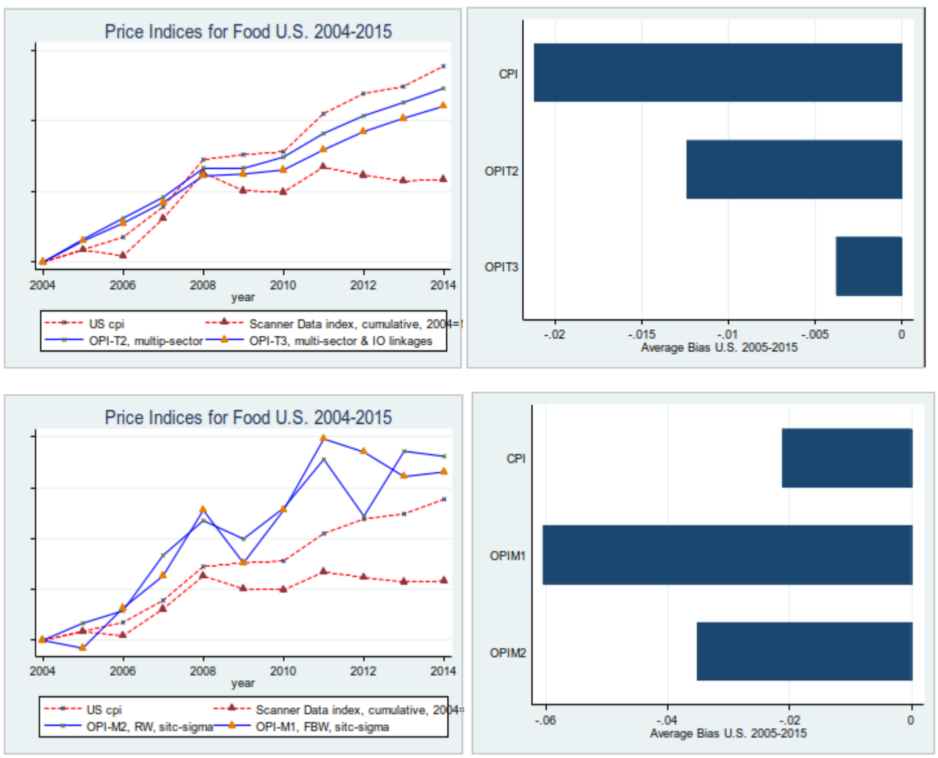
Moving on, we now compare OPI-M1 with OPI-M2. Figure 4 shows that the average bias of OPI-M2 is smaller in magnitude than OPI-M1, and column (1) of Table 5 shows that OPI-M2 has a smaller relative average bias than OPI-M1. Recall, from Sub-section 4.2, that OPI-M1 and -M2 have exactly the same extensive margin and the same import-share adjustment. We thus see, from Tables 4 and 5, that the RW correction for the intensive margin, used in OPI-M2, moves the overall price index closer to the true household price index by 22.0% $(94.1 - 72.0)$. This finding is complementary to the result in Redding and Weinstein (2020) that the RW procedure helps correct the upward bias in the FBW procedure in the price indices constructed from the Nielsen home scanner data for the U.S.[36] To further highlight the results for OPI-M2, we plot its year-dummy estimates and their 95 percent confidence intervals in the right panel of Figure 8, where we have included the estimates under CPI for comparison. We see that the year-dummy estimates of OPI-M2 are closer to zero than those under CPI.

[36] Redding and Weinstein (2020) measure varieties as bar codes and interpret their results to reflect relative taste changes. Our results may reflect other relative changes, such as quality, because our variety measure of HS6 by exporter is broader than bar codes.

6.3 Validation Using Nielsen Home Scanner Data

In this sub-section, we rank price indices by comparing them with the index constructed from the Nielsen home scanner data, which are widely regarded as the best price data available to researchers.[37] Our idea is that, because the Nielsen-home-scanner-data index is the best available measure of the true household price index, the ranking of OPI indices and CPI based on the scanner-data index ought to be similar to the one based on the estimation of the food Engel curve, and so provides an external validation of our results from the previous sub-sections.

Figure 9: The Nielsen Home-Scanner-Data Index and OPIs



Notes: This figure shows price indices obtained from Nielsen home-scanner data together with CPI and OPI indices. The left are for the U.S. food sector only, and the two graphs on the right are for the entire U.S. economy.

The Nielsen Consumer Panel data represent a longitudinal panel of 40,000 to 60,000 U.S. households, who use in-home scanners or mobile apps to record, and report to Nielsen, all their purchases for in-home personal use. The main strength of the Nielsen home scanner data is that they are at the barcodes level. In our data, from 2004-2015,[38] there are 0.6 to 0.7 million barcodes per year. Meanwhile, the Nielsen home scanner data cover about one third of all goods categories in U.S. CPI (RW). Because the mapping between the home-scanner-data products and the non-food merchandise sectors, $s = \{2, ..., S-1\}$, is not straightforward (e.g. Bai and Stumpner (2019)), we

[37] e.g. Broda and Weinstein (2006); Handbury et al. (2013); Handbury and Weinstein (2015). The sample of the Nielsen Retail Scanner data is not as representative as the home-scanner data (RW).

[38] Our access is provided by the Kilts Center for Marketing at the University of Chicago Booth School of Business, and our use of the data is bound by the data agreements with the Kilts Center. Additional information is available at http://research.chicagobooth.edu/nielsen.

construct the home-scanner-data index for food using RW's methodology,[39] and compare it with the food-sector indices of OPI and CPI. Because these OPI indices are at the sector level, their construction does not require the aggregation of equation (24).

The upper-left graph of Figure 9 plots the home-scanner-data index, for food, in red dashed line and red-triangle markers. It also plots the U.S. food CPI in red dashed line with blue-cross markers. The home-scanner-data indices are lower than U.S. CPI, consistent with RW.[40] The two solid blue lines show the following two OPI-T indices for food: the one with blue-cross markers is for OPI-T2, the model of which has multiple non-food merchandise sectors, and the one with orange-triangle markers is for OPI-T3, the model of which has multiple sectors and input-output linkages.[41] Although the two OPI-T indices lie above the U.S. CPI for 2005-2007,[42] they lie between CPI and the home-scanner-data index for the rest of the sample. Overall, OPI-T2 tracks the scanner data index more closely than CPI, and OPI-T3 tracks the home-scanner-data index even better. These results echo our findings in Table 4 and Figure 7.

The same intuition, from these previous results, also explains the patterns in the upper-left graph of Figure 9. Briefly speaking, the U.S. food CPI is subject to the quality and new-goods biases, while the home-scanner-data index, computed following RW, addresses such biases.[43] Both OPI-T indices capture the changes in foreign new goods and quality, and so they track the home-scanner-data index better than CPI. Relative to OPI-T2, OPI-T3 accounts for input-output linkages, and so captures the changes in foreign quality and new goods better, and tracks the home-scanner-data index more closely.

One may be concerned that our previous findings in Table 4 and Figure 7 are for a different sample, 1995-2015. We re-estimate the food Engel curve for the U.S. for 2005-2015,[44] compute the average biases for CPI, OPI-T2 and OPI-T3, for the full U.S. economy, and plot them in the upper-right graph of Figure 9. The ranking here is the same as in Figure 7.

The lower-left graph of Figure 9 compares the home-scanner-data index and U.S. CPI for food with the OPI-M indices for the food sector, shown in blue solid lines. The one with blue-cross markers is OPI-M1, based on the FBW formulation, and the one with orange-triangle markers is OPI-M2, based on the RW formulation. Both OPI-M indices lie above CPI. This result is not contradictory to our previous results regarding OPI-M in Figure 7 and Table 4, because those results are for the different sample of 1995-2015 and for the entire economy. To clarify this point, we re-

[39] We rely heavily on RW's codes, and apply the same data cuts (e.g. we drop all magnet data). We have been able to replicate RW's main results, such as their Table 1 and Figure 3.

[40] RW focus on the Laspereys index of the scanner data. We have found this index to be similar to U.S. CPI, consistent with Broda and Weinstein (2010) and RW.

[41] Recall that these food-sector indices of OPI-T2 and -T3 are computed using the overall CPI. This also explains why we do not include OPI-T1: its difference with food CPI is driven by the difference between the overall CPI and food CPI.

[42] Recall that the OPI-T indices for food, given in equations (15) through (17), are based on CPI for the entire U.S. economy, not CPI for food.

[43] The scanner data are at the barcodes level, and so some of the "quality" bias in raw CPI data may show up as new barcodes and new varieties in the scanner data.

[44] The year 2004 is not available in our U.S. household survey data, PSID. We obtained very similar estimates for β_F, θ, γ as compared with Table 4; e.g. the estimate for β_F is -0.0757.

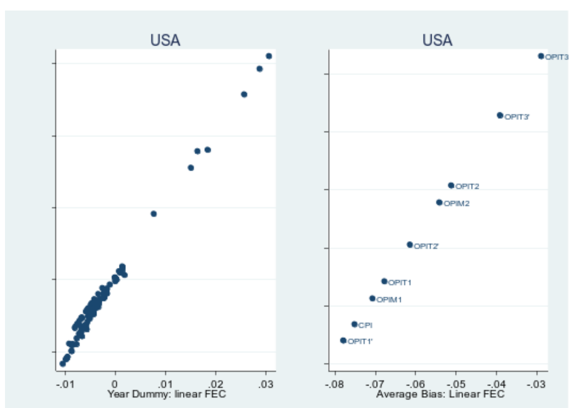
estimate the U.S. food Engel curve for 2005-2015, and plot the average biases of CPI, OPI-M1 and OPI-M2 in the lower right graph of Figure 9. Both OPI-M1 and -M2 have larger average biases (in magnitudes) than CPI.[45] We thus see, again, that the ranking based on the average bias from the food Engel curve is consistent with the ranking based on the home-scanner-data index.

There can be multiple reasons why the OPI-M indices do not track the home-scanner-data index as well as CPI during 2005-2015 for the U.S. As we discussed above, the raw data for OPI-M are not as good as the home-scanner data or the raw CPI data. This likely explains divergence between OPI-M2 and the home-scanner-data index, both of which are based on the RW formulation. In addition, gains from trade for the U.S. might have been lower during 2005-2015 than during 1995-2015,[46] and so they contribute less to the overall performance of the OPI-M indices for 2005-2015 than for 1995-2015.

In summary, the use of the Nielsen-home-scanner-data index delivers similar rankings between OPI and CPI, as compared with the use of the average bias, for the U.S. food sector during 2005-2015. This exercise provides a validation for our approach based on the food Engel curve. While the OPI indices for food do not track the home-scanner-data index perfectly, the OPI indices are readily available for multiple countries and multiple time periods, and provide full coverage of the entire national economy.

6.4 Quadratic Food Engel Curve

Figure 10: U.S. Year-Dummy Estimates and Average Biases with Quadratic Food Engel Curves



Notes: This figure shows estimates from quadratic food Engel curves versus linear food Engel curves for the U.S.

In this sub-section, we address the concern that our food Engel curve, equation (4), is linear in log household income, by augmenting equation (4) with quadratic log income, to obtain the

[45]The year-dummy estimates for 05-15 are very similar to those for 1995-2015 minus the year-2005 estimate. So one could visualize the OPI-M2 and CPI year dummies for 05-15 by shifting their plots in the U.S. graph in Figure 8 up, until the 2005 dummies for OPI-M2 and CPI are both at 0.

[46]For example, the China shock for the U.S. became much smaller after 2007 (e.g., see Autor et al., 2013).

estimates of year dummies and average biases under the following specification,

$$\lambda_{hr,F}^t = \alpha_{Fq} + \beta_{Fq1}(\ln Y_{hr}^t - \ln P^t + \sum d_q^t D^t) + \beta_{Fq2}(\ln Y_{hr}^t - \ln P^t + \sum d_q^t D^t)^2 + \theta_q Z_{hr}^t + \alpha_{qr} + \gamma_q \ln(P_{r,F}^t / P_{r,N}^t) + \varepsilon_{qh}^t. \tag{26}$$

Because the year-t bias estimate from equation (26) is $(-d_q^t)$, we scale d_q^t by β_F, the log-income coefficient of the linear food Engel curve, (4), to make them comparable to our previous results.

We first estimate equation (26) for the U.S., with constant prices and then CPI, and present the results in columns (3) and (4) of Table 3. Just as in the linear Engel curve specification, we obtain identical R^2 and identical estimates of β_{Fq}, θ_q and γ_q in columns (3) and (4). We also see that the coefficient estimates from the quadratic specification of (26) are very similar to our previous results, in columns (1) and (2) of Table 3, obtained from the linear specification of (4).

We then estimate equation (26) with the OPI indices and plot the year-dummy estimates, d_q^t, against our previous results, the estimates of d^t of (4), in the left graph of Figure 10. We include the year dummies with CPI and constant prices in the scatter plots as well. The two sets of estimates have a correlation coefficient of 0.996. We then compute the average biases of the OPI indices and CPI, obtained from equation (26), and plot them against our previous results, obtained from equation (4), in the upper graph of Figure 10. The correlation coefficient between the two sets of average biases is 1.000.

7 Results for Non-U.S. Countries

In this section, we present the results for countries other than the U.S. in our data as listed in Table 1.

7.1 Results for OPI-T

We first discuss the results for individual countries, highlighting their unique features relative to our results for the U.S. Imports account for larger shares of these countries' national economies. In addition, Peru implemented extensive trade policy changes in the 1990's, and so the changes in its OPI-T indices, relative to the benchmark year of 2001, are likely driven by these policy changes.

Table 6 shows, by country, the relative average bias and the fraction of CPI-average-bias explained for the three OPI-T indices. Columns (1) and (2) are based on overall CPI, and columns (3) and (4) are based on domestic CPI.[47] Figure 12 plots the average biases of CPI and OPI-T, by country, as well as their 95 percent confidence intervals. In the top panels, (a)-(d), we have constructed the OPI-T indices using CPI, and in the bottom panels, (e)-(h), we have used domestic CPI.

The U.K. When we use CPI to compute the OPI-T indices, as in columns (1) and (2) of Table 6 and Panel (a) of Figure 12, the average bias of OPI-T3 is smaller in magnitude than CPI, but is larger in magnitude than OPI-T2. In addition, the average bias of OPI-T3 is positive. One interpretation of these results is that the model element of input-output linkages, when used in

[47] Column (4) of Table 6 corresponds to the sum of columns (2) and (3) of Table 4.

Table 6: Mean Relative Average Bias and Percentage of CPI Bias Explained, non-U.S. Countries

	OPI-T with CPI		OPI-T with domestic CPI	
	(1) Relative Average Bias	(2) % CPI Bias Explained by OPI-T	(3) Relative Average Bias	(4) % CPI Bias Explained by OPI-T
U.K.				
OPI-T1	84.0%	16.0%	119.2%	-19.2%
OPI-T2	16.8%	83.2%	52.0%	487.1%
OPI-T3	55.1%	155.1%	19.9%	119.9%
S. Korea				
OPI-T1	97.4%	2.8%	65.9%	34.1%
OPI-T2	96.3%	3.7%	64.9%	35.1%
OPI-T3	86.2%	13.9%	54.8%	45.2%
Canada				
OPI-T1	113.5%	-13.5%	226.7%	-126.7%
OPI-T2	42.6%	57.4%	148.6%	-56.0%
OPI-T3	12.6%	112.6%	93.2%	-0.6%
Peru				
OPI-T1	48.6%	51.4%	17.7%	117.7%
OPI-T2	50.1%	150.1%	100.0%	200.0%
OPI-T3	265.7%	365.7%	325.9%	425.9%

Notes: Column (1) shows, for each OPI-T index, the mean value of its relative average bias across countries, where the relative average bias is computed using equation (7). Column (2) shows the mean value of the fractions of the CPI average bias that OPI-T indices explain, where the fractions are computed using equation (9). The countries and sample years used are as follows: S. Korea, 2006-2015, Canada 2000-2009, and the U.K., 1996-2008,

conjunction with the element of multiple sectors, leads to an over-prediction of the U.K.'s gains from trade for 1996-2008.

An alternative interpretation is that the use of CPI in OPI-T3 has double counted the gains from trade that are due to decreases in import prices. When we use domestic CPI instead, in Panel (e) of Figure 12 and columns (3) and (4) of Table 6, the average bias of OPI-T3 becomes smaller in magnitude and statistically insignificant as well. In comparison, the average bias of OPI-T2 becomes statistically significant and larger in magnitude. Therefore, this alternative interpretation says that the model element of input-output linkages moves the gains-from-trade prediction for the U.K. closer to the true household price index, just like it does for the U.S.

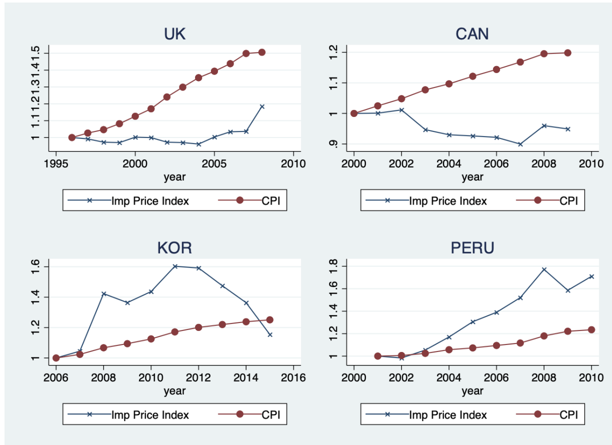
To show which interpretation is more consistent with data, we plot the U.K.'s CPI and import price index as reported by its national statistics in the top left panel of Figure 11. Importantly, like the U.S., the U.K. collects the raw data for its import price index through establishment surveys, and computes this index as Laspeyres.[48] This means that the U.K. import price index is compatible with its CPI. As a result, the figure shows that the U.K. import price index relative to its CPI decreased fairly substantially during 1996-2008, providing evidence that the results with domestic CPI are more appropriate for the U.K.

S. Korea. The results are qualitatively similar to the U.S.; i.e. all three OPI-T indices help explain a fraction of CPI's average bias, and this fraction increases from OPI-T1 to -T2 to -T3.

Like the case for the U.S. and U.K., S. Korea's import price index is Laspeyres and based on

[48]See the documentation at U.K.'s Office of National Statistics, https://www.ons.gov.uk/economy/inflationandpriceindices/methodolo

Figure 11: The Import Price Index and CPI



Notes: This figure shows for each of the countries, U.K., Canada, and South Korea, the import price index based on national statistics as well as CPI.

establishment surveys.[49] The left bottom panel ofFigure 11 shows that S. Korea's import price index increased relative to its CPI during 2006-2015, the pattern of which is opposite to that for the U.K. This increase may have resulted from the depreciation of the Korean won (recall that the import price index is in local currency). Therefore, when we use domestic CPI to compute OPI-T indices, the average biases we obtain are smaller in magnitude, as can be seen from comparing Panel (g) of Figure 12 vs. Panel (c), or columns (3) and (4) of Table 6 vs. columns (1) and (2).

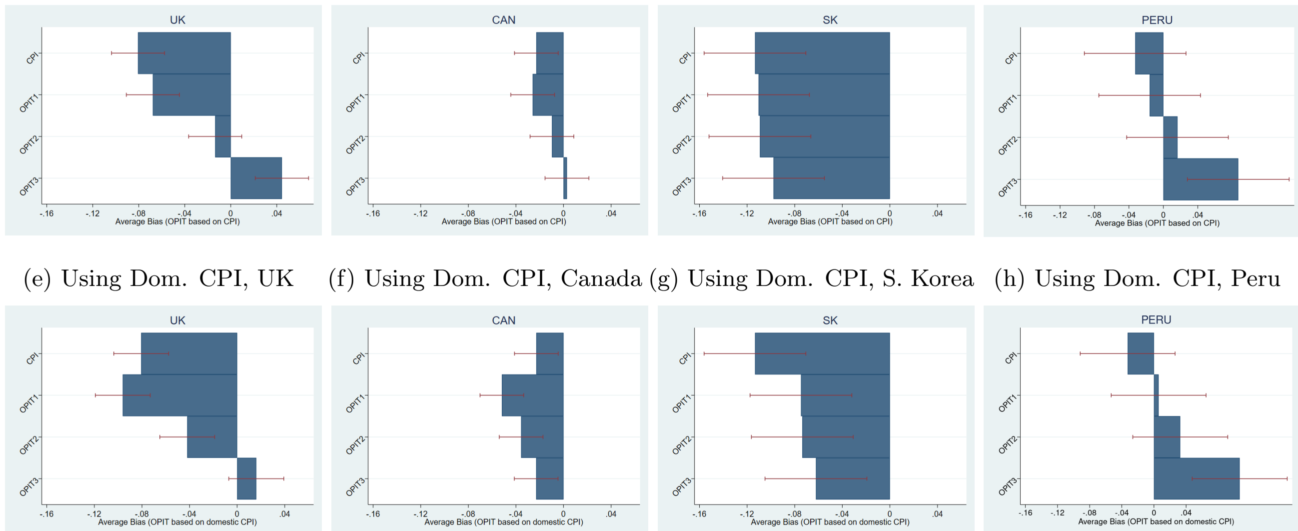
Canada. The results based on CPI are qualitatively similar to the U.S. In particular, the average biases of OPI-T2 and -T3 are both statistically insignificant, as can be seen in Panel (b) of Figure 12. When we use domestic CPI, however, we obtain substantially different results. The average biases of the OPI-T indices are often larger in magnitude than CPI, and they are all statistically significant. These differences arise because, according to Canada's official import price index, the prices of imports throughout 2003-2009 are lower than in the base year of 2000, as illustrated in Figure 11.

Our take is that the results based on overall CPI are probably more credible, because, unlike the U.S., U.K. and S. Korea, Canada collects the raw data for its import price index from customs data and computes this index using the unit values from these data.[50] This implies that Canada's official import price index is conceptually similar to the intensive margin of our OPI-M indices. As we will discuss in the next sub-section, for non-U.S. countries, the use of unit values from customs data could be subject to the issue of exchange-rate pass through.

Peru. Peru signed a number of preferential trade agreements with other Latin American countries in the 1990's, some of which went into effect shortly before 2001, the benchmark year of our sample (e.g. Estevadeordal et al., 2008). Given the small size of Peru's economy, these exogenous policy changes likely drive the changes in Peru's domestic shares relative to 2001, which account

[49]This is based on brief statements by the Bank of Korea, which produces this index.
[50]See the documentation at Statistics Canada, https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=1372645.

Figure 12: The Average Biases of OPI-T and CPI, non-US Countries



(a) Using CPI, UK (b) Using CPI, Canada (c) Using CPI, S. Korea (d) Using CPI, Peru

(e) Using Dom. CPI, UK (f) Using Dom. CPI, Canada (g) Using Dom. CPI, S. Korea (h) Using Dom. CPI, Peru

Notes: This figure shows average bias of OPI-T indices and of CPI for non-US countries. In the top row, OPI-T indices are constructed based on CPI, and in the bottom row they are based on constructed domestic CPI. Red lines are 95 percent confidence intervals.

for the differences between Peru's OPI-T indices and CPI. Like Canada, Peru uses the unit values from customs data in the computation of its import price index, and so we focus on the results for OPI-T based on overall CPI.[51]

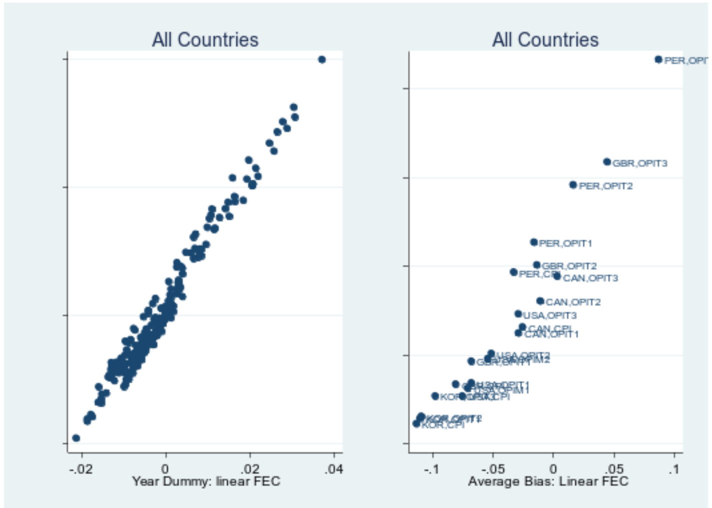
Panels (d) and (h) of Figure 12 shows that the average biases of CPI, OPI-T1 and OPI-T2 are all statistically insignificant for Peru, and the last panel of Table 6 shows that the use of both OPI-T1 and OPI-T2 reduces the magnitude of the average bias by about half, relative to CPI.[52] These results provide a useful validation for our earlier results, because the differences between Peru's OPI-T1 and OPI-T2 indices vs. CPI are likely driven by the aforementioned exogenous changes in Peru's trade policies.

On the other hand, the average bias of OPI-T3 is positive and significant, and larger in magnitude than CPI, suggesting that the use of input-output linkages might have delivered overly optimistic predictions about the changes of Peru's real consumption relative to 2001. This result qualifies our earlier results, where the magnitude of the average bias tends to be the smallest with OPI-T3. It also shows the usefulness of using the true household price index recovered from household consumption data as the benchmark for comparison, as is done in our approach, because the theoretical predictions with larger real-consumption gains from trade may not track observed household consumption data better.

[51] See, e.g., the Peruvian Central Bank's Guide to Methodology, https://www.bcrp.gob.pe/docs/Publicaciones/Guia-Metodologica/nota-semanal/Guia-Metodologica.pdf.

[52] The magnitude of the average bias is slightly larger with OPI-T2 than with OPI-T1. This result, however, is driven by the coefficient estimate of 2002. Without this year, the average bias becomes -0.06 for CPI (p-value = 0.068), -0.04 for OPI-T1 (0.23), -0.0002 for OPI-T2 (0.99), and 0.08 for OPI-T3 (0.01).

Figure 13: Non-U.S. Year-Dummy Estimates and Average Biases with Quadratic Food Engel Curves



In summary, we obtain qualitatively similar results for non-U.S. countries for OPI-T. Unlike the case for the U.S., however, the use of official import price indices, and whether they are compatible with official CPI, are potentially important for the results (e.g. those for Canada).

Finally, we show that our results for non-U.S. countries are also robust to the use of the quadratic food Engel curve. The left graph of Figure 13 plots the year-dummy estimates obtained from the quadratic food Engel curve of (26) against those for our main results, obtained from equation (4), where the OPI-T indices are based on CPI. We have included all countries, including the U.S., but excluded the OPI-M estimates for the non-U.S. countries, for reasons we will describe in the next sub-section. The correlation coefficient between the two sets of estimates is 0.98. The right graph of Figure 13 is a similar plot for the average biases, where the correlation coefficient is 0.96.

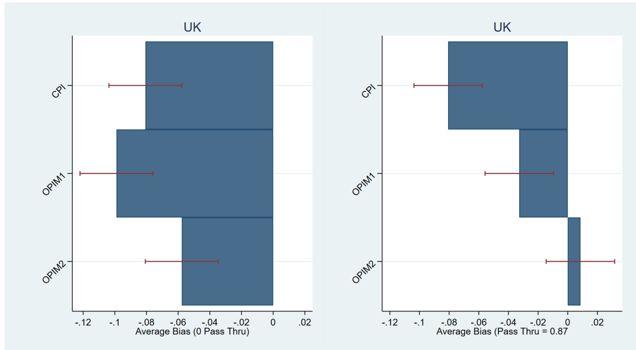
7.2 Results for OPI-M

For the non-U.S. countries, the OPI-M indices we compute are in U.S. dollars, because they are used to invoice many international transactions, and all the product-level prices we have in our import data are in U.S. dollars. Therefore, we need to convert the OPI-M indices into the local-currency-denominated prices that domestic households should face, as predicted by theory. To the best of our knowledge, this issue has not been dealt with in previous studies of the gains-from-trade literature that construct import price indices.[53] We first assume that household prices are unaffected by exchange rates (i.e. zero pass-through), and then apply estimates from the exchange-rate pass-through literature. We use the U.K. to illustrate our results. For most of our sample years, the British pound appreciated relative to the base year of 1996.

Figure 14 plots the average biases of the OPI-M indices and CPI for the U.K., as well as

[53]While some studies have applied the Feenstra (1994) methodology to non-U.S. countries, they focus on the extensive margin (e.g. Mohler and Seitz (2012); Broda et al. (2017)), to which the issue of currency conversion does not apply.

Figure 14: The Average Biases of OPI-M and CPI, the U.K.



Notes: This figure shows the average biases of OPI-M indices for the U.K. In the left panel, the pass-through of exchange rate onto domestically-nominated import prices is assumed to be zero. In the right panel, that is assumed to be 0.87 taken from its long-run estimate for the U.K in Burnstein and Gopinath (2014). Red lines represent 95 percent confidence intervals.

their 95 percent confidence intervals. In the left panel, we assume the pass-through rate of 0, and in the right panel, we apply the long-run pass-through estimate for the import price index from Burstein and Gopinath (2014), 0.87. We see that the average biases of the OPI-M indices change substantially from the left panel to the right. These results clearly illustrate the importance of the exchange-rate pass-through assumption. This assumption turns out to be even more important for the other non-US countries in our sample. Those economies are smaller in terms of GDP and they tend to be influenced more by exchange rate movements. Given that the exchange-rate pass-through literature has produced a relatively wide range of estimates,[54] we refrain from engaging in additional discussions about the results for the OPI-M indices for non-U.S. countries.

8 Conclusion

Recent years have witnessed rapid growth in the studies that predict how much consumers gain from international trade. In contrast, our knowledge remains limited as to what extent the gains-from-trade predictions by this literature are consistent with observed household consumption decisions. In addition, this literature has used a number of model specifications and generated a range of gains-from-trade predictions. It is unclear whether specific model elements render the overall predictions more consistent, or less consistent, with household consumption decisions.

In this paper, we incorporate the core predictions from this literature about real consumption into over-time changes in open-economy price indices, or OPI indices. We then use the food Engel curve, a strong empirical regularity in micro household survey data, to quantify the deviations of OPI indices from the true price index that households use to deflate their income, and compare them with

[54]For example, Burstein and Gopinath (2014) regress log changes in prices on eight quarterly lags in log changes in the trade-weighted nominal exchange rate. They define the sum of coefficients on all lags as the long-run pass through. The regression is done for a handful of countries, including the U.S., the U.K., and Canada. Burstein and Gopinath (2014) find that pass-through is larger in border prices than it is in retail prices, that pass-through deepens over time but is incomplete in the long-run, and that there is a marked difference in estimates across countries.

the deviations of CPI. We show that the OPI indices tend to follow the true price index more closely than official CPI; i.e. household consumption behavior is more consistent with the predictions of the gains-from-trade literature than the average price changes measured by CPI. We also show that the overall gains-from-trade predictions tend to better track household consumption decisions if we allow for multiple industries and input-output linkages in quantitative general-equilibrium models, or if we correct for demand residuals in the computation of import-price indices as in Redding and Weinstein (2020). Our results provide a validation check of the core predictions, as well as specific model elements and specifications, of the gains-from-trade literature.

More broadly, official CPI is extensively used for policy. While there is a broad agreement that CPI does not track household prices well over time due to quality and new-goods biases, there is no consensus about better alternatives. In turn, our results establish a connection between the gains-from-trade literature and the research on official CPI. The sufficient-static approach in the trade literature (e.g., ACR), a cornerstone of the gains-from-trade literature, provides a solution to the quality bias and new-goods bias of official CPI, both of which are difficult problems for the CPI literature. By operationalizing the complementarity between sufficient statistics from commonly-used trade theories and official CPI in a transparent and tractable way, our OPI-T indices provide one potential alternative to be used alongside CPI for policy analysis and discussion.

References

Almås, I. (2012). International income inequality: Measuring ppp bias by estimating engel curves for food. *American Economic Review*, 102(2):1093–1117. 1, 2

Almås, I., Beatty, T. K., and Crossley, T. F. (2018). Lost in translation: What do engel curves tell us about the cost of living? 14, 17

Amiti, M., Dai, M., Feenstra, R. C., and Romalis, J. (2020). How did china's wto entry affect us prices? *Journal of International Economics*, 126:103339. 1

Anderson, J. E. and Yotov, Y. V. (2020). Short run gravity. *Journal of International Economics*, 126:103341. 10

Antràs, P. and De Gortari, A. (2020). On the geography of global value chains. *Econometrica*, 88(4):1553–1598. 10

Argente, D. and Lee, M. (2021). Cost of living inequality during the great recession. *Journal of the European Economic Association*, 19(2):913–952. 14

Arkolakis, C., Costinot, A., and Rodriguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102(1):94–130. 1, 1

Arkolakis, C., Demidova, S., Klenow, P. J., and Rodríguez-Clare, A. (2008). Endogenous variety and the gains from trade. *American Economic Review*, 98(2):444–50. 6

Arkolakis, C., Ramondo, N., Rodríguez-Clare, A., and Yeaple, S. (2018). Innovation and production in the global economy. *American Economic Review*, 108(8):2128–73. 10

Armington, P. S. (1969). A theory of demand for products distinguished by place of production. *Staff Papers*, 16(1):159–178. 4.1

Atkin, D., Faber, B., Fally, T., and Gonzalez-Navarro, M. (2020). Measuring welfare and inequality with incomplete price information. Technical report, National Bureau of Economic Research. 1, 15, 17, 4.4

Auer, R., Burstein, A., Lein, S. M., and Vogel, J. (2022). Unequal expenditure switching: Evidence from switzerland. Technical report, National Bureau of Economic Research. 14, 23, 30

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *The American Economic Review*, 103(6):2121–2168. 6.1, 46

Bai, L. and Stumpner, S. (2019). Estimating us consumer gains from chinese imports. *American Economic Review: Insights*, 1(2):209–24. 1, 6.3

Bai, Y., Jin, K., and Lu, D. (2019). Misallocation under trade liberalization. Technical report, National Bureau of Economic Research. 10

Balistreri, E. J., Hillberry, R. H., and Rutherford, T. F. (2011). Structural estimation and solution of international trade models with heterogeneous firms. *Journal of international Economics*, 83(2):95–108. 6

Banks, J., Blundell, R., and Lewbel, A. (1997). Quadratic engel curves and consumer demand. *Review of Economics and statistics*, 79(4):527–539. 2, 4.4

Baqaee, D. and Burstein, A. T. (2021). Welfare and output with income effects and taste shocks. *NBER Working Paper*, (w28754). 1, 19, 22

Barrett, G. F. and Brzozowski, M. (2010). Using engel curves to estimate the bias in the australian cpi. *Economic record*, 86(272):1–14. 12

Beatty, T. K. and Larsen, E. R. (2005). Using engel curves to estimate bias in the canadian cpi as a cost of living index. *Canadian Journal of Economics/Revue canadienne d'économique*, 38(2):482–499. 12

Berndt, E. R. (2006). The boskin commission report after a decade: After-life or requiem? *International Productivity Monitor*, 12:61. 13, 5.3

Borusyak, K. and Jaravel, X. (2018). The distributional effects of trade: Theory and evidence from the united states. *Available at SSRN 3269579*. 14

Boskin, M. J., Dulberger, E. R., Gordon, R. J., Griliches, Z., and Jorgenson, D. W. (1997). The cpi commission: Findings and recommendations. *The American Economic Review*, 87(2):78–83. 13, 5.3

Broda, C., Greenfield, J., and Weinstein, D. E. (2017). From groundnuts to globalization: A structural estimate of trade and growth. *Research in Economics*, 71(4):759–783. 53

Broda, C. and Weinstein, D. E. (2006). Globalization and the gains from variety. *The Quarterly journal of economics*, 121(2):541–585. 1, 1, 21, 5.1, 6.2, 37, **??**

Broda, C. and Weinstein, D. E. (2010). Product creation and destruction: Evidence and price implications. *American Economic Review*, 100(3):691–723. 6.2, 40

Brown, D. K., Stern, R. M., et al. (1989). Us-canada bilateral tariff elimination: The role of product differentiation and market structure. *Trade policies for international competitiveness*, pages 217–54. 11

Burstein, A. and Gopinath, G. (2014). International prices and exchange rates. In *Handbook of international economics*, volume 4, pages 391–451. Elsevier. 7.2, 54

Caliendo, L., Dvorkin, M., and Parro, F. (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. 10

Caliendo, L., Dvorkin, M., Parro, F., et al. (2015). Trade and labor market dynamics. *Federal Reserve Bank of St. Louis Working Paper Series*, (2015-009). **??**

Caliendo, L. and Parro, F. (2015). Estimates of the trade and welfare effects of nafta. *The Review of Economic Studies*, 82(1):1–44. 1, 3

Chor, D. and Ma, L. (2020). Contracting frictions in global sourcing: Implications for welfare. Technical report. 10

Chung, C., Gibson, J., and Kim, B. (2010). Cpi mismeasurements and their impacts on economic management in korea. *Asian Economic Papers*, 9(1):1–15. 12

Clerc, M.-e. and Coudin, E. (2011). Does the cpi reflect changes in the cost of living in france? what the analysis of engel curves tells us. *Economie & Statistique*. 12

Costa, D. L. (2001). Estimating real income in the united states from 1888 to 1994: Correcting cpi bias using engel curves. *Journal of political economy*, 109(6):1288–1310. 1, 2

Costinot, A. and Rodríguez-Clare, A. (2014). Trade theory with numbers: Quantifying the consequences of globalization. In *Handbook of international economics*, volume 4, pages 197–261. Elsevier. 1, 3

Costinot, A. and Rodríguez-Clare, A. (2018). The us gains from trade: Valuation using the demand for foreign factor services. *Journal of Economic Perspectives*, 32(2):3–24. 9

Dabalen, A., Gaddis, I., and Nguyen, N. T. V. (2020). Cpi bias and its implications for poverty reduction in africa. *The Journal of Economic Inequality*, 18(1):13–44. 12

de Carvalho Filho, I. and Chamon, M. (2012). The myth of post-reform income stagnation: Evidence from brazil and mexico. *Journal of Development Economics*, 97(2):368–386. 12

Deaton, A. and Muellbauer, J. (1980). An almost ideal demand system. *The American economic review*, 70(3):312–326. 2, 2.1, 2.1

Deaton, A. S. and Muellbauer, J. (1986). On measuring child costs: With applications to poor countries. *Journal of Political Economy*, 94(4):720–744. 2

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934. 10

Du, X. and Wang, Z. (2021). Multinational production, innovation relocation, and the consequences of globalization. *Innovation Relocation, and the Consequences of Globalization*. 10

Eaton, J., Eslava, M., Jinkins, D., Krizan, C. J., and Tybout, J. R. (2021). A search and learning model of export dynamics. Technical report, National Bureau of Economic Research. 10

Eaton, J., Jinkins, D., Tybout, J. R., and Xu, D. (2022). Two-sided search in international markets. Technical report, National Bureau of Economic Research. 10

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779. 1, 3, 4.1

Engel, E. (1895). *Die Lebenskosten belgischer Arbeiter-Familien früher und jetzt*. Heinrich. 1

Estevadeordal, A., Freund, C., and Ornelas, E. (2008). Does regionalism affect trade liberalization toward nonmembers? *The Quarterly Journal of Economics*, 123(4):1531–1575. 7.1

Fajgelbaum, P. D. and Khandelwal, A. K. (2016). Measuring the unequal gains from trade. *The Quarterly Journal of Economics*, 131(3):1113–1180. 2, 14

Fally, T. and Sayre, J. (2018). Commodity trade matters. Technical report, National Bureau of Economic Research. 9

Farrokhi, F. (2020). Global sourcing in oil markets. *Journal of International Economics*, 125:103323. 9

Farrokhi, F. and Pellegrina, H. S. (2022). Trade, technology, and agricultural productivity. *STEG Working Paper*. 10

Farrokhi, F. and Soderbery, A. (2022). Trade elasticities in general equilibrium: Demand, supply, and aggregation. 9

Feenstra, R. C. (1994). New product varieties and the measurement of international prices. *The American Economic Review*, pages 157–177. 1, 1, 21, 6.2, 53

Feenstra, R. C., Luck, P., Obstfeld, M., and Russ, K. N. (2018). In search of the armington elasticity. *Review of Economics and Statistics*, 100(1):135–150. 27

Feenstra, R. C. and Weinstein, D. E. (2017). Globalization, markups, and us welfare. *Journal of Political Economy*, 125(4):1040–1074. 7

General Accounting Office (1995). Us imports: Unit values vary widely for identically classified commodities. Technical report, Report GAO/GGD-95-90. 6.2

Giri, R., Yi, K.-M., and Yilmazkuday, H. (2021). Gains from trade: Does sectoral heterogeneity matter? *Journal of International Economics*, 129:103429. 6, 32

Gordon, R. J. (2006). The boskin commission report: A retrospective one decade later. Technical report, National Bureau of Economic Research. 13, 5.3

Greenlees, J. S. and McClelland, R. (2011). New evidence on outlet substitution effects in consumer price index data. *Review of Economics and Statistics*, 93(2):632–646. 13

Hamilton, B. W. (2001). Using engel's law to estimate cpi bias. *American Economic Review*, 91(3):619–630. 1, 2, 2

Handbury, J., Watanabe, T., and Weinstein, D. E. (2013). How much do official price indexes tell us about inflation? Technical report, National Bureau of Economic Research. 1, 37

Handbury, J. and Weinstein, D. E. (2015). Goods prices and availability in cities. *The Review of Economic Studies*, 82(1):258–296. 37

Hausman, J. (2003). Sources of bias and solutions to bias in the consumer price index. *Journal of Economic Perspectives*, 17(1):23–44. 1, 6.1.2

Hertel, T. (2013). Global applied general equilibrium analysis using the global trade analysis project framework. In *Handbook of computable general equilibrium modeling*, volume 1, pages 815–876. Elsevier. 11

Hsieh, C.-T. and Ossa, R. (2016). A global view of productivity growth in china. *Journal of international Economics*, 102:209–224. 6

Hsu, W.-T., Riezman, R. G., and Wang, P. (2019). Innovation, growth, and dynamic gains from trade. Technical report, National Bureau of Economic Research. 10

Imbs, J. and Mejean, I. (2015). Elasticity optimism. *American Economic Journal: Macroeconomics*, 7(3):43–83. 5.1, ??

Jaravel, X. and Sager, E. (2019). What are the price effects of trade? evidence from the us and implications for quantitative trade models. 7

Kehoe, T. J., Pujolas, P. S., and Rossbach, J. (2017). Quantitative trade models: Developments and challenges. *Annual Review of Economics*, 9:295–325. 1

Kehoe, T. J., Rossbach, J., and Ruhl, K. J. (2015). Using the new products margin to predict the industry-level impact of trade reform. *Journal of International Economics*, 96(2):289–297. 1

Klenow, P. J. and Kryvtsov, O. (2008). State-dependent or time-dependent pricing: Does it matter for recent us inflation? *The Quarterly Journal of Economics*, 123(3):863–904. 6.1.2, 6.2

Krolikowski, P. M. and McCallum, A. H. (2021). Goods-market frictions and international trade. *Journal of International Economics*, 129:103411. 10

Kucheryavyy, K., Lyn, G., and Rodríguez-Clare, A. (2016). Grounded by gravity: A well-behaved trade model with industry-level economies of scale. Technical report, National Bureau of Economic Research. 9

Larsen, E. R. (2007). Does the cpi mirror the cost of living? engel's law suggests not in norway. *Scandinavian Journal of Economics*, 109(1):177–195. 12

Levchenko, A. A. and Zhang, J. (2016). The evolution of comparative advantage: Measurement and welfare implications. *Journal of Monetary Economics*, 78:96–111. 6

Lileeva, A. and Trefler, D. (2010). Improved access to foreign markets raises plant-level productivity... for some plants. *The Quarterly journal of economics*, 125(3):1051–1099. 29

Lind, N. and Ramondo, N. (2018). Trade with correlation. Technical report, National Bureau of Economic Research. 10

Markusen, J. R., Rutherford, T. F., and Hunter, L. (1995). Trade liberalization in a multinational-dominated industry. *Journal of International Economics*, 38(1-2):95–117. 11

McCaig, B. (2011). Exporting out of poverty: Provincial poverty in vietnam and us market access. *Journal of International Economics*, 85(1):102–113. 29

Melitz, M. J. and Redding, S. J. (2015). New trade models, new welfare implications. *American Economic Review*, 105(3):1105–46. 9

Melser, D. and Syed, I. A. (2016). Life cycle price trends and product replacement: Implications for the measurement of inflation. *Review of Income and Wealth*, 62(3):509–533. 13

Mohler, L. and Seitz, M. (2012). The gains from variety in the european union. *Review of World Economics*, 148(3):475–500. 53

Moulton, B. R. (1996). Bias in the consumer price index: what is the evidence? *Journal of Economic perspectives*, 10(4):159–177. 1, 6.1.2

Moulton, B. R., Moses, K. E., Gordon, R. J., and Bosworth, B. P. (1997). Addressing the quality change issue in the consumer price index. *Brookings Papers on Economic Activity*, 1997(1):305–366. 13

Nakamura, E. and Steinsson, J. (2008). Five facts about prices: A reevaluation of menu cost models. *The Quarterly Journal of Economics*, 123(4):1415–1464. 6.1.2

Nakamura, E., Steinsson, J., and Liu, M. (2016). Are chinese growth and inflation too smooth? evidence from engel curves. *American Economic Journal: Macroeconomics*, 8(3):113–44. 1, 2, 2.1

Ossa, R. (2015). Why trade matters after all. *Journal of International Economics*, 97(2):266–277. 6, 6.1.3

Redding, S. and Weinstein, D. E. (2018). Accounting for trade patterns. *Princeton University, mimeograph*. 21

Redding, S. J. and Weinstein, D. E. (2020). Measuring aggregate price indices with taste shocks: Theory and evidence for ces preferences. *The Quarterly Journal of Economics*, 135(1):503–560. 1, 4.2, 4.2, 6.2, 6.2, 36, 8

Sacerdote, B. (2017). Fifty years of growth in american consumption, income, and wages. Technical report, National Bureau of Economic Research. 12

Shapiro, J. S. (2016). Trade costs, co 2, and the environment. *American Economic Journal: Economic Policy*, 8(4):220–54. 10, **??**

Shapiro, M. D. and Wilcox, D. W. (1996). Mismeasurement in the consumer price index: An evaluation. *NBER macroeconomics annual*, 11:93–142. 13

Vartia, Y. O. (1976). Ideal log-change index numbers. *scandinavian Journal of statistics*, pages 121–126. 1

Appendices

A Data Appendix

A.1 Additional Details about Data

A.1.1 BACI-CEPII data

Our variety-level trade data come from BACI-CEPII, which has annual values and quantities of bilateral trade at the level of 6-digit Harmonized System (HS6) codes. We use BACI data to compute variety-level unit values and import-value shares, the variables $p^t_{j,gs}$ and $\lambda^t_{j,gs}$ in equations (19) - (21), by defining a variety j as a pair of export country and HS6 product.

While the computation of the OPI-M indices accommodates the change of the variety set within a given SITC good over time, it requires that the common set of varieties be non-empty. In addition, the computation of the intensive margin of OPI-M2 involves taking the simple geometric mean within the common set, which might be an issue if the common set contains too few varieties.

We thus merge certain SITC goods in the raw BACI data, to ensure that the common set contains at least 20 varieties for each SITC good by year. Table A1 lists the SITC goods that are involved in the merges for the U.S.

A.1.2 STAN and UNIDO data

We obtain OECD STAN data on gross production, exports and imports by sector by year. The STAN sectors are aggregate ISIC (International Standard Industrial Classification) version 4. We have 1 sector for the food category, F, which combines agricultural and manufactured food, and 11 sectors for the category of tradeable merchandise, G. Since STAN does not cover Peru, we use UNIDO data for Peru. The UNIDO sectors are ISIC version 3, and we convert them into the same codes as STAN.

Our STAN and UNIDO data provide the values of the variables $\pi^{D,t}_s$ in equations (14) - (16), as gross output minus exports divided by apparent consumption, which is gross output minus exports plus imports, as well as the variable $\pi^{M,t}_s$ in equations (22) and (23), as $1 - \pi^{D,t}_s$. Below, we explain how we use extrapolations to deal with the following two data issues: missing gross-output data, and negative domestic shares.

Missing gross-output data. Case 1 involves the following countries, years and sectors: U.S., 1995 and 1996, 31-32 (furniture & other manufacturing) and 01-03 (food & agriculture); U.K., 1996, 05-08 (mining), and 10-12 (food & agriculture). In this case, we regress gross output on values of imports and exports for each sector-by-country using the years with non-missing data. The R^2 of these regressions are high (e.g. it is 0.96 for U.S. 01-03 and 0.86 for U.K. 05-08). We extrapolate the missing gross-output values using the regression coefficients and the data for imports and exports.

Case 2 involves 31-32 (furniture & other manufacturing) for Canada for multiple years. When the gross-output values of 31-32 are missing, they are aggregated into 31-33. We compute the mean output share of 31-32 within 31-33 using the years with non-missing data, and extrapolate the missing data for 31-32 using this mean output share and the gross-output values of 31-33.

Negative domestic shares. In STAN, like in other industry-output data, the value of gross output could be smaller than the value of export, because of such issues as re-exports, allocation of secondary activities, and so on (e.g. the STAN database documentation, 2005). Likewise for UNIDO.

Case 1 is for the U.K. for the following years and sectors: 2011, 26-28 (machinery & electronics), 2012, 31-32 (furniture & other manufacturing), and 2013, 24-25 (metals). In these cases, $\pi_s^{D,t} < 0$ in the raw data, which is inconsistent with trade theory.

In the raw data, the sector 24-25 shows a large jump in exports in 2013, and 31-32 shows a large jump in both imports and exports in 2012. Meanwhile, the sector 26-28 does not show major jumps in imports, exports, or gross output in 2011.

For the sector 24-25, we regress exports on imports and gross output using the years other than 2013 ($R^2 = 0.82$), and then replace the raw export data in 2013 with the predicted value from this regression. For 31-32, we replace the raw import and export values in 2012 with their mean values in 2011 and 2013. For 26-28, we fit a lowess regression for the series of domestic-share data, and then replace the negative value in 2011 with the predicted value from lowess.

Case 2 is for Peru for 24-25 (metals) for the years of 2008-2009, and 2011-2017. In the raw data, imports, exports and domestic production do not show large jumps during these years. We regress domestic production on imports and exports using the other years in our data ($R^2 = 0.98$), and use this relationship to extrapolate domestic production. These predicted values of domestic production deliver positive $\pi_s^{D,t}$, except for 2014 and 2015. We next fit a lowess regression for $\pi_s^{D,t}$, using all years of data, and replace the negative $\pi_s^{D,t}$ values of 2014 and 2015 with their positive predicted values.

A.1.3 CPI Data

We obtain the prices and weights of official CPI from the national statistical agencies, for the three categories of F, G, and V (non-tradable services). Note that F includes food and beverages at home, and that food and beverages away from home are not in F, but in V.

Unlike the other countries in our sample, the U.K. excludes owners' housing from its official CPI. Although U.K. published CPIH, a variant of its official CPI that includes owners' housing, in 2013, CPIH was suspended in 2014 due to data quality issues. As a result, we adjust U.K.'s official CPI to include owners' housing in the following steps.

In step 1, we regress the service-category weight of U.S. CPI on this weight in the U.S. national accounts, NIPA ($R^2 = 0.85$), and then use the regression coefficients, together with U.K.'s NIPA service weight, to extrapolate U.K.'s CPI service weight. Before the extrapolation, the U.K.'s CPI service weights are around 40%, and after the extrapolation, they are around 60%, in line with the other high-income countries in our data, the U.S., Canada, and S. Korea.

In step 2, we regress the index of owners' rent in U.S. CPI on the Case-Shiller Housing Index for the U.S. ($R^2 = 0.89$), and use the regression coefficients, together with U.K.'s Halifax Housing Price Index, to extrapolate the index of owners' rent for U.K. CPI.

In step 3, the final step, we adjust U.K.'s official CPI using the service-category weights and owners' rent index that we computed in steps 1 and 2 above.

Our CPI data provide the values of the CPI-weight and CPI-price variables in equations (22) and (23).

A.1.4 Peruvian NHS Data

The Peruvian National Household Survey (NHS) is available for 1998-2010. The years of 1998-2000, however, have very small sample sizes, relative to the other years, as we show in Table A2. We thus use the data of 2001-2010.

A second issue with the NHS data is that, while the methodology of computing household expenditures is the same from 2004 throughout 2010, it is different for 2001-2003. Fortunately, the NHS data for 2004 is available under both the old methodology (used for 2001-2003) and the new one. Comparing these two data sets, we see that they cover identical individuals and households and report identical nominal income, but report different household expenditures. We regress the new-methodology food expenditure and total expenditure on their old-methodology counterparts ($R^2 = 0.99$ and 0.97, respectively), and use these relationships to convert the 2001-2003 expenditure data to the same basis as the 2004-2010 data.

A.1.5 Other Data

Our data on input-output linkages come from WIOD (World Input-Output Database). We use WIOD to compute the Leontief-inverse matrix, and obtain the values of its elements, the variables $\tilde{\alpha}_{ks}$ in equation (16). Since WIOD does not cover Peru, we use GTAP data for Peru.

We obtain import shares by country by year, β^t_M in equation (25), from the national-accounts data of WDI (World Development Indicators).

A.2 Additional Details about Parameter Values

A.2.1 Industry-Mean Substitution Elasticity for OPI-M

As discussed in Section 4, we use goods-specific substitution elasticities, σ_g, to compute the OPI-M indices, and one may be concerned about the large number of parameter values involved. In this sub-section, we construct the OPI-M indices with sector-level substitution elasticities, σ_s, instead. To do so, we maintain the same goods layers within sectors as before, and apply σ_s to the goods-level price indices. This change affects the extensive margins of both OPI-M indices, as shown in equation (19), as well as the intensive margin of OPI-M2, as shown in equations (19) and (21). Meanwhile, this change has no effect on the intensive margin of OPI-M1, as shown in equations (21) and (20), or the import-share adjustment of both indices, as shown in equations (22) and (23).

We obtain similar year-dummy estimates and similar average biases for the U.S. Pooling across both OPI-M1 and -M2, the correlation coefficient between the two sets of year-dummy estimates is 0.98, and that between the two sets of average biases is 0.92.

A.2.2 Industry-level Trade Elasticities across the Literature

Table 1 reports estimates of the trade elasticity, $\theta_s = \sigma_s - 1$, as we use in this paper and across a few studies in the literature.

The last row of Table A3 shows that the correlation between the literature's values and ours is fairly high (except for Bagwell et al. 2018). XXX copied from what we had before ... I am not sure about it though XXX

Table 1: Trade Elasticity Estimates across Select Papers

	Industry	This paper based on BW-2006	Imbs and Mejean - Feenstra	Imbs and Mejean - CP	Shapiro-2016	CP-2015
1	Food	2.9	6.1	4.03	4.3	2.6
2	Textile and Apparel	2.4	6.0	5.73	14.3	5.6
3	Wood and Paper	1.6	3.2	12.36	5.9	10.0
4	Petroleum	8.0	8.5	41.77	8.9	51.1
5	Chemicals	2.1	4.2	5.16	1.6	4.8
6	Plastics	0.9	3.5	2.37	1.6	1.7
7	Minerals	1.0	4.7	1.71	8.9	2.8
8	Metals	4.5	3.8	13.13	12.9	6.1
9	Machinery and Electronics	1.0	6.0	12.47	10.8	8.2
10	Transport equipment	2.3	4.9	5.53	6.9	0.7
11	Furniture and Other Mfg	1.0	3.6	7.74	12.8	5.0
12	Mining	25.6	9.6	24.50	3.5	15.7

Notes: This table shows each industry's trade elasticity estimates, θ_s, as we use in this paper based on BW-2006 (Broda and Weinstein, 2006), and across select papers in the literature: (Imbs and Mejean, 2015) based on Feenstra's procedure and based on Caliendo and Parro's procedure, (Shapiro, 2016), and (Caliendo et al., 2015).

B Theory Appendix

B.1 Formulas for OPI-M

Nested CES Structure. Sector-level consumption bundle aggregates imported and domestic bundles in the following CES fashion:

$$C_s^t = \left[\left(C_s^{M,t}\right)^{(\sigma_s-1)/\sigma_s} + \left(C_s^{D,t}\right)^{(\sigma_s-1)/\sigma_s} \right]^{\sigma_s/(\sigma_s-1)}$$

where σ_s is the elasticity of substitution between imported and domestic sector–level bundles. In turn, the imported bundle of sector s, $C_s^{M,t}$, aggregates over imported bundle of goods within sector s,

$$C_s^{M,t} = \left[\sum_{g \in \Omega_s} \left(C_{gs}^{M,t} \right)^{(\sigma_s^M - 1)/\sigma_s^M} \right]^{\sigma_s^M/(\sigma_s^M - 1)}$$

where σ_s^M is the elasticity of substitution between imported bundle of goods within sector s. The imported bundle of good g in sector s aggregates over varieties j,

$$C_{gs}^{M,t} = \left[\sum_{j \in \Omega_{gs}^{M,t}} \left(c_{j,gs}^t \right)^{(\sigma_g - 1)/\sigma_g} \right]^{\sigma_g/(\sigma_g - 1)}$$

where σ_g is the elasticity of substitution between varieties within good g.

Weights in Intensive Margin of the Price Index. Below, we derive the price index that corresponds to the consumption bundle of the lowest tier, $C_{gs}^{M,t}$. Derivations for upper tiers are similar. To simplify the notation, below we drop superscript M, and subscript gs, and work with:

$$C^t = \left[\sum_{j \in \Omega^t} \left(b_j^t \right)^{1/\sigma} \left(c_j^t \right)^{(\sigma - 1)/\sigma} \right]^{\sigma/(\sigma - 1)} \tag{B.1}$$

The associated price index is:

$$P^t \equiv P(\{p_j^t\}, \{b_j^t\}, \Omega^t) = \left[\sum_{j \in \Omega^t} b_j^t (p_j^t)^{1-\sigma} \right]^{1/(1-\sigma)} \tag{B.2}$$

Share of expenditure on variety j equals:

$$\lambda_j^t \equiv \lambda(\{p_j^t\}, \{b_j^t\}, \Omega^t) = \frac{b_j^t (p_j^t)^{1-\sigma}}{P_t^{1-\sigma}} \tag{B.3}$$

And, the change to the price index is:

$$\hat{P}^t = \frac{P(\{p_j^t\}, \{b_j^t\}, \Omega^t)}{P(\{p_j^{t-1}\}, \{b_j^{t-1}\}, \Omega^{t-1})} \tag{B.4}$$

We only focus on the intensive margin of price changes since our derivations for the extensive margin is the precisely the same as the ones in the previous literature. Focusing on the intensive margin, suppose the set of varieties purchased in t and $t-1$ are the same, $\Omega^t = \Omega^{t-1} = \Omega$. By equation (B.3), $P^t = (\lambda_j^t / b_j^t)^{1/(1-\sigma)} p_j^t$. Using this expression and the definition (B.4),

$$\hat{P}^t = \Big(\frac{\lambda_j^t / b_j^t}{\lambda_j^{t-1} / b_j^{t-1}} \Big)^{1/(\sigma - 1)} \Big(\frac{p_j^t}{p_j^{t-1}} \Big)$$

Taking logs:

$$\ln \hat{P}^t = \frac{1}{\sigma - 1} \ln \left(\frac{\lambda_j^t / b_j^t}{\lambda_j^{t-1} / b_j^{t-1}} \right) + \ln \left(\frac{p_j^t}{p_j^{t-1}} \right)$$

By reorganizing,

$$\frac{1}{\ln \left(\frac{\lambda_j^t / b_j^t}{\lambda_j^{t-1} / b_j^{t-1}} \right)} \ln \hat{P}^t - \frac{1}{\ln \left(\frac{\lambda_j^t / b_j^t}{\lambda_j^{t-1} / b_j^{t-1}} \right)} \ln \left(\frac{p_j^t}{p_j^{t-1}} \right) = \frac{1}{\sigma - 1}$$

Multiply this equation by $(\lambda_j^t - \lambda_j^{t-1})$ and sum over varieties $j \in \Omega$,

$$\sum_{j \in \Omega} \frac{\lambda_j^t - \lambda_j^{t-1}}{\ln \left(\frac{\lambda_j^t / b_j^t}{\lambda_j^{t-1} / b_j^{t-1}} \right)} \ln \hat{P}^t - \sum_{j \in \Omega} \frac{\lambda_j^t - \lambda_j^{t-1}}{\ln \left(\frac{\lambda_j^t / b_j^t}{\lambda_j^{t-1} / b_j^{t-1}} \right)} \ln \left(\frac{p_j^t}{p_j^{t-1}} \right) = \sum_{j \in \Omega} (\lambda_j^t - \lambda_j^{t-1}) \frac{1}{\sigma - 1}$$

From $\sum_{j \in \Omega} \lambda_j^t = \sum_{j \in \Omega} \lambda_j^t - \lambda_j^{t-1} = 1$, it follows that $\sum_{j \in \Omega} (\lambda_j^t - \lambda_j^{t-1}) = 0$. Therefore,

$$\ln \hat{P}^t \sum_{j \in \Omega} \frac{\lambda_j^t - \lambda_j^{t-1}}{\ln(\lambda_j^t / b_j^t) - \ln(\lambda_j^{t-1} / b_j^{t-1})} = \sum_{j \in \Omega} \frac{\lambda_j^t - \lambda_j^{t-1}}{\ln(\lambda_j^t / b_j^t) - \ln(\lambda_j^{t-1} / b_j^{t-1})} \ln \left(\frac{p_j^t}{p_j^{t-1}} \right)$$

Hence, the change to price index equals:

$$\ln \hat{P}^t = \sum_{j \in \Omega} d_j^t \ln \left(\frac{p_j^t}{p_j^{t-1}} \right)$$

where the weight on variety j, d_j^t , is given by:

$$d_j^t = \frac{\frac{\lambda_j^t - \lambda_j^{t-1}}{\ln(\lambda_j^t / b_j^t) - \ln(\lambda_j^{t-1} / b_j^{t-1})}}{\sum_{j \in \Omega} \frac{\lambda_j^t - \lambda_j^{t-1}}{\ln(\lambda_j^t / b_j^t) - \ln(\lambda_j^{t-1} / b_j^{t-1})}} \tag{B.5}$$

Equation (B.5) reproduces the weights in OPI-M FBW and RW (equations 20 and 21). The former is the case under the assumption that $b_j^t = b_j^{t-1}$.

Equivalence to RW. Note that we could rewrite the change in the price index as follows:

$$\ln \hat{P}^t = \frac{1}{\sigma - 1} \ln \left(\frac{\lambda_j^t}{\lambda_j^{t-1}} \right) + \ln \left(\frac{p_j^t / (b_j^t)^{\frac{1}{\sigma - 1}}}{p_j^{t-1} / (b_j^{t-1})^{\frac{1}{\sigma - 1}}} \right)$$

Here, the shift in our perspective is to think of $(b_j^t)^{\frac{1}{\sigma-1}}$ as a price-equivalent taste parameter. That is, to think of the whole term $p_j^t / (b_j^t)^{\frac{1}{\sigma-1}}$ as price itself. For a generic variable x, define $\bar{x}$ as:

$$\bar{x}^t = \prod_{j \in \Omega} \left(x_j^t \right)^{\alpha_j}$$

for a set of $\{\alpha_j\}_{j\in\Omega}$ satisfying $\alpha_j > 0$ for all $j \in \Omega$ and $\sum_{j\in\Omega} \alpha_j = 1$. We follow RW and assume that $\alpha_j = 1/|\Omega|$ for all varieties j, where $|\Omega|$ is the number of varieties in the common set Ω, and that average taste remains unchanged, i.e., $\bar{d}_{t-1} = \bar{d}_t$. Therefore,

$$\ln \hat{P}^t = \frac{1}{\sigma - 1} \ln\left(\frac{\bar{\lambda}^t}{\bar{\lambda}^{t-1}}\right) + \ln\left(\frac{\bar{p}^t}{\bar{p}^{t-1}}\right) \tag{B.6}$$

which reproduces equation (9) in RW. Our equivalent formula is based on recovering taste parameters, and plug them back as weights. First, we can recover demand parameters according to:

$$\left(\ln b_j^t - \ln b_j^{t-1}\right) = \ln\left[\left(\frac{\lambda_j^t}{\bar{\lambda}^t}\right) \Big/ \left(\frac{\lambda_j^{t-1}}{\bar{\lambda}^{t-1}}\right)\right] - (1-\sigma)\ln\left[\left(\frac{p_j^t}{\bar{p}^t}\right) \Big/ \left(\frac{p_j^{t-1}}{\bar{p}^{t-1}}\right)\right]$$

Replacing the above into equation (B.6) reproduces equation (B.5).