Online Appendix to "The Distributional Effects of Trade: Theory and Evidence from the United States"

Kirill Borusyak, UCL and CEPR
Xavier Jaravel, LSE and CEPR
June 2021

A Proofs of Section 2.1 Results

Proof of Equation (2).

For a differentiable indirect utility function $\mathcal{V}$, equivalent variation EV_i for consumer i solves, by definition:

$$\mathcal{V}(p^0, X_i^0 + EV_i) - \mathcal{V}(p^0, X_i^0) = \mathcal{V}(p^0 + dp, X_i^0 + dX_i) - \mathcal{V}(p^0, X_i^0),$$

which for small shocks to prices and total expenditures implies

$$\frac{\partial \mathcal{V}}{\partial X} EV_i = \frac{\partial \mathcal{V}}{\partial X} dX_i + \frac{\partial \mathcal{V}}{\partial p} dp,$$

$$EV_i = dX_i - \left(-\frac{\partial \mathcal{V}/\partial p}{\partial \mathcal{V}/\partial X}\right) dp = dX_i - \sum_\omega q_\omega^i dp_\omega,$$

where q_ω^i is the initial consumption of ω, and the last equality holds by Roy's identity. Thus,

$$d\log \mathcal{W}_i \equiv \frac{EV_i}{X_i} = \frac{dX_i}{X_i} - \sum_\omega s_\omega^i \frac{dp_\omega}{p_\omega} = d\log X_i - \sum_\omega s_\omega^i d\log p_\omega.$$

Proof of Proposition 1.

Under Assumption 1, prices are equal to marginal costs. For variety ω produced in country $\gamma(\omega)$, the marginal cost (or unit cost) is denoted by m_ω, and the cost minimization problem implies

$$m_\omega = M_\omega \left(\left\{ m_\ell \tau_{\ell\gamma(\omega)} \right\}_{\ell \in \Omega}, w_{\gamma(\omega)} \right), \tag{8}$$

where M_ω is the cost function, Ω is the set of all varieties produced anywhere, $\tau_{\ell\gamma(\omega)}$ are the iceberg trade costs for delivering input ℓ (produced in country $\gamma(\ell)$) to country $\gamma(\omega)$, and $w_{\gamma(\omega)}$ is the vector of factor prices in country $\gamma(\omega)$. We assume system (8) has a unique solution around the initial equilibrium. Since factor prices do not change by Assumption 2, Shephard's lemma implies:

$$
\begin{aligned}
d\log m_\omega &= \sum_{\ell\in\Omega} \beta_\ell^\omega \left(d\log \tau_{\ell\gamma(\omega)} + d\log m_\ell\right) \\
&= \sum_{\ell\in\Omega} \beta_\ell^\omega d\log \tau_{\ell\gamma(\omega)} + \sum_{\ell\in\Omega} \beta_\ell^\omega \sum_{k\in\Omega} \beta_k^\ell d\log \tau_{k\gamma(\ell)} + \dots, \qquad (9)
\end{aligned}
$$

where β_ℓ^ω is the direct cost share of intermediate input ℓ in the production of ω.[40] Intuitively, the change in the unit production cost of ω depends on changes in the costs of all intermediate inputs through changes in trade costs, including higher-order terms along domestic and international supply chains.

Since we consider counterfactual shocks to trade costs between foreign suppliers c and and the Home country only (rather than between foreign suppliers), the only non-zero terms in (9) correspond to inputs directly imported from c to H. By Assumption 3, varieties ω produced abroad and imported into H use no inputs from Home. Therefore, the unit production costs for imported products remain unchanged, i.e. $d\log m_\omega = 0$ for all imported varieties. By perfect competition and complete pass-through, consumer prices in H for varieties imported from c change in proportion to changes in trade costs: $d\log p_\omega = d\log\tau$ for $\omega\in\Omega_c$, with $d\log p_\omega = 0$ for all other imported varieties.

For varieties that are domestically produced, the change in consumer price reflects changes in domestic production costs through intermediate inputs, i.e. $d\log p_\omega = d\log m_\omega$, and (9) simplifies to

$$
\begin{aligned}
d\log m_\omega &= \underbrace{\sum_{\ell\in\Omega_c} \beta_\ell^\omega d\log\tau}_{\text{direct impact on unit cost}} + \underbrace{\sum_{\ell\in\Omega_H} \beta_\ell^\omega d\log m_\ell}_{\text{indirect impact via domestic IO linkages}} \\
&= \left(\sum_{\ell\in\Omega_c} \beta_\ell^\omega + \sum_{\ell\in\Omega_H} \beta_\ell^\omega \sum_{k\in\Omega_c} \beta_k^\ell + \dots\right) d\log\tau \equiv \widetilde{IP}_{\omega c}^{\text{Int}} d\log\tau,
\end{aligned}
$$

Here the second line is a sum across all domestic supply chains fragments starting from a good imported from c and leading to the production of ω. Summation across them yields the overall share of inputs imported from c in ω's total production cost,

[40] The direct cost share is defined as $\beta_\ell^\omega = m_\ell \tau_{\ell\gamma(\omega)} q_\ell^\omega / m_\omega$, where q_ℓ^ω is the unit requirement of input ℓ in production of ω in equilibrium.

$\widetilde{IP}^{\text{Int}}_{\omega c}$, that satisfies the recursive definition of footnote 11 and measures the sensitivity of the output price to the change the iceberg costs.

Given the expressions above for price changes above, and since total expenditures do not change when earnings do not change,[41] Proposition 1 follows by (2).

Proof of Equation (3).

Define $\widetilde{IP}_{\omega c}$ as $\widetilde{IP}^{\text{Int}}_{\omega c}$ for $\omega \in \Omega_H$, 1 for $\omega \in \Omega_c$, and 0 otherwise. Then we have:

$$
\begin{aligned}
ImpSh^i_c - ImpSh^0_c &= \sum_\omega \left(s^i_\omega - s^0_\omega\right) \widetilde{IP}_{\omega c} \\
&= \sum_r \left(s^i_r ImpSh^i_{rc} - s^0_r ImpSh^0_{rc}\right) \\
&= \sum_r \left(s^i_r - s^0_r\right) ImpSh^0_{rc} + \sum_r s^i_r \left(ImpSh^i_{rc} - ImpSh^0_{rc}\right).
\end{aligned}
$$

B Parametric Frameworks

In this appendix, we provide details for the setting and results from Section 4: first for the AIDS demand system of Fajgelbaum and Khandelwal (2016, henceforth FK), and then for the nested NNHCES demand system we employ instead.

B.1 AIDS Demand System

Preferences. We first briefly review the AIDS demand system used by FK and their estimation procedure, based on bilateral trade data from the World Input-Output Database at the country-industry level. The demand system, with the "extended Cobb-Douglas" restrictions imposed by FK, is given by the following indirect utility function:

$$
\mathcal{V}(w,p) = \frac{\log w - a(p)}{\exp\left(\sum_{j,c} \beta_{jc} \log p_{jc}\right)},
$$

where j indexes $J = 35$ industries (including both goods and services and assuming that there are no intermediate goods), c indexes $C = 40$ countries, a constraint

[41] We maintain the proportionality of expenditures and earnings even in Section 5.1 where we allow for trade imbalances and thus the budget constraint violated at the aggregate country level.

$\sum_{j,c}\beta_{jc}=0$ is imposed, and

$$a(p)=\underline{\alpha}+\sum_{j,c}\alpha_{jc}\log p_{jc}+\frac{1}{2}\sum_{j}\gamma_j\left(\frac{1}{C}\left(\sum_{c}\log p_{jc}\right)^2-\sum_{c}\left(\log p_{jc}\right)^2\right).$$

By Roy's identity and as long as an interior solution exists, demand can be written in the share form as

$$s_{jc}\left(w,p\right)=\left(\alpha_{jc}-\beta_{jc}a(p)\right)+\beta_{jc}\log w-\gamma_j\left(\log p_{jc}-\frac{1}{C}\sum_{c'}\log p_{jc'}\right). \quad (10)$$

We return to the issue of corner solutions below.

Assuming all consumers in the same country n face the same prices p^n, the expenditure share on each variety is linear in log-income, $\frac{\partial s_{jc}(w,p^n)}{\partial \log w}=\beta_{jc}$. As a consequence, there exists an income level w_n such that country's aggregate expenditure share s^n_{jc} observed in the trade data equals $s_{jc}(w_n,p^n)$. When the income distribution within a country is not observed, this representative consumer income can be approximated, e.g. by calibrating a log-normal income distribution as in FK.[42] Predicting expenditure shares for consumers with income $w\neq w_n$ is straightforward given β_{jc}:

$$s_{jc}(w,p^n)=s^n_{jc}+\beta_{jc}\left(\log w-\log w_n\right). \quad (11)$$

The demand system thus extrapolates from the observed expenditure share for the representative agent, s^n_{jc}, to the rest of the income distribution via the estimated Engel curve.

Estimation. A key challenge when estimating the demand parameters with cross-country data is that prices are unobserved. FK address this challenge by assuming a simple structure of iceberg trade costs, whereby the price p^n_{jc} of variety (j,c) in country n is given by

$$\log p^n_{jc}=\log m_{jc}+\rho' d_{jcn}+\epsilon^n_{jc}. \quad (12)$$

Here m_{jc} is a unit production cost, ρ captures the relationship between trade costs and measures of proximity between the countries, d_{jcn}, such as the log of geographic

[42] There is a minor error in how this approximation is done by FK, which we fixed in our replication. Specifically, with $\log w\sim\mathcal{N}\left(\mu_n,\sigma_n^2\right)$, it is easy to verify that $\log w_n=\mu_n+\sigma_n^2$, while the code by FK instead used $\log w_n=\mu_n+\sigma_n$, meaning that the representative consumer was attributed to an incorrect percentile of the income distribution unless $\sigma_n=1$. However, for many countries including the U.S. the calibrated σ_n is not far from one.

distance or an indicator for common language, all interacted with industry indicators, and ϵ_{jc}^{n} is an error term. FK then plug in (12) into (10) evaluated at $w = w_n$ to obtain the non-homothetic gravity equation; see their Section IV.A for details of the estimation procedure.

Mechanical pro-poor expenditure channel within industries. We now show that, in the presence of home bias and income-inelastic tradable industries, the AIDS demand system mechanically generates pro-poor differences in import shares across the income distribution, which arise within industries.

First, note that in equation (11) the expenditure share of a variety is less sensitive to consumer income (in terms of elasticities) in countries that have a higher expenditure share on that variety. Log-differentiating (11) around $w = w_n$, we have

$$\frac{\partial \log s_{jc}\left(w_n, p^n\right)}{\partial \log w} = \frac{\beta_{jc}}{s_{jc}^{n}}. \tag{13}$$

Using this result, consider the predictions of AIDS regarding import shares within goods-producing industries across the income distribution. Denoting $\beta_{j,-n} = \sum_{c \neq n} \beta_{jc}$, $\beta_j = \sum_c \beta_{jc}$, and similarly $s_{j,-n}^{n} = \sum_{c \neq n} s_{jc}^{n}$ and $s_j^n = \sum_c s_{jc}^{n}$, (11) implies that the import share within tradable industry j is

$$IP_j^n(w, p^n) \equiv \frac{s_{j,-n}^{n}\left(w, p^n\right)}{s_j^n(w, p^n)} = \frac{s_{j,-n}^{n} + \beta_{j,-n}\left(\log w - \log w_n\right)}{s_j^n + \beta_j\left(\log w - \log w_n\right)}. \tag{14}$$

In the data, all but one goods-producing industries are income-inelastic, i.e. have $\beta_j < 0$. For income-inelastic industries, differentiating (14) yields the following condition for the import share in industry j to decline with income w: $\frac{\partial IP_j^n(w,p^n)}{\partial \log w} < 0$ if and only if

$$\frac{s_{j,-n}^{n}}{s_j^n} < \frac{\beta_{j,-n}}{\beta_j}. \tag{15}$$

Whether this condition holds in the data generally depends on the Engel curve parameters. With the estimates of FK, this condition holds for 599 out of the 600 pairs of (j, n) within goods-producing income-inelastic industries. This is due to home bias: the import share in industry j on the left-hand side tends to be far below one; the median import share among these (j, n) pairs is 39%. In contrast, the right-hand side of (15) tends to be close to one, since $\beta_{j,-n}$ and β_j are sums of 39 and 40 country parameters that differ only by one term (corresponding to the domestic Engel curve parameter, β_{jn}). Intuitively, (15) almost always holds because the income elasticity of foreign varieties is magnified due to home bias, per equation (13). In the single exception case, the estimated domestic Engel curve parameter, β_{jn}, is sufficiently large, overturning this effect.

We have thus shown that with AIDS the fraction of import spending within goods-producing industries is predicted to fall with income, because of the combination of home bias and the fact that these industries are almost all income-inelastic. However, applying the same derivations to services may seem to imply that an offsetting pattern should arise for income-elastic services, i.e. low-income consumers should have lower import shares within services. In fact, corner solutions prevent this offsetting effect from operating in practice.

As mentioned, equation (10) only applies to interior solutions. However, corner solutions are common in the estimated demand system of FK: for example, a U.S. consumer at the 25th percentile of the income distribution is predicted to have zero shares on 1,015 out of 1,365 foreign varieties, particularly in services (658 out of 741 varieties).[43] This happens because, according to (11), the expenditure share on income-elastic varieties, characterized by $\beta_{jc} > 0$, can turn negative for incomes just slightly below w_n, specifically when $w < w_n \exp\left(-s^n_{jc}/\beta_{jc}\right)$. This issue is most relevant for foreign services which have particularly low observed trade shares s^n_{jc} and tend to be income-elastic. Appendix A of FK describes an iterative procedure which removes these negative shares, yielding the appropriate corner solution. This procedure sets the negative shares to zero (e.g., for an imported service variety) and scales down the shares of other varieties in the same industry (e.g., for the domestic service variety). Therefore, for lower-income consumers the fraction of foreign services is not significantly lower than for the representative consumer, as it rapidly hits a

[43] These zero shares are due to the extrapolation in equation (10): there are *no* zero shares in the observed data s^n_{jc} for $n =$ U.S. Corner solutions are key for understanding two important results obtained by FK. First, without corner solutions it would be impossible to have import shares fall with income in all countries. Indeed, assuming (10) holds and aggregating it across all domestic varieties ($c = n$) yields that the total domestic expenditure share grows with income if and only if domestic varieties are more income elastic than the world average: $\frac{\partial \sum_j s_{jn}}{\partial \log w} = \sum_j \beta_{jn}$, where the world average of the right-hand side terms, $\frac{1}{C}\sum_c \sum_j \beta_{jn}$, is zero. Yet, the import shares in the estimated FK model do fall with income in all countries, e.g. comparing the 25th and 75th percentiles (see Panel A of our Figure 4 for the U.S. as an example). Second, for the same reason it would be impossible for import shares to have a region of growth at the top of the distribution, producing an overall U-shape, while this is again the case in the estimates of FK for all countries (see again Panel A of Figure 4 for the U.S.).

corner and stays at zero. In contrast, the share of foreign goods in the consumption basket (and within the total spending on goods) falls with income according to (11).[44]

B.2 Nested Non-Homothetic CES

We now consider a nested version of the non-homothetic CES utility function of Comin et al. (2021), which we define recursively by

$$\begin{aligned} \mathcal{U} &= \left(\sum_j Q_j^{(\varepsilon-1)/\varepsilon} \right)^{\varepsilon/(\varepsilon-1)}, \\ Q_j &= \left(\sum_c \left(a_{jc} \mathcal{U}^{\varphi_{jc}(\xi_j-1)} \right)^{1/\xi_j} Q_{jc}^{(\xi_j-1)/\xi_j} \right)^{\xi_j/(\xi_j-1)}. \end{aligned} \tag{16}$$

Here the first line is a CES aggregate of consumption across industries (outer nest), with elasticity ε, and the second line defines sectoral consumption Q_j as an aggregate of quantities Q_{jc} across country-specific varieties with elasticity $\xi_j > 1$ and taste shifters a_{jc}. Primitive parameters $\varphi_{jc} < 1$ determine how non-homothetic tastes $a_{jc}\mathcal{U}^{\varphi_{jc}(\xi_j-1)}$ vary with consumer utility; we will show that high φ_{jc} translates into high income elasticity of the variety.[45]

It is straightforward to derive utility-dependent price indices:

$$\begin{aligned} p_j^* &= \left(\sum_c a_{jc} \mathcal{U}^{\varphi_{jc}(\xi_j-1)} p_{jc}^{1-\xi_j} \right)^{1/(1-\xi_j)}, \\ \pi^* &= \left(\sum_j p_j^{*1-\varepsilon} \right)^{1/(1-\varepsilon)}, \end{aligned} \tag{17}$$

[44] The same mechanism explains the slight U-shape in import shares found by FK. For income levels sufficiently above w_n the shares of income-inelastic imported goods hit the zero bound. Those high-income consumers are therefore predicted to buy a lot of foreign services, and this cannot be compensated by a lower share of foreign goods (which is already 0%), such that the import share begins to rise. Such U-shape is an inherent feature of AIDS under home bias: it would arise whenever there are *any* differences in income elasticities across industries; it is more severe for incomes below the representative agent's because home bias is particularly strong for income-elastic services.

[45] Our way of writing the utility function differs slightly from that of Comin et al. (2021) to better resemble traditional nested CES when $\varphi_{jc} \equiv 0$. A single-tier version of (16) is equivalent to equation (1) in Comin et al. (2021) with $1 - \varphi_{jc}$ as relevant income elasticity parameters. As in Comin et al. (2021), a parameter restriction $\varphi_{jc} < 1$ is required for integrability.

such that the agent's utility always satisfies at the optimal consumption bundle:

$$\mathcal{U} = w/\pi^*. \tag{18}$$

The Hicksian expenditure shares are then given by

$$s_{jc} \equiv \underbrace{\frac{a_{jc}\mathcal{U}^{\varphi_{jc}(\xi_j-1)}p_{jc}^{1-\xi_j}}{p_j^{*1-\xi_j}}}_{s_{c|j}} \cdot \underbrace{\frac{p_j^{*1-\varepsilon}}{\pi^{*1-\varepsilon}}}_{s_j}. \tag{19}$$

Income and substitution patterns with NNHCES. We first show how expenditure shares vary with income and prices. Define $\bar{\varphi}_j = \sum_c s_{c|j}\varphi_{jc}$ and $\bar{\varphi} = \sum_j s_j\bar{\varphi}_j$. Log-differentiating (17) we have:

$$d\log p_j^* = \sum_c s_{c|j}\left(d\log p_{jc} - \varphi_{jc} d\log\mathcal{U}\right) = d\log p_j - \bar{\varphi}_j d\log\mathcal{U}, \tag{20a}$$

$$d\log\pi^* = \sum_j s_j d\log p_j^* = d\log\pi - \bar{\varphi} d\log\mathcal{U}, \tag{20b}$$

where $\mathrm{d}\log\mathcal{U}$ is the log-change in cardinal utility, while $d\log p_j = \sum_c s_{c|j} d\log p_{jc}$ and $d\log\pi = \sum_j s_j d\log p_j$ are the industry-level and overall Laspeyres price indices for the consumer, respectively. Together with (18), (20b) implies

$$d\log\mathcal{U} = d\log w - d\log\pi^* = \frac{d\log w - d\log\pi}{1-\bar{\varphi}}. \tag{21}$$

This equation relates changes in the cardinal utility to observable objects only: the money metric of the welfare gain $d\log\mathcal{W}$ (change in the total expenditure minus the Laspeyres price index) and spending shares at the original equilibrium (which enter $\bar{\varphi}$).

We can now express changes in demand in terms of observables: log-differentiating (19) and plugging in (20) and (21) yields

$$\begin{aligned} d\log s_{jc} &= \varphi_{jc}\left(\xi_j - 1\right) d\log\mathcal{U} + \left(1-\xi_j\right)\left(d\log p_{jc} - d\log p_j^*\right) + \left(1-\varepsilon\right)\left(d\log p_j^* - d\log\pi^*\right) \\ &= \left(1-\xi_j\right)\left(d\log p_{jc} - d\log p_j\right) + \left(1-\varepsilon\right)\left(d\log p_j - d\log\pi\right) \\ &\quad + \left(\psi_{jc} - 1\right)\left(d\log w - d\log\pi\right), \end{aligned} \tag{22}$$

where the income elasticity of variety jc is given by

$$\psi_{jc} = 1 + \frac{\left(\xi_j - 1\right)\left(\varphi_{jc} - \bar{\varphi}_j\right) + \left(\varepsilon - 1\right)\left(\bar{\varphi}_j - \bar{\varphi}\right)}{1-\bar{\varphi}}. \tag{23}$$

According to (22), the change in the expenditure share on variety jc has three components. The first two are identical to conventional nested CES, capturing the substitution effects across varieties within the industry and across industries, respectively. The third term is the income effect, shaped by the income elasticity ψ_{jc}. When welfare increases ($d\log w - d\log \pi > 0$), so does the spending share on income-elastic products (those with $\psi_{jc} > 1$).

Equation (23) implies that NNHCES, in contrast with AIDS, does not produce the mechanical relationship between the income elasticity and the spending shares on that variety. Indeed, it is immediate from (23) that if two varieties in the same industry j from countries c and c' have the same $\varphi_{jc} = \varphi_{jc'}$, then $\psi_{jc} = \psi_{jc'}$ for every consumer in the world.[46] More generally, among two varieties in the same industry, the one with higher φ_{jc} is always more income-elastic for every consumer: $\psi_{jc} - \psi_{jc'} = (\varphi_{jc} - \varphi_{jc'}) \frac{\xi_j - 1}{1-\bar{\varphi}}$, with $\frac{\xi_j - 1}{1-\bar{\varphi}}$ always positive. While this demand system is flexible and can capture that certain varieties are high- or low-income elastic (with correspondingly high or low φ_{jc}), it does not hard-wire differences between foreign and domestic varieties.

Baseline estimation. We now estimate the key φ_{jc} parameters by adapting the non-homothetic CES estimation strategy proposed by Comin et al. (2021, Appendix A.2.2) to accommodate nests and the cross-country setting of FK. We assume that the observed equilibrium is generated by identical NNHCES preferences, and each country n is populated by representative agents with known income w_n.[47] We further assume that price differences across countries follow equation (12). We focus on estimating φ_{jc}, assuming that $\xi_j \equiv \xi$ and ε are known (we use $\xi = 3.5$ and $\varepsilon = 2$ as in Section 5.2).

We also impose a normalization. Like in the single-tier version of Comin et al. (2021), preferences are over-parameterized, in that rescaling all $1 - \varphi_{jc}$ by a positive constant only changes the cardinal value of utility but does not affect preferences or Marshallian demand. We therefore assume, without loss of generality, that the

[46] Like in every non-homothetic demand system, income elasticities still vary across income groups, via $\bar{\varphi}_j$ and $\bar{\varphi}$ in (23).

[47] Because non-homothetic CES preferences do not admit a positive representative consumer (Mas-Colell et al. 1995, p.116), this should be viewed as an approximation. However, the same issue arises for AIDS when corner solutions exist for some consumers in the country, as is the case for FK. For comparability with FK we use the same representative consumer (see footnote 42). Estimation of either demand system without invoking representative consumers would be a useful avenue for future work.

average of $1-\varphi_{jc}$ across all varieties, weighted by their output shares λ_{jc} in the world output at the observed equilibrium, equals one. For any variable z_{jc} we denote $\check{z}=\sum_{jc}\lambda_{jc}z_{jc}$, with conventions that $\log\check{z}=\sum_{jc}\lambda_{jc}\log\check{z}_{jc}$ (with the log taken before averaging) and $\check{z}^n=\sum_{jc}\lambda_{jc}z^n_{jc}$.

We now derive the estimating equation. First, using (19) and (18), we define augmented expenditure shares in each country n which eliminate the complications related to the two-tier structure of our demand system:

$$S^n_{jc}\equiv s^n_{c|j}\left(s^n_j\right)^{(1-\xi)/(1-\varepsilon)}=a_{jc}\mathcal{U}_n^{(\varphi_{jc}-1)(\xi-1)}\left(\frac{p^n_{jc}}{w_n}\right)^{1-\xi}$$

or, in the log form,

$$\log S^n_{jc}=\log a_{jc}+(\xi-1)\left(\log w_n-\log p^n_{jc}\right)+(\varphi_{jc}-1)(\xi-1)\log\mathcal{U}_n. \qquad (24)$$

Averaging (24) across j,c with weights λ_{jc} and using the imposed parameter normalization $\check{\varphi}=0$, we solve for the unknown $\log\mathcal{U}_n$ in each country:

$$\log\check{S}^n-\log\check{a}-(\xi-1)(\log w_n-\log\check{p}^n)=-(\xi-1)\log\mathcal{U}_n. \qquad (25)$$

We can now substitute for utility in (24) to obtain a function of expenditure shares and prices only. Plugging (25) and rearranging terms yields

$$\begin{aligned}\log S^n_{jc}-\log\check{S}^n=&(\log a_{jc}-\log\check{a})-(\xi-1)\left(\log p^n_{jc}-\log\check{p}^n\right)\\&-\varphi_{jc}\left(\log\check{S}^n-\log\check{a}-(\xi-1)(\log w_n-\log\check{p}^n)\right).\end{aligned}$$

Finally, plugging in the structure of iceberg trade costs from (12), we obtain[48]

$$\begin{aligned}\log S^n_{jc}-\log\check{S}^n=&FE_{jc}-(\xi-1)\rho'\left(d_{jcn}-\check{d}_n\right)\\&+\varphi_{jc}\left[(\xi-1)\log w_n-\log\check{S}^n-(\xi-1)\rho'\check{d}_n\right]+\tilde{\epsilon}^n_{jc}, \qquad (26)\end{aligned}$$

where FE_{jc} is the variety fixed effect invariant to the importer,[49] and the error is given by

$$\tilde{\epsilon}^n_{jc}=-(\xi-1)\rho'(\epsilon_{jcn}-\check{\epsilon}_n)-\varphi_{jc}(\xi-1)\rho'\check{\epsilon}_n.$$

We estimate equation (26) by nonlinear least squares where unknown parameters

[48]We note that because of (12) the model does not permit zero trade shares, which are found in the trade data, particularly for some service industries. We therefore winsorize the shares from below at the first % percentile of non-zero shares (1.04×10^{-10}), which affects 9.4% observations, almost all of which correspond to foreign varieties in service industries.

[49]Specifically, $FE_{jc}=\log a_{jc}+(\varphi_{jc}-1)\log\check{a}-(\xi-1)\log m_{jc}-(\xi-1)(\varphi_{jc}-1)\sum_{jc}\lambda_{jc}\log m_{jc}$.

are φ_{jc} and FE_{jc} for each of the 1,400 varieties, as well as 140 gravity parameters ρ. Specifically, we use the gravity variables d_{jcn} from FK (log of geographic distance and indicators for common border and common language) as well as the indicator for $c = n$ to capture border effects; these four variables are interacted with 35 industry indicators to allow for industry-specific elasticities, as in FK.[50]

Identification of the key parameters φ_{jc} in equation (26) is very intuitive: a variety is more income elastic if richer countries buy relatively more of it, conditionally on prices captured by the gravity terms. This logic applies because the terms in square brackets recover the utility of the representative agent by adjusting her nominal income for price differences, and φ_{jc} is estimated, given ρ, by a cross-country regression of relative shares (adjusted for the nesting structure) on this imputed utility measure.

Constrained estimation. As explained in the main text, baseline estimates of NNHCES predict that goods overall are too income-inelastic, compared with what we see in the CEX data. We therefore re-estimate the parameters φ_{jc} keeping within-industry differences in income elasticities as in the baseline estimation but adjusting across-industry differences. We match the income elasticity of tradable goods for the U.S. representative consumer to the value of 0.864 obtained from our CEX-IO data of Section 2.2 (see Appendix S.2.5).

Concretely, equation (23) shows how within- and across-industry differences in φ_{jc} parameters respectively translate into within- and across-industry differences in income elasticities. We therefore take the baseline estimates $\varphi_{jc}^{\text{baseline}}$ and ρ^{baseline} and look for across-industry shifts $\Delta\varphi_j$ such that $\varphi_{jc} = \varphi_{jc}^{\text{baseline}} + \Delta\varphi_j$ best fits (26) while imposing

$$\frac{\sum_{j\in\text{goods}}\sum_c s_{jc}^{\text{US}}\psi_{jc}^{\text{US}}}{\sum_{j\in\text{goods}}\sum_c s_{jc}^{\text{US}}} = 0.864 \tag{27}$$

Given $\varphi_{jc}^{\text{baseline}}$ and ρ^{baseline}, estimation of (23) becomes linear.[51]

Inferring expenditure shares across the income distribution. We finally explain how we use the estimated demand system to impute the expenditure shares for

[50]To ease computation we note that, conditionally on ρ, (26) is linear. We therefore perform nonlinear search over ρ only. Moreover, to satisfy the the integrability constraint $\varphi_{jc} < 1$ and avoid numerical instability, we impose the constraints $\varphi_{jc} \leq 0.9$; these constraints are binding for only 5.5% of varieties, many of which in the pseudo-industry "Private Households."

[51]Goods-producing industries are those coded 1–16 in the data of FK. We continue to require that $\sum_{j,c}\lambda_{jc}\varphi_{jc} = 0$ and $\varphi_{jc} \leq 0.9$.

consumers at various levels of consumer income, as in Panels B and C of Figure 4. This procedure is based on exact hat algebra and does not involve any approximation. We will use hats here to denote exact log-differences between the representative consumer with income w_n (the initial point) and a consumer with income level w (the end point, marked by primes). For example, $\hat{w} = \log \frac{w}{w_n}$ and $\hat{s}_{jc} = \log \frac{s_{jc}(w,p^n)}{s^n_{jc}}$.

Log-differencing (19) between w_n and w and noting that price indices p^*_j and π^* are income-dependent implies:

$$\hat{s}_{jc} = \varphi_{jc} (\xi_j - 1) \hat{\mathcal{U}} + (\xi_j - 1) \hat{p}^*_j + (1 - \varepsilon) \left(\hat{p}^*_j - \hat{\pi}^* \right). \tag{28}$$

To solve for $\hat{\mathcal{U}}$, $\hat{\pi}^*$, and $\hat{p}^*_j$, we use (17) and (18) to write

$$(1 - \xi_j) \hat{p}^*_j = \log \frac{\sum_c a_{jc} (\mathcal{U}')^{\varphi_{jc}(\xi_j - 1)} p_{jc}^{1-\xi_j}}{\sum_c a_{jc} \mathcal{U}^{\varphi_{jc}(\xi_j - 1)} p_{jc}^{1-\xi_j}} = \log \sum_c s_{c|j} \exp \left(\varphi_{jc} (\xi_j - 1) \hat{\mathcal{U}} \right) \tag{29a}$$

and similarly

$$(1 - \varepsilon) \hat{\pi}^* = \log \sum_j s_j \exp \left((1 - \varepsilon) \hat{p}^*_j \right), \tag{29b}$$

$$\hat{\mathcal{U}} = \hat{w} - \hat{\pi}^*. \tag{29c}$$

Combining the three lines in (29) we arrive at an equation in $\hat{\mathcal{U}}$ only:

$$\hat{\mathcal{U}} = \hat{w} - \frac{1}{1 - \varepsilon} \log \sum_j s_j \exp \left(\frac{1 - \varepsilon}{1 - \xi_j} \log \sum_c s_{c|j} \exp \left(\varphi_{jc} (\xi_j - 1) \hat{\mathcal{U}} \right) \right),$$

which is solved numerically for each $\hat{w}$ of interest. Substituting $\hat{\mathcal{U}}$ back to (29a), (29b), and (28), we retrieve the extrapolated shares as $s_{jc}(w, p^n) = s^n_{jc} \exp(\hat{s}_{jc})$ for each variety.

C Details and Proofs of Section 5.1 Results

Appendix C.1 proves Proposition 2 and characterizes the welfare changes for each worker after a uniform trade shock in GE. Appendix C.2 then considers special cases: with NNHCES demand system and with the structure of the labor market as in each of our two calibrations. Finally, Appendix C.3 proves the first-order approximation for the effects of trade shocks on inequality, equation (7).

C.1 Proof of Proposition 2

We consider changes in the $I \times 1$ vector of wages w and the $J \times 1$ vector of value-added by industry (measured in monetary terms) VA. We allow not only trade costs but also the $I \times 1$ labor supply vector L (measured in efficiency units) to change. This additional generality will be useful to define the macro elasticity of factor demand.

We first derive two equations, respectively characterizing labor and product market equilibria in log-changes:

$$d\log w = \boldsymbol{E} \cdot d\log VA + \boldsymbol{V} \cdot d\log w - d\log L, \tag{30}$$

$$d\log VA = \eta \cdot (-d\log \tau) + \boldsymbol{G} \cdot d\log w + \boldsymbol{D} \cdot d\log VA, \tag{31}$$

where $\boldsymbol{E}$, $\boldsymbol{V}$, $\boldsymbol{G}$, and $\boldsymbol{D}$ are matrices that we characterize and discuss later (see equations (35), (36), (44), and (45), respectively). We prove Proposition 2 using these equations. We then apply equation (2) to characterize the welfare change for each worker.

Labor market equilibrium and proof of (30). Let $v_{i|j}$ be the share of value added from industry j that accrues to labor type i (with $\sum_i v_{i|j} = 1$) and, conversely, $e_{j|i}$ be the share of total labor income of type i that stems from industry j (with $\sum_j e_{j|i} = 1$). We start from an accounting identity, that the total wage payments of type i labor equal the sum of wage payments across industries, which can be expressed as:

$$w_i L_i = \sum_j v_{i|j} \cdot VA_j, \tag{32}$$

with summation across $j \in \mathcal{J}_i$. Log-differentiating it yields

$$d\log w_i + d\log L_i = \sum_j e_{j|i} \left(d\log v_{i|j} + d\log VA_j \right). \tag{33}$$

We now argue that changes in the composition of payroll across types in a given industry, $d\log v_{i|j}$, depend fully on the wage changes without direct effects of trade costs or total labor supply. This follows from our assumption on the production function, in which all labor inputs enter via an aggregator F^{VA}. Thus, the optimal composition of labor per unit of value added solves

$$W_j \equiv \min_{L_1^j \ldots L_I^j} \sum_i w_i L_i^j \quad \text{s.t. } F_j^{VA}(L_1^j, \ldots, L_I^j) \geq 1. \tag{34}$$

This problem yields within-industry payroll shares $v_{i|j}(w)$, which are homogeneous functions of degree 0 that depend on wages only and capture patterns of labor substitution within the industry. This problem also yields the value-added cost index W_j which we will use later. Thus,

$$d \log v_{i|j} = \sum_{i'=1}^{I} \frac{\partial \log v_{i|j}}{\partial \log w_{i'}} d \log w_{i'}.$$

Together with (33), this implies (30), with matrix

$$\boldsymbol{E} = \left(e_{j|i}\right)_{i,j} \tag{35}$$

collecting worker exposures to different industries and matrix

$$\boldsymbol{V} = \left(\sum_{j} e_{j|i} \frac{\partial \log v_{i|j}}{\partial \log w_{i'}}\right)_{i,i'} \tag{36}$$

collecting cross-industry averages of labor substitution elasticities between types i and i'. Intuitively, the wage of type i workers increases if the industries in which these workers are employed expand, and falls if either supply of these workers or wages of substitutable workers grow.

Product market equilibrium and proof of (31). To derive equation (31) for value-added changes in each industry, we first solve for the price changes after the shock, similar to Proposition 1 but allowing for wage changes. We then use price and income elasticities, as well as the structure of foreign demand and domestic intermediate demand, to translate the price and consumer income changes into VA changes.

Changes in prices. We first explain how Assumption 4 implies that relative price indices and relative product demand do not change in foreign countries in response to the counterfactual shock. Consider some foreign variety ω belonging to industry j. Since exports to Home are assumed to be a small fraction of ω's worldwide sales, shocks to trade costs with H have negligible effects on the total demand for ω. Likewise, shocks to wages in H have a negligible impact on total demand for ω. Moreover, since imports from Home are a small fraction of absorption abroad, shocks to trade costs and to wages in H have negligible impacts on industry-j consumer price indices in all foreign countries. Thus, the demand for variety ω from consumers outside H remains unchanged after the shocks. These observations have direct implications for factor prices, since factor demand arises from the relative demand for goods: absent changes in relative demand, relative foreign factor prices stay constant.[52]

Turning to the domestic economy, Proposition 1 extends naturally to characterize changes in the industry consumer prices P_{jH} and the prices of domestic varieties, p_{jH}. Log-differentiating the consumer price index (i.e. the CES aggregator of consumer prices across selling countries for a given industry), we have by Roy's identity:

$$d\log P_{jH} = IP_{jc} d\log\tau + (1 - IP_j)\, d\log p_{jH}. \tag{37}$$

By Shephard's lemma and using perfect competition,

$$d\log p_{jH} = (1-\beta_j)\, d\log W_j + \sum_{\ell=1}^{J} \beta_\ell^j d\log P_{jH}. \tag{38}$$

Denote the domestic input requirement (i.e., input-output) matrix by $\boldsymbol{B} = \left(\beta_\ell^j\right)$, and by $\tilde{\boldsymbol{B}} = \left(\mathbb{I}_J - \operatorname{diag}(1 - IP_j)\, \boldsymbol{B}'\right)^{-1}$ its Leontief inverse matrix, such that $\tilde{\boldsymbol{B}}y$ is a weighted sum of variable y in the reference industry j and in all upstream industries in the domestic supply chain of j. Solving the system of (37)–(38) yields

$$d\log p_{jH} = \widetilde{IP}_{jc}^{\text{Int}} d\log\tau + \left(1 - \widetilde{IP}_j^{\text{Int}}\right) d\log\tilde{W}_j \qquad \text{and} \tag{39a}$$

$$d\log P_{jH} = \widetilde{IP}_{jc} d\log\tau + \left(1 - \widetilde{IP}_j\right) d\log\tilde{W}_j, \tag{39b}$$

where $\left\{\widetilde{IP}_{jc}\right\}_{j=1}^{J} = \tilde{\boldsymbol{B}} \cdot \{IP_{jc}\}_{j=1}^{J}$ collects the IO-adjusted shares of imports from c in industry absorption, $\left\{\widetilde{IP}_j^{\text{Int}}\right\} = \boldsymbol{B} \cdot \left\{\widetilde{IP}_j\right\}$ collects the shares of imported inputs in the costs of domestic varieties, and $d\log\tilde{W}_j$ is the average change in the value added cost in the domestic part of the supply chain resulting in j, defined by

$$\left\{\left(1 - \widetilde{IP}_j\right) d\log\tilde{W}_j\right\} = \tilde{\boldsymbol{B}} \cdot \left\{\left(1 - \widetilde{IP}_j\right)(1-\beta_j)\, d\log W_j\right\}. \tag{40}$$

Domestic price changes in (38) imply consumer price changes for domestic varieties in foreign countries: after a bilateral liberalization, prices change by $d\log p_{jH}$ in

[52] More formally, one could consider the effects of a trade shock in a sequence of economies with the share of Home in imports and exports abroad converging to zero, while domestic trade shares do not change along the sequence to match our data. In the limit, the effects of the trade shock on foreign prices become negligible, even though the responses of the demand for Home's varieties and relative goods and factor prices at Home remain non-vanishing.

countries other than c and by $d\log p_{jH} + d\log\tau$ in c. Equation (37), in turn, yields the Laspeyres price index for a domestic consumer with income x, as

$$d\log P_x = \sum_j s_j^x d\log P_{jH} = ImpSh_c^x d\log\tau + \sum_j s_j^x \left(1 - \widetilde{IP}_j\right) d\log\tilde{W}_j. \tag{41}$$

Changes in industry sizes. To characterize the change in industry VA, as required by (30), we first observe that it equals the change in the value of industry output Y_{jH}, i.e. $d\log VA_j = d\log Y_{jH}$. This follows since production functions are Cobb-Douglas in value added and inputs. To characterize changes in domestic output, we start from the product market clearing condition: domestic output can be sold to domestic final and intermediate consumers, or as exports. That is, $Y_{jH} = Y_{jH}^{\text{Final}} + Y_{jH}^{\text{Int}} + Y_{jH}^{\text{Export}}$, where $Y_{jH}^{\text{Int}} = \sum_k Y_{jH}^k$ measures total intermediate sales as a sum across domestic downstream industries k. The change in total sales is thus determined by the shares of different modes of sales at the initial equilibrium and by the changes in each component after the shock.

We use the IO table to measure the composition of different modes of sales. A challenge arises because the IO table does not fully report modes of sales. Specifically, the IO table reports the share of exports *in output* and the share of final consumers and each downstream industry k *in absorption*. Modes of sales can be computed using the proportionality condition (see footnote 26). Specifically, we introduce the intermediate absorption coefficients $\delta_j^k = Y_j^k/\text{Absorption}_j$ which measure the share of industry j's absorption that is used as intermediate inputs to downstream industry k. While β_k^j characterize industry j's suppliers, δ_j^k characterize its buyers. By proportionality, shares δ_j^k can be applied to the domestic sales of *domestic* varieties specifically, i.e. $Y_{jH}^k/\left(Y_{jH}^{\text{Final}} + X_{jH}^{\text{Int}}\right) = \delta_j^k$. Therefore, the share of domestic output that goes to k equals $Y_{jH}^k/Y_{jH} = DomSalesSh_j \cdot \delta_j^k$. Similarly, the share of domestic output that is sold to domestic final consumers is $DomSalesSh_j \cdot (1 - \delta_j) \equiv DFS_j$, where $\delta_j = \sum_k \delta_j^k$ measures the share of intermediate sales in absorption. As a result,

$$d\log VA_j = ExSh_j \cdot d\log Y_{jH}^{\text{Export}} + DFS_j \cdot d\log Y_{jH}^{\text{Final}} + \sum_{k=1}^{J} DomSalesSh_j \delta_j^k \cdot d\log Y_{jH}^k. \tag{42}$$

We now turn to the changes in each component of sales in (42). First, consider exports to some country $c' \neq H$. Since the consumer price for the domestic variety in j changes in country c' by $d\log p_{jH} + \mathbf{1}[c' \in c]\, d\log\tau$, purchases by final and intermediate buyers in c' change by

$$d\log Y_{jH}^{\text{Export},c'} = d\log Y_j^{c'} + (1-\xi_j)\left(d\log p_{jH} + \mathbf{1}[c' \in c]\, d\log\tau - d\log P_{jc'}\right),$$

where $Y_j^{c'}$ is the total spending on all varieties of j by all buyers in c' and $P_{jc'}$ is the industry price index in that country. By Assumption 4, $d\log P_{jc'} = 0$ and $d\log Y_j^{c'} = 0$. Thus, exports to an individual country change by $d\log Y_{jH}^{\text{Export},c'} = (1-\xi_j)\left(d\log p_{jH} + \mathbf{1}[c' \in c]\, d\log\tau\right)$. Aggregating across foreign countries, we have

$$ExSh_j \cdot d\log Y_{jH}^{\text{Export}} = (1-\xi_j)\left(ExSh_j d\log p_{jH} + ExSh_{jc} d\log\tau\right).$$

Second, domestic final sales in (42) are the total of purchases by various consumer groups defined by type i and initial income level x, Y_{jH}^{ix}, and thus

$$d\log Y_{jH}^{\text{Final}} = \sum_{x,i} \mu_{x,i|j} d\log Y_{jH}^{ix},$$

where $\mu_{x,i|j}$ captures the composition of final buyers of industry j by income and labor market type.[53] By the assumption of CES preferences within industries, $d\log Y_{jH}^{ix} = d\log Y_j^{ix} + (1-\xi_j)\left(d\log p_{jH} - d\log P_{jH}\right)$, where Y_j^{ix} measures total spending by the consumer group on industry j varieties, domestic or foreign. By definition of income and price elasticities,

$$\begin{aligned} d\log Y_j^{ix} &= \psi_{jx} d\log w_i + \sum_{k=1}^{J} \varepsilon_{xjk} d\log P_{kH} \\ &= d\log w_i + (\psi_{jx}-1)\left(d\log w_i - d\log P_x\right) + \sum_{k=1}^{J} \varepsilon_{xjk}\left(d\log P_{kH} - d\log P_x\right). \end{aligned} \tag{43}$$

Here in the first line we equated expenditure and wage changes using the assumption that each consumer spends a constant multiple of their income. The second line used $\psi_{jx} + \sum_k \varepsilon_{xjk} = 1$, which follows because increasing income and prices proportionately does not change expenditure shares.

Finally, for intermediate sales in (42) we use the Cobb-Douglas assumption again.

[53] Because labor market data (e.g. industries) are not available for the consumers in the CEX, we do not observe $\mu_{x,i|j}$ directly. However, with identical non-homothetic preferences the industry does not matter for consumption baskets conditionally on income. Thus, we measure $\mu_{x,i|j}$ as the product of the share of income decile x in the CEX expenditures on industry j, $\mu_{x|j}$, and the share of type-i workers in the total payroll of workers in income decile x in the ACS, $v_{i|x}$.

The share of spending by industry k on *all* varieties of j is fixed, so the change in expenditures equals the change in k's value added: $d\log Y_j^k = d\log Y_{kH} = d\log VA_k$. But substitution between domestic and foreign varieties implies that domestic sales of j to k change by

$$d\log Y_{jH}^k = d\log VA_k + (1-\xi_j)\left(d\log p_{jH} - d\log P_{jH}\right).$$

Equation (31) now follows by plugging price changes derived above into the expressions for the changes in exports, domestic final sales, and domestic intermediate sales, plugging those in turn into (42), and rearranging terms. Specifically, all terms that enter with $-d\log\tau$ are collected in the η vector, yielding (5). The terms with $d\log VA_k$, arising from intermediate demand only, define the $\boldsymbol{D}$ matrix:

$$\boldsymbol{D} = \left(DomSalesSh_j \cdot \delta_j^k\right)_{j,k}. \tag{44}$$

Pre-multiplication by its Leontief inverse $\tilde{\boldsymbol{D}} = (\mathbb{I}_J - \boldsymbol{D})^{-1}$ is interpreted as the IO adjustment that accounts for the propagation of shocks from downstream industries up through changes in domestic intermediate demand. For example, the elements of $\tilde{\boldsymbol{D}} \cdot ExSh$ are the shares of domestic output that is exported either directly or indirectly (by selling to domestic downstream industries that export).

Finally, collecting the terms related to wage changes defines the $\boldsymbol{G}$ matrix, as follows:

$$\begin{aligned}(\boldsymbol{G}w)_j \equiv{}& (1-\xi_j)\left(ExSh_j + DomSalesSh_j IP_j\right)\left(1-\widetilde{IP}_j^{\text{Int}}\right) d\log\tilde{W}_j \\ &+ DFS_j \sum_{x,i} \mu_{x,i|j}\left[d\log w_i + (\psi_{jx}-1)\left(d\log w_i - \sum_{\ell=1}^{J} s_\ell^x\left(1-\widetilde{IP}_\ell\right) d\log\tilde{W}_\ell\right)\right. \\ &\left. + \sum_{k=1}^{J}\varepsilon_{xjk}\left(\left(1-\widetilde{IP}_k\right) d\log\tilde{W}_k - \sum_{\ell=1}^{J} s_\ell^x\left(1-\widetilde{IP}_\ell\right) d\log\tilde{W}_\ell\right)\right],\end{aligned} \tag{45}$$

with $d\log\tilde{W}_k$ linearly related to $d\log w$ via (40). The first line of (45) captures the loss of competitiveness of domestic varieties (relative to foreign varieties in the same industry) in both domestic and foreign markets when domestic wages grow. The second line captures the change in domestic final demand when consumer incomes change, as well as income effects from changing both consumer income and inflation. The third line captures the substitution effects driven by domestic wage changes.

Proof of Proposition 2. Letting $\tilde{\boldsymbol{V}} = (\mathbb{I}_I - \boldsymbol{V})^{-1}$, (31) implies

$$d\log VA = \tilde{\boldsymbol{D}}\left(-\eta \cdot d\log \tau + \boldsymbol{G} \cdot d\log w\right).$$

and, from (30),

$$\begin{aligned} d\log w &= \tilde{\boldsymbol{V}}\left(\boldsymbol{E} \cdot d\log VA - d\log L\right) \\ &= \tilde{\boldsymbol{V}}\left(-\boldsymbol{E}\tilde{\boldsymbol{D}}\eta \cdot d\log \tau + \boldsymbol{E}\tilde{\boldsymbol{D}}\boldsymbol{G} \cdot d\log w - d\log L\right) \\ &= \tilde{\boldsymbol{G}}\left(\boldsymbol{E}\tilde{\boldsymbol{D}}\eta \cdot (-d\log \tau) - d\log L\right). \end{aligned} \tag{46}$$

Here

$$\tilde{\boldsymbol{G}} = \left(\mathbb{I}_I - \tilde{\boldsymbol{V}}\boldsymbol{E}\tilde{\boldsymbol{D}}\boldsymbol{G}\right)^{-1}\tilde{\boldsymbol{V}} \tag{47}$$

captures the GE response of factor prices to an exogenous decline in factor supply and therefore can be interpreted as the (negative of the) inverse labor demand elasticity matrix. With $d\log L = 0$, equation (46) reduces to (4), establishing Proposition 2. We note that the $\tilde{\boldsymbol{G}}$ matrix generalizes the macro elasticity of factor substitution that Oberfield and Raval (2020) derived for a closed economy with homothetic preferences and only two factors.

Welfare effects in general equilibrium. Given wage changes characterized by Proposition 2, we can obtain welfare changes for workers of type i with initial income x. We have:

$$\begin{aligned} d\log \mathcal{W}_{ix} = {} & (d\log w_i - d\log \bar{w}) - ImpSh_c^x d\log \tau \\ & + ImpSh^x d\log \bar{w} - \sum_j s_j^x \left(1 - \widetilde{IP}_j\right)\left(d\log \tilde{W}_j - d\log \bar{w}\right), \end{aligned} \tag{48}$$

where $ImpSh^x$ is the total share of imports from all foreign countries in the consumption baskets. The first term here is the earnings channel, capturing the gap between wage growth of type i relative to the economy overall, with the latter defined as $d\log \bar{w} = \sum_i v_i d\log w_i$ where v_i denotes the initial payroll share of type i in the economy. The second term is the partial equilibrium effect on prices from Proposition 1. The third term is a terms-of-trade adjustment: if domestic wages grow on average, all imports become relatively cheaper. The final term captures the idea that, if some group of consumers tends to buy goods from industries where wages grow relatively more after the shock (directly or in their supply chains), this group will benefit less. For instance, if college graduates buy goods produced with skilled labor, then increases in the skill premium generate an offsetting effect on inequality via the expenditure channel; we label this mechanism a "segregation effect."[54]

Proof of equation (48). We have:

$$\begin{aligned}
d\log \mathcal{W}_{ix} &= d\log w_i - d\log P_x \\
&= d\log w_i - ImpSh_c^x d\log\tau - \sum_j s_j^x \left(1-\widetilde{IP}_j\right) d\log \tilde{W}_j \\
&= d\log w_i - ImpSh_c^x d\log\tau - \left(1-ImpSh^x\right) d\log\bar{w} \\
&\quad - \sum_j s_j^x \left(1-\widetilde{IP}_j\right)\left(d\log \tilde{W}_j - d\log\bar{w}\right) \\
&= \left(d\log w_i - d\log\bar{w}\right) - ImpSh_c^x d\log\tau + ImpSh^x d\log\bar{w} \\
&\quad - \sum_j s_j^x \left(1-\widetilde{IP}_j\right)\left(d\log \tilde{W}_j - d\log\bar{w}\right).
\end{aligned}$$

Here the first line used the Roy identity, the second equation used (41), the third line used the definition of $ImpSh^x = \sum_j s_j^x \widetilde{IP}_j$, and the last line rearranged the terms.

C.2 Special Cases

We now consider special cases of Proposition 2. We first characterize the substitution effects in η (see equation (5)) under the nested non-homothetic CES demand system our calibrations employ. We then discuss how the general formulas simplify in our two calibrations. In particular, for the worker-level calibration we explain why wages in less traded sectors are more sensitive to labor demand shocks. For the calibration across education groups, we show how the local elasticities of labor substitution in all industries can be summarized with a single "macro" elasticity.

Substitution effects with NNHCES. As described in Section 5.2, we discipline substitution effects by NNHCES preferences, analogous to those of Appendix B.2 but using a different tier structure. With the assumption on CES aggregation across varieties in each industry, we require the income elasticities parameters φ_j to be the same for those varieties (see footnote 26). Relative to Appendix B.2, we allow for

[54] For the analyses of this effect, see Clemens et al. (2018) and Wilmers (2017), as well as our early draft (Borusyak and Jaravel 2018). In Figures 6B(ii) and 7B we combine the last two terms into the "price effects of wages" total.

additional flexibility by adding goods and services as a separate upper tier, indexed by r:

$$\mathcal{U} = \left(\sum_{r=\text{Goods, Services}} Q_r^{(\rho-1)/\rho} \right)^{\rho/(\rho-1)},$$
$$Q_r = \left(\sum_{j \in r} \left(\mathcal{U}^{\varphi_j(\varepsilon_r - 1)} \right)^{1/\varepsilon_r} Q_j^{(\varepsilon_r-1)/\varepsilon_r} \right)^{\varepsilon_r/(\varepsilon_r-1)}, \tag{49}$$

where Q_j is a CES aggregator across country varieties: $Q_j = \left(\sum_c a_{jc}^{1/\xi_j} Q_{jc}^{(\xi_j-1)/\xi_j} \right)^{\xi_j/(\xi_j-1)}$.

We now show that, for industry j that belongs to sector r, substitution effects in (5) satisfy:

$$\sum_{k=1}^{J} \varepsilon_{xjk} \left(\widetilde{IP}_{kc} - ImpSh_c^x \right) = (1-\varepsilon_r) \left(\widetilde{IP}_{jc} - ImpSh_{rc}^x \right)$$
$$+ (1-\rho) \left(ImpSh_{rc}^x - ImpSh_c^x \right), \tag{50}$$

where $ImpSh_{rc}^x = \sum_{k \in r} s_{k|r}^x \widetilde{IP}_{kc}$, and similarly for the substitution effects in the last line of (45):

$$\sum_{k=1}^{J} \varepsilon_{xjk} \left(\left(1 - \widetilde{IP}_k\right) d\log \tilde{W}_k - \sum_{\ell=1}^{J} s_\ell^x \left(1 - \widetilde{IP}_\ell\right) d\log \tilde{W}_\ell \right) =$$
$$(1-\varepsilon_r) \left(\left(1 - \widetilde{IP}_j\right) d\log \tilde{W}_j - \sum_{k \in r} s_{k|r}^x \left(1 - \widetilde{IP}_k\right) d\log \tilde{W}_k \right)$$
$$+ (1-\rho) \left(\sum_{k \in r} s_{k|r}^x \left(1 - \widetilde{IP}_k\right) d\log \tilde{W}_k - \sum_{k=1}^{J} s_k^x \left(1 - \widetilde{IP}_k\right) d\log \tilde{W}_k \right) \tag{51}$$

Proof of (50)–(51). Analogously to (22), after a set of income and price changes, changes in expenditures of consumer i with income x on the aggregate good of industry j within sector r are given by

$$d\log Y_j^{ix} = d\log w_i + (\psi_{xj} - 1) \left(d\log w_i - \sum_{k=1}^{J} s_k^i d\log P_{kH} \right)$$
$$+ (1-\varepsilon_r) \left(d\log P_{jH} - \sum_{k \in r} s_{k|r}^i d\log P_{kH} \right) + (1-\rho) \left(\sum_{k \in r} s_{k|r}^i d\log P_{kH} - \sum_{k=1}^{J} s_k^i d\log P_{kH} \right). \tag{52}$$

We use this expression instead of the more general (43) and follow the remaining part of the proof of Proposition 2, plugging in prices from (39) and isolating the terms with $d \log \tau$ and $d \log \tilde{W}_k$. Then the substitution effects from the second line of (52) yield (50)–(51).

Worker-level calibration. We now consider how Proposition 2 applies to our worker-level calibration, which assumes no mobility of workers across industries.

In this setting, a labor type directly maps to an industry and $I = J$. Absent labor substitution, $\boldsymbol{V} = 0_{I \times I}$, and each labor type is exposed just to its own industry, $\boldsymbol{E} = \mathbb{I}_I$. By equation (30), wages are proportional to industry value added, $d \log w = d \log VA - d \log L$.

We now show that in industries with lower trade shares (i.e., export shares and import penetration rates), wages are more responsive to shifts in labor demand, compared to more traded industries. We prove this result in a restricted model, in which only the export and import competition effects arise, while intermediate inputs, income, and substitution channels are shut down. We find in our calibration of Section 5.3 that this result holds qualitatively even when all channels are operative.

Formally, suppose $\tilde{\boldsymbol{D}} = \mathbb{I}_J$, $\psi_{xj} \equiv 1$, and $\varepsilon_{xjk} \equiv 0$ and consider a set of shifts to labor demand $d \log L_j^D$ (or, equivalently, a similar reduction in labor supply). Then we prove that, for $d \log w = \tilde{\boldsymbol{G}} \cdot d \log L^D$,

$$d \log w_j = \frac{d \log L_j^D}{1 + (\xi_j - 1) T_j} + \frac{DomSalesSh_j}{\zeta_2 (1 + (\xi_j - 1) T_j)} \cdot \left(\sum_{k=1}^{J} \frac{e_k d \log L_k^D}{1 + (\xi_k - 1) T_k} \right), \qquad (53)$$

where $T_j = ExSh_j + IP_j \cdot DomSalesSh_j$, $\zeta_2 = 1 - \sum_j \frac{e_j DomSalesSh_j}{1 + (\xi_j - 1) T_j} \in (0, 1)$, and e_j is the payroll share of industry j in the economy.

The first term in (53) shows that wages in more traded industries are less responsive to shifts in labor demand in their own industry, via the T_j term which increases in both the export share and the import penetration rate. The second term shows that they are also less sensitive to the economy average shift in labor demand, via both higher T_j and lower $DomSalesSh_j$.

Proof of equation (53). Under the above conditions, $DFS_j = DomSalesSh_j$. In the absence of non-homotheticities, $\sum_x \mu_{xi|j} = e_i$ for any j. Thus, equation (45) simplifies to

$$(\boldsymbol{G}\cdot d\log w)_j = (1-\xi_j)\,T_j d\log w_j + DomSalesSh_j \cdot \sum_i e_i d\log w_i.$$

In matrix form,

$$\boldsymbol{G} = -\operatorname{diag}\left[(\xi_j - 1)\,T_j\right] + DomSalesSh \cdot e'.$$

By the Sherman-Morrison formula in linear algebra, its Leontief inverse equals

$$\tilde{\boldsymbol{G}} = \operatorname{diag}\left[1+(\xi_j-1)\,T_j\right]^{-1} + \frac{\operatorname{diag}\left[1+(\xi_j-1)\,T_j\right]^{-1} DomSalesSh \cdot e' \operatorname{diag}\left[1+(\xi_j-1)\,T_j\right]^{-1}}{1-e'\operatorname{diag}\left[1+(\xi_j-1)\,T_j\right]^{-1} DomSalesSh}.$$

Expanding these terms, $\tilde{\boldsymbol{G}}\cdot d\log L^D$ satisfies (53).

Calibration across education groups. We next consider the setting with full labor mobility across industries and two labor types (e.g. education groups), which we denote H and L (high and low skilled). We show that labor substitution elasticities σ_j of all industries enter the $\boldsymbol{V}$ matrix (and therefore $\tilde{\boldsymbol{G}}$) only through a scalar parameter σ_{macro}, which we refer to as the macro elasticity of labor substitution. Specifically,

$$\boldsymbol{V} = (\sigma_{\text{macro}} - 1)\begin{pmatrix} -v_L & v_L \\ v_H & -v_H \end{pmatrix}, \tag{54}$$

where[55]

$$\sigma_{\text{macro}} - 1 = \sum_j e_j \frac{v_{H|j} v_{L|j}}{v_H v_L}\left(\sigma_j - 1\right). \tag{55}$$

.

We then use this result to show that the skill group that is initially specialized in industries that will grow faster after the shock will experience a higher wage growth:

$$d\log\frac{w_H}{w_L} = \frac{1}{\sigma_{\text{macro}}}\left(\sum_j v_{H|j} d\log VA_j - \sum_j v_{L|j} d\log VA_j\right). \tag{56}$$

Proof of (54)–(56). By definition of the labor substitution elasticity in j,

$$d\log\frac{v_{H|j}}{v_{L|j}} = (1-\sigma_j)\,d\log\frac{w_H}{w_L}.$$

[55] We note that $\sigma_{\text{macro}} - 1$ is generally not a weighted average of $\sigma_j - 1$: the sum of weights, $\sum_j e_j \frac{v_{H|j} v_{L|j}}{v_H v_L}$, is smaller than one unless all industries have the same skill composition.

Since $v_{L|j} = 1 - v_{H|j}$ and using $d \log \frac{z}{1-z} = \frac{1}{1-z} d \log z$, we obtain:

$$d \log v_{H|j} = v_{L|j} \left(1 - \sigma_j\right) d \log \frac{w_H}{w_L}. \tag{57}$$

Thus,

$$\boldsymbol{V}_{HH} = \sum_j e_{j|H} v_{L|j} \left(1 - \sigma_j\right) = v_L \sum_j e_j \frac{v_{H|j}}{v_H} \frac{v_{L|j}}{v_L} \left(1 - \sigma_j\right) = -v_L \left(\sigma_{\text{macro}} - 1\right),$$

where the first equality follows by definition of and (57), the second one rewrites $e_{j|H} = \frac{e_j v_{H|j}}{v_H}$, and the last uses the definition of σ_{macro}. The other elements of $\boldsymbol{V}$ are obtained analogously, yielding (54). Plugging in (54) into (30) for $d \log w_H$ and $d \log w_L$ and taking the difference, one obtains (56).

C.3 Proof of Equation (7)

We consider a sequence of $d \log \mathcal{W} = W dt$ for a fixed random variable W and $dt \to 0$. Then:

$$\begin{aligned}
\text{SD}\left(\log X + W dt\right) &= \sqrt{\text{Var}\left[\log X\right] + 2\text{Cov}\left[\log X, W\right] dt + \text{Var}\left[W\right] dt^2} \\
&= \text{SD}\left(\log X\right) \cdot \sqrt{1 + 2\frac{\text{Cov}\left[\log X, W\right]}{\text{Var}\left[\log X\right]} dt + o(dt)} \\
&= \text{SD}\left(\log X\right) \left(1 + \frac{\text{Cov}\left[\log X, W\right]}{\text{Var}\left[\log X\right]} \cdot dt\right) + o(dt) \\
&= \text{SD}\left(\log X\right) + \text{Corr}\left[\log X, W\right] \cdot \text{SD}\left(W\right) dt + o(dt) \\
&= \text{SD}\left(\log X\right) + \text{Corr}\left[\log X, d \log \mathcal{W}\right] \cdot \text{SD}\left(d \log \mathcal{W}\right) + o(dt).
\end{aligned}$$

References

Borusyak, K., and X. Jaravel. 2018. "The Distributional Effects of Trade: Theory and Evidence from the United States." *Mimeo.*

Clemens, J., J. D. Gottlieb, M. Olsen, and H. David. 2018. "The Spillover Effects of Top Income Inequality." *Mimeo.*

Comin, D., D. Lashkari, and M. Mestieri. 2021. "Structural Change with Long-run Income and Price Effects." *Econometrica* 89:311–374.

Fajgelbaum, P., and A. K. Khandelwal. 2016. "Measuring the Unequal Gains from Trade." *Quarterly Journal of Economics* 131:1113–1180.

Mas-Colell, A., M. D. Whinston, and J. R. Green. 1995. *Microeconomic Theory.* Vol. 1. New York: Oxford University Press.

Oberfield, E., and D. Raval. 2020. "Micro Data and Macro Technology." *Mimeo.*

Wilmers, N. 2017. "Does Consumer Demand Reproduce Inequality? High-income Consumers, Vertical Differentiation and the Wage Structure." *Americal Journal of Sociology* 123:178–231.

Supplementary Material (Not For Publication) for "The Distributional Effects of Trade: Theory and Evidence from the United States"

Kirill Borusyak, UCL
Xavier Jaravel, LSE and CEPR

June 2021

S Supplementary Materials

S.1 Extensions to Proposition 1

We now discuss how Proposition 1 can be extended to several important deviations from Assumptions 1–3.

GVCs. Two adjustments to Proposition 1 are required when the Home economy is involved in global value chains. First, some imported products are subject to iceberg trade costs multiple times, and the resulting price increases accumulate. Equation (9) shows how import shares need to be measured in that case: every imported intermediate input should be counted as many times as it crosses the border from c into H. Second, with GVCs, bilateral shocks to iceberg costs have additional price effects, as imported products may contain exported domestic products; again, equation (9) accommodates that case.

We note that our analysis, both in the baseline case and with GVCs, considers shocks to trade costs. However, it can easily be applied to studying productivity shocks in foreign countries. If, say, labor productivity increases in c, the relevant measure of the import share will be based on the fraction of value added in the costs of the final output that originated from c, regardless of the route it took to arrive in H.

Markups. Suppose firms are monopolists in the markets for each variety they sell but price takers in the markets for factors and intermediate inputs they buy, and that the set of available varieties does not change after the trade shock. If firms charge constant markups μ_ω, which are not affected by the trade shocks (but could differ across firms, as with monopolistic competition in multi-sectoral models), then Proposition 1 continues to hold.[56]

In contrast, Proposition 1 has to be adjusted if markups are not constant and respond endogenously to the counterfactual shocks. For standard demand systems with endogenous markups, the pass-through of marginal cost shocks into prices is incomplete (Arkolakis and Morlacco 2017). This could affect the expenditure channel if products purchased by different consumer groups systematically differ in their pass-through rates or in the length of supply chains, as incomplete pass-through at multiple stages of domestic production generates more attenuation. Another implication of endogenous markups for the expenditure channel is that, following a trade liberalization, firms might change their markups even absent a marginal cost change, because demand for each variety shifts. In particular, domestic prices could change through reduced markups due to increased foreign competition.

Endogenous Variety. While in our benchmark model the set of products in the consumption basket is assumed exogenous, Proposition 1 continues to hold in several classes of models where product variety changes in response to a trade liberalization. First, suppose some foreign varieties are not purchased by a consumer at the initial equilibrium because their prices are above reservation prices, but these varieties start to be consumed (with small shares) when prices fall after the trade liberalization. This could occur, for instance, in models with an Armington product space but non-CES

[56]From the measurement perspective, however, this case is more complex than perfect competition because the cost shares β_ω^ℓ no longer equal the revenue shares of the same inputs (see Tintelnot et al. 2020).

demand that features finite reservation prices. In this setting, the implied welfare effect of the new products is second order. Indeed, the envelope theorem underlying Proposition 1 continues to apply when expenditure shares on some products are null, as long as prices change continuously.

Second, a similar logic applies to Eaton and Kortum (2002) type models, with or without the Frechet distribution of productivity. Although in this model some imported products enter the domestic market with non-negligible quantities, they replace domestic varieties which are only marginally more expensive. Thus, by grouping perfectly substitutable varieties across countries, one can view consumer prices and expenditure shares as continuous in trade costs. The envelope theorem then continues to apply, even though prices and shares jump at the producing country-variety level, and this extensive margin response does not yield first-order welfare effects.

Finally, even if entry of foreign varieties generates first-order welfare gains, as in Melitz (2003) and Chaney (2008), it may be accompanied by exit of varieties produced domestically or in third countries. These effects exactly offset each other in some cases, as we show next.

Increasing Returns and Selection Effects. In this final extension, we show how Proposition 1 extends to a setting with increasing returns to scale and endogenous variety stemming from selection into exporting. We study the Generalized Melitz-Pareto model of Kucheryavyy et al. (2020), which enriches the standard Melitz-Pareto model of Chaney (2008) and Melitz (2003) by decoupling the scale and trade elasticities. This model is isomorphic to the Armington model with external scale economies and a generalized Krugman model (Kucheryavyy et al. 2020). As in Proposition 1, we consider partial equilibrium changes. We focus on selection and scale effects but abstract from changes in the numbers of potential entrants or in market size. This is without loss in single-sector settings; with multiple sectors, home market effects need to be accounted for but their welfare consequences were shown to be small (Costinot and Rodríguez-Clare 2015).

We focus on a particular industry, suppressing the j index. We assume that there is a continuum of varieties with (homothetic) nested CES preferences, in which the inner nest aggregates within each country with elasticity ς and the outer nest aggregates across countries with elasticity $\epsilon \leq \varsigma$, with $\epsilon = \varsigma$ corresponding to the standard Melitz model. In country c, an exogenous number N_c of potential entrants

ω draw productivity $z(\omega)$ from the Pareto distribution with the shape parameter $\kappa > \varsigma - 1$ and scale parameter $z_{\min,c}$ and choose which markets to sell in and how much to produce using labor as the single factor. For a firm born in c, the marginal cost of selling in H is $m_c \tau_{cH}/z(\omega)$, where m_c is the unit cost of production. Besides the iceberg cost τ_{cH}, exporting involves a fixed cost F_{cH} (expressed in monetary terms; in partial equilibrium, it is irrelevant whether this fixed cost is paid in the exporting or importing county's labor).

We now show that the industry consumer price index in the Home country, P_H, following a set of changes $d \log \tau_{cH}$ in iceberg trade costs from various countries c to H satisfies[57]

$$d \log P_H = \sum_c IP_c d \log \tau_{cH}, \tag{S1}$$

where IP_c is the expenditure share in H on goods imported from c within the industry. Aggregating price index changes across industries and focusing on a change in the costs of importing from a single country or group of countries, one obtains Proposition 1.

Proof of Equation (S1). With nested CES demand, the optimal gross markup is $\frac{\varsigma}{\varsigma-1}$. The consumer price index is given by:

$$P_H^{1-\epsilon} = \sum_c P_{Hc}^{1-\epsilon}, \tag{S2}$$

where the price aggregate of all varieties imported from c, P_{Hc}, is given by

$$\begin{aligned} P_{Hc}^{1-\varsigma} &= N_c \int_{z_{cH}}^{\infty} \left(\frac{\varsigma}{\varsigma-1}\right)^{1-\varsigma} m_c^{1-\varsigma} \tau_{cH}^{1-\varsigma} z^{\varsigma-1} \cdot \frac{\kappa z^{-\kappa-1}}{z_{\min,c}^{-\kappa}} dz \\ &= \left(\frac{\varsigma}{\varsigma-1}\right)^{1-\varsigma} \frac{\kappa z_{c,\min}^{\kappa}}{\kappa-(\varsigma-1)} \cdot N_c m_c^{1-\varsigma} \tau_{cH}^{1-\varsigma} z_{cH}^{-\kappa+(\varsigma-1)}, \end{aligned} \tag{S3}$$

with z_{cH} denoting the productivity cutoff for exporting from c to H (or choosing to sell domestically in the case of $c = H$).

Log-differentiation of P_{Hc}, holding m_c and N_c fixed but allowing z_{cH} to respond, yields

$$d \log P_{Hc} = d \log \tau_{cH} + \frac{\kappa - (\varsigma - 1)}{\varsigma - 1} d \log z_{cH}.$$

Intuitively, when trade costs decrease, consumer prices at home fall at the intensive

[57] Summation here includes $c = H$, although $d \log \tau_{HH} = 0$ for shocks to international trade.

margin, with complete pass-through, and there is a correction for the number of exported varieties through love of variety.

Log-differentiating (S2) further yields the change in the aggregate domestic price index:

$$d\log P_H = \sum_c IP_c d\log P_{Hc} = \sum_c IP_c \left(d\log \tau_{cH} + \frac{\kappa - (\varsigma - 1)}{\varsigma - 1} d\log z_{cH} \right). \qquad \text{(S4)}$$

To solve for changes in exporting cutoffs across countries, we consider the indifference condition for the decision to sell in H:

$$F_{cH} = Y_H \left(\frac{P_{Hc}}{P_H} \right)^{1-\epsilon} \left(\frac{\frac{\varsigma}{\varsigma - 1} m_c \tau_{cH} / z_{cH}}{P_{Hc}} \right)^{1-\varsigma}.$$

With marginal costs m_c, fixed costs F_{cH}, and total market size Y_H held fixed, log-differentiating the indifference condition implies

$$\begin{aligned} 0 &= (1 - \epsilon)\left(d\log P_{Hc} - d\log P_H \right) + (1 - \varsigma)\left(d\log \tau_{cH} - d\log z_{cH} - d\log P_{Hc} \right) \\ &= (1 - \epsilon) \left(d\log \tau_{cH} + \frac{\kappa - (\varsigma - 1)}{\varsigma - 1} d\log z_{cH} - d\log P_H \right) + \kappa d\log z_{cH}, \end{aligned}$$

and therefore

$$d\log z_{cH} = \zeta_1 \left(d\log \tau_{cH} - d\log P_H \right), \qquad \text{for } \zeta_1 = \left(1 + \frac{\kappa}{\epsilon - 1} - \frac{\kappa}{\varsigma - 1} \right)^{-1} > 0. \quad \text{(S5)}$$

Intuitively, less productive firms find it profitable to sell in H when trade costs fall, but the opposite happens when the domestic market becomes more competitive as measured by a lower consumer price index. Plugging (S5) into (S4) yields

$$\begin{aligned} d\log P_H &= \sum_c IP_c \left(d\log \tau_{cH} - \zeta_1 \frac{\kappa - (\varsigma - 1)}{\varsigma - 1} \left(d\log \tau_{cH} - d\log P_H \right) \right) \\ &= \left(1 + \zeta_1 \frac{\kappa - (\varsigma - 1)}{\varsigma - 1} \right) \sum_c IP_c d\log \tau_{cH} - \zeta_1 \frac{\kappa - (\varsigma - 1)}{\varsigma - 1} d\log P_H \\ &= \sum_c IP_c d\log \tau_{cH}. \end{aligned}$$

Thus, the response of the domestic price index is governed by import shares in the same way as without selection forces. This happens because two effects offset each other exactly. Lower trade costs induce entry of new varieties from countries where trade costs are falling. At the same time, higher competition pushes less productive firms from all countries out of the H market.

S.2 Data Replication Appendix

S.2.1 CEX-IO Data

Overview. To measure import shares by consumer group at the industry level, we merge consumption data from the Consumer Expenditure Survey (CEX) to the import shares measured using the U.S. Input-Output table. We focus on the year 2007, the most recent year for which the detailed IO table is available, although we check robustness to other years with more aggregated data.

The CEX is a survey by the U.S. Bureau of Labor Statistics that measures detailed expenditures on all goods and services for a representative panel of households. We use the Integrated survey of the CEX, which combines the complete coverage of the Interview survey with the high resolution of the Diary survey on a subset of the most frequently purchased items. These data include 668 detailed spending categories, while recording household characteristics, such as income and education. We pool data from 2006–2008 to increase sample size.

The IO table from the Bureau of Economic Analysis (BEA), in turn, allows us to measure direct and indirect import shares by industry, with 389 industries in total. BEA data are the most detailed available accounts of the entire U.S. economy. For each industry we compute import penetration as the fraction of imports in absorption (defined as output plus imports minus exports). There are two advantages of using the BEA data to measure import penetration: trade in services is accounted for and trade flows are measured from the same data as domestic output, which improves consistency. Then we build the input requirement matrix, which measures the composition of suppliers for each buying industry. We use it to construct the share of indirect imports (imports of intermediate inputs) in domestic production. Combining direct and indirect imports, we obtain the total import share in absorption of each industry.

We pay special attention to the "distribution margins," which refer to the costs of retailing, wholesaling, and transportation and by definition have a low import share. For example, when consumers buy apparel, much of their spending is effectively devoted to distribution margins. The IO table reports that imported final products constitute 84% of total absorption in the apparel industry; when accounting for domestic distribution costs, the import share shrinks to only 36%. We combine "producer-value" and "purchaser-value" IO tables to implement this adjustment.

We use additional tabulations from the U.S. Census Bureau to compute import shares for specific trading partners: China, NAFTA countries (Mexico and Canada), and 34 developed economies (OECD members as of 2017, excluding NAFTA, plus Taiwan and Singapore). Specifically, we compute the shares of these countries in the total 2007 U.S. imports in each industry. We then distribute the overall import penetration reported in the IO table across trade partners using these shares,[58] and combine direct and indirect imports from a given trade partner using the input requirement matrix.

Finally, we match CEX spending categories to 170 final industries in the IO table by building a manual concordance.[59] We use personal final consumption from the IO table as a measure of total spending in the industry and decompose it by income and education groups using CEX-based shares. This approach parallels Lebow and Rudd (2003) who show that reweighting the CEX using BEA spending shares yields more accurate inflation estimates, correcting non-classical measurement error in the CEX (e.g., Garner et al. 2009).[60]

Column 1 of Table S1 presents summary statistics for our linked dataset with 170 final industries. We classify manufacturing, agriculture, and mining into goods and all other industries into services. For some analyses, we further classify goods and services into 24 and 15 subsectors, respectively, listed in Table S2. We next provide details on the CEX and the IO table.

Consumer Expenditure Survey. The CEX is a stratified household survey conducted by the U.S. Bureau of Labor Statistics that measures the universe of personal spending by households with over 600 detailed product categories. The CEX consists

[58]These data were made available by Schott (2008) and converted into NAICS industry codes by Pierce and Schott (2012). Trade flow statistics are available only for trade in goods. We therefore assign zero direct imports from specific trading partners in service industries. This does not constitute an important limitation for China and Mexico. For instance, China constitutes less than 3% of total U.S. imports of services according to the BEA International Services tables for 2007. This limitation is likely to be more important when considering trade with developed economies.

[59]We note that our analysis is fully consistent with the IO structure of the economy, such that the same industry may have both final and intermediate sales. By final industries we mean industries with non-zero consumption in the CEX, without a need to classify industries into different types.

[60]Measurement error in the CEX does not create biases for our results as long as it has the multiplicative structure proposed and justified by Aguiar and Bils (2015): there may be industry- and consumer group-specific biases but no interactions between them. Industry-specific biases are corrected by the BEA weights, while consumer-group-specific biases only result in a re-scaling of consumption across groups without systematic effects on the expenditure composition of each group.

of two separate parts, the interview and diary surveys, which we use in combination. Quarterly interviews cover the complete range of expenditures, whereas diaries focus on some categories, such as food and clothing, in much greater detail. The interview panel includes around 6,900 households per quarter, each surveyed for four consecutive quarters. Diaries are collected for roughly the same number of distinct households per year but capture only two weeks of consumption. We select categories of spending (UCC) from both surveys according to the Integrated Stub file provided by the BLS, so that they cover all categories without double-counting.

The key advantage of the CEX is that consumption structure can be measured separately for different groups of households. We split households by bins of household income before tax, converted to the 2007 prices using the U.S. Consumer Price Index.[61]. As a measure of income, we use variable `FINCBTXM` in the interview survey and `FINCBEFX` in the diary survey. Eleven income bins are defined by the following cutoffs (in $000): 10, 20, 30, 40, 50, 60, 75, 90, 110, and 150. For the analysis in Section 5, we also split the households into deciles of earnings, with thresholds, in $000, of around 13.5, 21, 28.8, 37.3, 47.0, 58.7, 73.0, 92.5, and 127.4. We further split panelists by education of the household's reference person answering the interview (variable `EDUC_REF`), defining college education as bachelor's degree or higher. For Figure S5, we also the following variables: the mean age of the household heads (`age_ref` and `age2`), Census region (`region`), home ownership (`cutenure`), and family size (`fam_size`).

To increase the sample size, we combine data from 2006–2008. We drop all households with reported income below $5,000 because of concerns about misreporting and temporary unemployment. Our final interview sample includes 87,238 household-quarters with average annualized spending of $34,976 (excluding diary categories), while the diary sample has 31,727 household-weeks spending $12,554 per household per year.

Expenditure on housing services requires special treatment. The range of CEX spending categories includes rents and mortgage interest, but not the mortgage principal payments. However, an addendum section of the interview survey provides information on the self-reported rental value of owned property. In our static setup that is the closest analog to annual expenditures on housing for home-owners, so we

[61] Source: FRED database, https://fred.stlouisfed.org/series/CPALTT01USA661S.

add imputed rents to the set of UCC we consider. Aguiar and Bils (2015) follow a similar approach.

We build a manual concordance from 668 CEX consumption categories into 170 IO industries. We thank James O'Brien for providing us with the concordance between CEX interview categories and the 2012 version of NAICS from Levinson and O'Brien (2017). We use this concordance, converted into 2007 IO codes, as a starting point. We manually extend it to diary categories as well as missing interview ones. The concordance is many-to-one, with a few exceptions where we allocate CEX consumption by each group equally across the corresponding IO codes. In most cases, our concordance is consistent with, but much finer than, the concordance from CEX to NIPA personal consumption expenditure categories provided by the BLS and used by Buera et al. (2018) and Jaimovich et al. (2015), among others.

BEA Input-Output Table. We use the most detailed IO table for the U.S., which is available in 2007. While BEA publishes annual tables with 71 relatively coarse three-digit industries (which we leverage in Figure S6), the 2007 one is disaggregated into 389 six-digit industries. These industries are groups of six-digit NAICS codes: while NAICS includes 581 goods and 565 service industries, the IO classification includes 258 and 122, respectively, plus 9 special industries such as government and non-comparable imports. Some IO industries are as detailed as NAICS (e.g. Electronic computer manufacturing), but in other cases aggregation is quite strong (e.g. 24 NAICS codes within Apparel manufacturing become a single category).

We classify all industries into goods or services. Manufacturing, agriculture, and mining are classified into goods, while all other industries into services. Construction is sometimes viewed as a good-producing industry (Comin et al. 2021) and sometimes as a service industry (Cravino and Sotelo 2019). We treat construction as an industry ultimately providing shelter for households and businesses, therefore we classify it into services. Goods and services are further classified into 24 and 15 subsectors, corresponding to three-digit IO codes and two-digit NAICS codes, respectively. We assign Management and Administrative services (NAICS industries 55 and 56) to the subsector of Professional, Scientific, and Technical Services (code 54).

The use of the IO table is complicated by two considerations. First, the same product ("commodity") can be produced by different industries: for example, SUVs are manufactured by both SUV and car manufacturing establishments. We follow the standard procedure to address this issue by using the Supplementary Tables after Redefinitions (Horowitz and Planting 2009) and combining the Make and Use tables to produce a square commodity-by-commodity use matrix.

Second, distribution industries—wholesale, retail, and transportation—require special attention. BEA has two approaches for these industries, neither of which is fully consistent with our model. The standard "producer-value" table models the distribution margin (i.e., the cost of wholesaling, retailing, and transportation) as a flow going directly from the distribution industries to the buyers (whether final or intermediate); in this case, the data are aggregated across the various commodities that have a distribution margin. With this approach, it is not possible to see which group of consumers pays for retailing services (e.g., for the apparel they buy), which have low import shares. The import share of apparel, from the buyer perspective, has a large upward bias for the same reason. The supplementary "purchaser-value" IO table instead includes the distribution margin in absorption of each commodity, which resolves the problem under the proportionality assumption that domestic and imported apparel have the same fraction of retailing cost. Yet, this table is not consistent with the production side of our model. For instance, domestic apparel producers face very strong import competition, which occurs before both domestic and imported apparel is retailed.

We address these issues by constructing an "augmented" IO table which yields correct measures of exposure to trade for both consumers and producers. To do so, we create two versions for each industry. The "producer version" hires primary factors and purchases intermediate inputs, produces output, exports, and gets imported. Then, the entire value of domestic absorption is sold to the "purchaser version" of that industry, which also buys distribution services from the corresponding industries and sells the combined outcome to final consumers and to producer versions of industries using its output as an intermediate input. Only the distribution margin of exporting (e.g., wholesaling of exported goods) is recorded as direct exports of the distribution industries. All the formulas of our model apply to this IO table. Although import penetration and export shares are zero in purchaser industries, value added is zero there as well, so the labor market exposures can be computed using the augmented table. Similarly, the relevant measure of import shares is computed in purchaser industries, which is indeed where all final consumption is concentrated.

In constructing the augmented IO table, we use both the producer- and purchaser-value tables from the BEA. Moreover, to measure the distribution margin for domestically sold and exported goods by each commodity, we employ the Margin Details table separately published by the BEA. Unfortunately, that table does not distinguish between modes of transportation and types of retailing, which have different IO codes. Therefore, we aggregate those industries in the entire analysis, resulting in 381 industries instead of 389. We keep transportation industries 485000 (Transit and ground passenger transportation) and 492000 (Couriers and messengers) intact because they do not constitute distribution margins, as reflected by the fact that their producer and purchaser output values are the same.

S.2.2 Nielsen-Census Data

Overview. Consumer packages goods are goods typically purchased in supermarkets.[62] To estimate import shares for them by consumer group, we use detailed expenditure data from the Nielsen Homescan Consumer Panel (henceforth Nielsen) and match them to the confidential U.S. Census Bureau data on domestic production and imports at the firm level.

To measure the direct and indirect import share of each barcode, we find the product's manufacturer or distributor in the confidential U.S. Census data.[63] We proxy for a product's import share by the ratio of imports to total sales of the corresponding firm. This measure captures imports of both final products and intermediate inputs (except those imported through a domestic intermediary). It is also available for imports from China, NAFTA, and 34 developed economies specifically.

To implement this idea, we build a novel match between Nielsen barcodes and firms in the Census datasets, in three steps. We start by assigning barcodes in the

[62] The Nielsen data cover a set of products regardless of whether they were purchased in a supermarket or elsewhere. For discussions of the share of overall consumption covered in Nielsen, see Broda and Weinstein (2010), Kaplan and Schulhofer-Wohl (2017), and Jaravel (2019).

[63] While barcodes uniquely identify products, they are not informative about the country of origin, especially in the U.S. Although the first three digits identify the country in which the barcode was registered (Bems and Di Giovanni 2016), most foreign products in the U.S. have domestic barcodes. For some barcodes, the country of origin can instead be obtained from the product label (Antoniades and Zaniboni 2016; Auer et al. 2021). However, we are not aware of a possibility to automate collecting this information in the U.S., while manual collection is infeasible due to the massive number of products sold in the country. Neither of these approaches would also be informative about the indirect import share of domestic products.

Nielsen data to firms in the GS1 US database. GS1 is a non-profit organization than maintains the barcode system; to sell products in supermarkets, a manufacturer or a distributor has to purchase a block of barcodes from GS1. Each barcode can only be registered by one firm.

Next, we link firms in GS1 to Census Bureau's confidential Business Register (also called SSEL) by name and address. With a small fraction of exceptions, all firms in the GS1 data have a U.S. address—thus, foreign firms do not tend to register barcodes without an affiliate or an intermediary in the U.S. In turn, SSEL provides a comprehensive list of names and addresses of U.S. firms and establishments. We develop a set of consecutive rules for exact and fuzzy matching and verify match quality by manual inspection of a sample of firms.

We finally link SSEL firms to the quinquennial Economic Censuses from 2007 and 2012 and the transaction-level data on imports and exports of goods from the U.S. Customs (LFTTD) using unique firm identifiers. From the Economic Censuses we obtain the total value of a firm's sales, while LFTTD yields the total value of its imports. Dividing imports (overall or by trading partner) by sales, we get our measure of the import share. In this process, we use all Economic Censuses, including Censuses of Wholesale, Retail, and other sectors, and not just the more commonly used Census of Manufactures. Observing non-manufacturing firms is useful for us, as importing of final products is often done by wholesalers and retailers.[64] To reduce noise, we pool three years of the Nielsen data, 2006–2008 and 2011–2013, for each Economic Census year. Overall, out of the total number of 23,300 Nielsen firm-years, we successfully match 12,700, covering 83% of sales.

We adopt a square-root weighting scheme to reduce measurement error. As we cannot attribute a multi-product firm's imports to a particular product it sells, or even to consumer packaged goods overall for firms operating in multiple industries, our proxy for the import share is likely to be noisier for large firms. Large multi-product firms play a large role when measuring the share of import spending by consumer group, a consequence of "granularity" in firm-level datasets (Gabaix 2011). Our solution, similar to Caron et al. (2014), is to reduce the influence of large firms by rescaling each firm's Nielsen sales to its square root and adjusting expenditures

[64] In Appendix S.2.2, we track several products photographed in a Walmart store to verify that domestically produced goods are normally registered by the manufacturer, while imported products are registered by the distributor, often a wholesaler.

on its products by each consumer group proportionately.[65]

Column 2 of Table S1 presents summary statistics for the linked dataset, while Table S7 details them by the three product classes. The average import share is 11.1%, similar to the entire consumption basket reported in Column 1, but there are large differences across product classes. We next provide details on the data sources, the matching process, and present further match statistics and examples.

Data sources. The Nielsen company asks around 55,000 U.S. households per year to record all purchases within certain classes of products. Consumers scan purchased goods using handheld barcode scanners provided by Nielsen. They also manually enter products that do not have barcodes, such as fresh produce. Nielsen obtains price information from a combination of store data and manual entry by households. The stratified sample of households is representative of the U.S. population in terms of income, education, age, race, household size, and other characteristics when using the Nielsen-provided projection weights.

GS1 maintains the concordance between barcodes and firm names and addresses; the version we obtained is complete as of February 2016. We drop 5.2% Nielsen barcodes which we could not link to GS1 (they constitute 1.8% of total sales in Nielsen). In most cases GS1 firms are located within the U.S., although there are some exceptions, mostly with Canadian addresses. We drop firms with addresses outside 50 U.S. states and Washington, D.C. or with missing state information, which constitute 4.3% of all Nielsen firms but only 0.75% of total sales.

We use three data sources on the Census side. Business Register, or SSEL, is the comprehensive list of establishments, with names and addresses, assembled using Census surveys, Internal Revenue Service tax data, and other data sources at the annual frequency (DeSalvo et al. 2016). Because firms change names and addresses over time, while GS1 provides only one observation per firm, we use addresses in the SSEL for all years from 1991–2014, which improves the quality of the merge.

The Economic Census is the survey of all business establishments in the U.S. It is conducted by the Census Bureau in years that end with 2 or 7, and participation is

[65] Granularity is a substantial challenge: in the full Nielsen sample, top 50 firms capture 46% of sales in an average year; with square-root weights, they take up only 9%. In unreported results we verify that all findings are very similar, both qualitatively and quantitatively, when firms are weighted by their Nielsen sales to the power of 1/4 or 3/4, or by the square-root of the firm's sales in the Economic Census.

required by law. The content of the questionnaire varies across sectors and industries but all of them include questions on the total revenue. We primarily use Censuses of Manufactures, Wholesale, and Retail. Establishments in Services, Finance, and Utilities are also part of our Economic Census sample, but they are rarely matched to Nielsen.

Finally, LFTTD (Linked/Longitudinal Firm Trade Transaction Database) is the microdata on all international trade transactions, based on the import declarations and shippers export declarations. It has been matched to the Census by firm identifier (see Bernard et al. 2009).

Sample construction. We predict total sales of each Nielsen barcode by applying projection weights provided by Nielsen to the purchases by each household and, using the GS1 crosswalk, aggregate them to firms and firm-by-product module cells. We classify households into college- and non-college by using education of both male and female heads. If they are both present but only one has college degree, we attribute half of the purchases to each education group. Income is reported in 16 discrete bins, and we use their midpoints.[66] Income is reported with a two-year lag, so we use the value from two years after, whenever available.

We apply several filters to Nielsen. First, we drop households with reported income below $5,000. Second, we drop "magnet data"—products that do not use standard barcodes, such as fresh fruits and vegetables. Finally, we also drop firm-years with less than five unique barcode-household pairs and those with total unweighted spending by Nielsen panelists under $100—we label those as "tiny" Nielsen firms. From now on, we will suppress mentioning years.

We then compute import shares for each Census firm. The numerator is total imports from LFTTD. To measure the total firm output in the denominator, we aggregate revenue of all establishments belonging to the firm. However, this creates double-counting if a manufacturing company ships its products to its own wholesalers or retailers and then sells them. Therefore, we only count the total revenue in the largest 2-digit NAICS sector in which the firm operates, although the results are not substantially different without this correction. We drop firms for which imports

[66] The cutoffs in $000 are: 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, and 100. In some years, the top-income group is decomposed further, but we use a consistent classification. We assign the top-income group the value of $140,000, based on the average income in the years when we have more detailed data.

exceed 200% of annual sales, indicating an imperfect match between LFTTD and the Census.

Finally, we merge name and addresses in GS1 with the Census firms—a procedure we describe next. Once done, we implement a consistency filter. Some firms, particularly large ones, span many industries, so their scope may not be covered well by the set of products covered by Nielsen. As a result, the overall importing behavior may be a very bad proxy for the set of products covered by Nielsen. We therefore require that Nielsen sales of a firm are within the range of 1% and 300% of the Census sales. Although still wide, this range excludes strongest violations of consistency in both directions and makes our results robust to using the square-root of Nielsen or Census sales as weights.

Merging process. We match names and addresses between GS1 and each year of SSEL from 1991–2014 separately. The process consists of three steps. First, we pre-process names and addresses in both datasets to maximize the probability of exact matches. Second, we develop a series of matching rules and apply them starting from the strictest, giving priority to multi-establishment Census firms. Third, because names and addresses change over time, some matches will only be found in some years. We extrapolate them to other years whenever possible. We now describe each step in detail.

Pre-processing. We use the algorithms from the `reclink2` package from Wasi and Flaaen (2015), with minor modifications. For company names, the `stnd_compname` command removes special symbols, makes standard substitutions (e.g., INTL to International), and isolates the entity type (e.g., INC) into a separate variable. Pre-processing of addresses is particularly important. The `stnd_address` command parses them into several parts: the main address variable (where special symbols are removed, street types are converted to their abbreviations, e.g., Street into ST, etc.), as well as the post office box, unit (e.g. SUITE 1400), and building numbers, if present. We implement an important addition to this parsing procedure by also extracting the house number from the address. We define it as the number at the beginning of the address or, if the address starts with a letter, the largest number in the address.[67]

Matching algorithm. The SSEL consists of records of three types: multi-unit (one per establishment for firms with multiple establishments), "submaster" (one per tax identifier of a multi-unit firm, created for consistency with the IRS), and single-unit. We give priority to multi-unit and submaster records by first attempting to match GS1 firms to them. For GS1 firms that are still not merged, we try matching to single-unit firms that are part of the LBD (the Longitudinal Business Database, which links SSEL records across years). The lowest priority is given to single-unit firms outside of the LBD.[68]

Within each priority level, we apply consecutive matching rules, starting from the strictest one. Once a GS1 firm finds an SSEL match, it is removed from the process. This guarantees that each GS1 firm is matched to only one Census firm, except for rare cases when we find several matches using the same matching rule. At the same time, we allow several GS1 firms to be matched to the same Census firm, as should be the case for subsidiaries of the same firm that appear in GS1 separately.

We developed seven matching rules by trial and error and manually checked samples of matched firms to verify that each of them mostly produces correct matches. Each rule requires an exact match and non-missing values for some key variables, an exact match on additional variables where missing values are allowed, and a bigram probabilistic ("fuzzy") match on other variables with a specified match score threshold. The implementation is again based on the `reclink2` package from Wasi and Flaaen (2015). While we kept its logic, we substantially improved computational efficiency in our own `reclink4` command.

Table S8 lists the rules. The two strictest rules require a non-missing match for the 9-digit zipcode (ZIP+4). Although available only for some firms, it generally identifies the building or a post box precisely. The first rule additionally requires an exact (possibly missing) match for the firm name, house number, address, PO Box,

[67] Extracting the largest number is inspired by the the addresses of foreign firms are treated in the LFTTD (see Kamal and Monarch 2016). With fuzzy matching, matching on the house number ensures that buildings like 47 Main St. and 49 Main St. are distinguished. It is also very useful for parts of Wisconsin and Illinois which use alphanumeric addresses, e.g. "W190 N10768 Commerce Cir, Germantown, WI."

[68] One SSEL record may list up to two addresses per establishment (physical and mailing) and sometimes specifies two zipcodes (one reported and one inferred automatically based on the rest of the address). We use all available versions of the address to increase the probability of the match.

unit, and building, standardized as previously described, while the second rule only requires an exact match on the house number, while the other variables are matched in a fuzzy way. The least restrictive seventh rule requires exact matches on the firm name, its entity type, and state, still delivering high quality of matches for the records that have not been matched using stricter rules.

Extrapolation of matches. Matching with GS1 is done separately for each year of the SSEL. If a GS1 firm does not find any SSEL match in a given year t, we turn to the matches that were found for this firm in other years, with preference to the closest years (starting with $t+1$, then using $t-1$, $t+2$, $t-2$, $t+3$, etc.). If some match is found in year t', we check in the LBD whether the matched firm existed in t and, if so, use this match for year t.

Match statistics. Panel A of Table S9 shows that the majority of Nielsen firms, excluding tiny ones, is matched, covering over 83% of total Nielsen sales.[69] In 2007, there were 26,900 Nielsen firms, and elimination of the tiny ones leaves us with 11,000 without any significant loss in total projected sales. Out of them we are able to find a Census match in the same year of the Census Business Register for 7,600, while using names and addresses from other years adds another 600 firms, making it 8,200 total. Although all firms are supposed to fill out Census forms, not all of them do, so we find 7,200 Nielsen firms in at least one of the Censuses, and of them 6,100 pass the consistency filter. Although there are a few cases where we find two Census matches for the same Nielsen firms, the number of Nielsen firms with single matches is the same 6,100 after rounding. Statistics are similar for 2012, increasing the sample size to 12,700 firm-years.

Panel B of Table S9 shows merging statistics starting from Census firms. Since Nielsen only covers consumer packaged goods, we do not expect a high match rate in most industries. However, Nielsen coverage is strongest for food, alcohol, and tobacco. This panel starts from all 51,500 firms in the Census of Manufactures in the corresponding NAICS codes 311 and 312. Out of them, 8,900 (or 17.3%) are merged to any Nielsen firm, including the tiny ones, and the merged ones account for 79% of the total sales. After dropping small Nielsen firms and implementing the

[69]The match rate is above 83% of sales for food and health and household products, but a bit worse for general merchandize, at 76%.

consistency filter, we match only 9.3% of the firm count but still 58.7% of sales by all manufacturers in the industry. Note that we also merge many wholesalers and retailers selling food, not accounted for in this table.

Table S10 shows that multi-establishment firms are a minority in the matched sample (29%), but they cover 93% of sales. Within both multi- and single-establishment matched firms, the strictest matching rule 1 captures the largest share of firms, but all rules contribute to the sample.

Table S11 shows the fractions of firms operating in different sectors, defined by their 2-digit NAICS codes, in the sample.[70] The manufacturing sector constitutes the largest fraction of the sample (57.2% with square-root weighting), followed by wholesaling (29.0%) and retailing (8.7%). The smaller share of retailers is in part determined by their large average sales, which imply that the square-root weighting scheme reduces their importance. At the same time, it is important to understand that most products sold by retailers are registered by other firms. We discuss below examples of products showing that this is true even for products manufactured for and distributed exclusively by Walmart. Among the 3-digit NAICS codes, Food Manufacturing and Nondurable Goods Wholesalers are the most prevalent ones, followed by Chemical Manufacturing (which includes soap, shampoos, etc.) and Beverage and Tobacco Manufacturing.

The last column of Table S11 presents a nice test on the quality of the match. Nielsen data allow us to identify products that are branded by the retail chain that sells them ("private label brands"). We find that over 99% sales of barcodes registered by food and beverage stores, according to their main NAICS code in the Economic Census, are private label brands. For comparison, this share is only 7.9% for wholesalers and mere 1.2% for manufacturers.

Table S12 examines how representative the matched sample is. Panel A compares firms in Nielsen, excluding tiny ones, that found a match to those that did not. Median firms in the merged sample have about twice as large Nielsen sales relative to the firms that did not find a match. Matched firms also sell to slightly, but statistically significantly, poorer and less educated consumers. For example, 29.1% of sales of matched firms is to college graduates, as opposed to 30.7% for firms that

[70]Because Census data provides NAICS codes for establishments not firms, we classify firms by the 2- and 3-digit NAICS in which they have the largest payroll, excluding NAICS code 55 "Management of Companies and Enterprises".

we did not match. However, these differences can largely be explained by the size difference; they are reduced when controlling for a quadratic term in log Nielsen sales. Panel B provides evidence on sample selection for the firms in the Census of Manufactures producing food, alcohol, and tobacco. Again, merged firms are much larger, with median sales of $13.3 million, payroll of $1.9 million and 54 employees, as opposed to $606,000 sales, $113,000 payroll and 4 employees for a median Census firm that we did not merge. Comparing these sets of firms by skill intensity (the payroll share of non-production workers) does not reveal statistically or economically significant differences.

Examples. A few examples illustrate the way in which our Nielsen-Census linked dataset labels products as domestic, imported, or as using imported intermediate inputs. We visited a Walmart store and photographed a sample of products, which we identify as domestic and imported by looking at their labels. Then, we identified these products in the GS1 database using their barcodes and searched for the information about the firms that registered them on the Internet. Figure S15 shows pictures of five products that illustrate well different situations we observed.[71]

Panels A and B show two plates labeled as "Made in the USA"; one is from an independent brand and the other is distributed by Walmart. According to the GS1 data, they were respectively registered by World Kitchen, LLC and Merrick Engineering Inc.. An Internet search shows that both of these companies are U.S.-based manufacturing firms, so our Nielsen-Census linked dataset will label them as domestic products (unless these companies use a lot of imported intermediate inputs).

The three other products on Figure S15 are all imported. The bed sheets in Panel C belong to the same brand as the plates in Panel B, they are distributed by Walmart, but their label indicates "Made in China". In GS1, we see that their barcode was registered by Jiangsu Royal Home USA, Inc. Internet sources indicate that this firm belongs to the NAICS code 423220 "Home furnishing merchant wholesalers" and imports its products from China. Our Nielsen-Census linked dataset will therefore label this product as imported.

The plates in Panel D, also made in China, are registered by First Design Global, Inc, which (again, according to Internet search) is a U.S.-based manufacturing firm

[71] We have not used any Nielsen or Census data in this section. These products may or may not be in our final sample.

but imports tableware and kitchenware from China. We will therefore attribute these plates partially to imports, using as a proxy the ratio of imports to total sales for this firm. This proxy does not bias our estimates of the expenditure channel if there is no systematic correlation between import shares and buyer characteristics within firms.

Finally, the Canadian hair conditioner from Panel E is distributed by Walmart and, unlike the aforementioned products, was registered by Walmart itself. Therefore, in the Nielsen-Census linked dataset, we proxy for the probability that it is imported by the fraction of Walmart's direct imports relative to its total Census sales. This may be an underestimate if Walmart's direct imports in the Customs data mostly cover its own-registered products, whereas its sales include all products, e.g. those from all previous pictures.

S.2.3 Datasets for Motor Vehicles

Linked CEX-Ward's dataset. To measure purchases of motor vehicles by consumer group, we use the OVB file ("Owned Vehicles Detailed Questions") from the CEX Interview Survey, which asks respondents to provide information about all vehicles they own, including the brand, whether the vehicle was purchased new or used, and in some cases the price. The respondents are asked to list all types of vehicles, but we focus on cars and trucks (which are mostly light trucks, i.e. SUVs, although in some cases could be medium-duty trucks as well), excluding motorcycles, boats, etc. As previously, we classify households into groups according to bins of household income or to the college education of the respondent after converting income to the 2007 prices as in Appendix S.2.1.

The data are available since 2006 but we use it for 2009–2015 for consistency with the Ward's sample. Each household is expected to participate in the survey for four consecutive quarters, so to avoid duplication we only use the most recent survey in which the OVB survey is filled. If the household reports several vehicles in that survey, we use all of them. Like in other datasets we build, we drop vehicles owned by households with income before tax below $5,000.

Data on importing come from Ward's Automotive Yearbooks. We use the electronic versions of the 2011, 2013, 2014, and 2016 yearbooks. Each of them shows the statistics for the previous two years, thus covering the entire 2009–2015 period. In each year we use five Ward's tables. Two are on sales in the U.S. (U.S. Car Sales by

Line by Month and same for Light Trucks): for each model (also called "lines," e.g. Chevrolet Camaro) they decompose the number of cars and light trucks sold in the U.S. into those built within and outside NAFTA. The other three tables (U.S. Vehicle Production by Line by Month and same for Mexico and Canada) report production by country and model, allowing us to decompose vehicles assembled within NAFTA into those built in the U.S., Canada, and Mexico.

We define a model as imported if it was assembled outside of the United States. Most models are either only imported or only assembled domestically, but in a few cases assembly occurs both in the U.S. and abroad. In such cases, we classify the model as "partly imported": our proxy for its direct import share is the fraction of vehicles of this model that were assembled outside the U.S. We then aggregate models to brands with total U.S. purchases as weights.

Specifically, we first aggregate all years of Ward's data to measure, for each model, the number of vehicles sold in the U.S., the share assembled outside NAFTA, and the shares of assembly within NAFTA that comes from the U.S., Canada, and Mexico separately. We then compute the domestic share of each model sales as the product of those from within NAFTA (from tables on sales) and the share of U.S. within NAFTA production (from production tables). For one model only (BMW Z4), the sales table reports some NAFTA production, but production data are missing, in which case we checked the country of production manually. At the end we aggregate all models by brand using sales weights from Ward's.[72]

We dropped a small fraction of CEX purchases for brands that we do not observe in Ward's because their production was discontinued before 2009. Oldsmobile is the most frequent brand we have to drop. All dropped brands combined constitute less than 1.5% of the sample. We also keep four brands (Daewoo, MG, Austin-Healey, Zenn) which are in CEX but not in Ward's, and are fully imported. This results in the sample of 45 brands, listed in Table S3. Column 3 of Table S1 presents summary statistics.

[72]We attribute all imports from countries other than NAFTA to the 34 developed economies. Ward's data do not report origin countries for cars imported from outside NAFTA. However, according to the BACI database on bilateral trade flows, out of all U.S. imports of motor vehicles (HS code 8703) in 2009–2016 from outside NAFTA, less than 3% were countries other than our 34 (mainly from South Africa).

Linked CEX-Census dataset. Import shares constructed with the CEX-Ward's dataset do not include imports of intermediate inputs used to produce domestically assembled vehicles. To address this potential limitation, we match the CEX to the confidential Census of Manufactures and LFTTD, where the fraction of imported car parts in the value of manufacturer's sales can be measured and compared to the value of imported assembled cars. This requires aggregating the CEX data from brands to firms.[73]

We use the 2012 version the Census of Manufacturers and the Customs data for the same year. We match these data to expenditure shares from the CEX. To increase sample size, we use all years of the CEX when the brand variable is available, from 2006 to 2015. In this analysis, we include cars only, not light trucks.

To match domestic car producers in the CEX, we first link each car brand to the firm that owned it in 2012, using the Ward's Automotive Yearbook and Internet search. Then we manually search for firm names in the 2012 Business Register (SSEL)—the list of all establishments in the U.S., and obtain the firm identifier or identifiers for all firms that participated in the Census.

Our sample includes two types of observations. If a firm has no production in the U.S., we keep its brands separately and assign 100% imports, both direct and total. And if a firm has some U.S. production (and participated in the 2012 Census of Manufacturers),[74] we aggregate its brands together and measure import shares.

The value of imports of assembled cars is defined as total imports in the Customs data in the Harmonized Trade Classification (HS) code 8703 "Motor cars and other motor vehicles principally designed for the transport of persons".[75] Imports of car parts are defined as those in HS codes 8706 (chassis fitted with engines), 8707 (bodies for motor vehicles), 8708 (parts and accessories of motor vehicles), 84 (machinery), 85 (electrical machinery and equipment), 90 (measuring and other instruments), 39 (plastics), 40 (rubber), 73 (articles of iron and steel), 83 (miscellaneous articles of base metal), and 94 (furniture).

[73] The CEX-Ward's sample shows that there can be meaningful variation across brands within a firm: for example, the share of vehicles assembled abroad is about 20% for Buick and 40% for Chevrolet, which are both produced by GM.

[74] Participation in the quinquennial Census is required by law, so the vast majority of firms reply. However, not all of them do, and the information on participation is confidential.

[75] This HS code includes some vehicles besides cars (e.g. SUVs and ambulances), which may create some upward bias.

We measure car sales by the sum of total shipments of domestically assembled cars and the imports of assembled cars. The former is defined as the total value of shipments from all of the firm's establishments which belong to NAICS code 33611 (Automobile and light duty motor vehicle manufacturing) in the Census of Manufactures. Then the direct (total) import share is the ratio of imports of cars (cars plus parts) in car sales. Note that while we use counts of vehicles in the CEX and Ward's data (due to data availability), here import shares as defined by value.

S.2.4 ACS Data

To measure the composition of workers by industry, education, and earnings, we use the 2007 American Community Survey from IPUMS — the long form of the population census answered by a random 1% sample of the U.S. population every year (Ruggles et al. 2015). We select only employed workers and drop the public administration sector. For the analyses of earnings, we rely on the `incwage` variable that captures total pre-tax wage and salary income during the previous calendar year; we only consider workers with earnings of at least $5,000. We split them into ten deciles, with the cutoffs, in $000, of 10.7, 16.0, 21.3, 27.0, 32.2, 40.0, 49.0, 60.0, and 85.0.

Since industries in ACS are more aggregated than IO codes (there are 253 codes overall, recorded in the variable `ind`), we have built a weighted crosswalk from ACS industries to IO codes. First, for each ACS industry we find the set of corresponding NAICS industries using a crosswalk provided by IPUMS.[76] Second, we allocate each ACS code to those NAICS industries with weights proportional to the total payroll by NAICS, which we obtain from the 2007 Quarterly Survey of Employment and Wages.[77] Third, we aggregate NAICS industries to IO codes. Finally, in parallel to our approach to the CEX (see Appendix S.2.1) we reweight the ACS to match the total compensation of employees by IO code from the IO table.

[76]Only in one case (NAICS industry 519130) the same NAICS code corresponds to two IND codes. We split this NAICS code into two proportionately to the IND payroll.

[77]The QCEW tabulations are published by the Bureau of Labor Statistics based on unemployment insurance statistics.

S.2.5 Estimation of Income Elasticities

Here we describe the procedure used to estimate income elasticities for each IO industry in Section 5, based on the CEX-IO data. Our approach uses the definition of the income elasticity as describing the relationship between spending and consumer expenditure (the Engel curve). As long as different consumers in the same country have the same preferences, face the same prices, and income elasticities $\psi_{xj} \approx \psi_j$ do not have much variation across income levels x, cross-sectional data allow us to estimate ψ_j directly. By taking this approach, we avoid making any parametric assumptions on the utility function and estimating demand structurally. Intuitively, higher-income consumers have larger expenditure shares on income-elastic products. Using this logic, we first compute the income *semi*-elasticity for each spending category by regressing spending shares on the logged total expenditure and then convert the estimates to elasticities and aggregate them into the IO industries.

Specifically, we split households in the CEX sample into 11 bins by the reported pre-tax household income and compute consumption shares across all spending categories (UCC) j for each of the bins i separately (s_j^i) and overall (s_j). Then for each spending category we estimate the income *semi*-elasticity by regressing, across income bins, spending shares on the log of total expenditure in this income group, averaged across households:

$$s_j^i = \text{constant}_j + \psi_j^{\text{semi}} \log \text{Expenditures}_i + \text{error term}_{ij}.$$

Observations are weighted by the number of households in each income bin. For an income-elastic spending category, the share is increasing in the total expenditures, so $\psi_j^{\text{semi}} > 0$, and the reverse holds for income-inelastic products. We then convert the semi-elasticity into the elasticity ψ_j for an average consumer of product j:

$$\psi_j = 1 + \frac{\hat{\psi}_j^{\text{semi}}}{s_j}.$$

The intermediate step with semi-elasticities guarantees that the spending-weighted average of income elasticities across all spending categories is equal to one, as it should be theoretically:

$$\sum_j \psi_j s_j = \sum_j s_j + \sum_j \hat{\psi}_j^{\text{semi}} = 1 + 0 = 1,$$

where $\sum_j \hat{\psi}_j^{\text{semi}} = 0$ because spending shares sum up to a constant (one) for each income group, and the regression of a constant on $\log \text{Expenditures}_i$ yields a zero slope.

Expenditures are used on the right-hand side instead of income because in the CEX, total expenditures do not vary one-to-one with reported income. That relationship is increasing but much less than proportionate, which may be a consequence of imperfect measurement of income—either because current income is not a good proxy for permanent income, or for pure measurement error reasons. In either case, income elasticity estimates would be biased towards one if income was used on the right-hand side.

We winsorize a small number of ψ_j to be between -1 and 3. At the end we convert the UCC-level income elasticities to IO codes in the same way as we do for the expenditure shares.

S.2.6 Ranges of Substitution Elasticities

Our Section 5 calibrations require four substitution elasticities: between domestic and foreign varieties (ξ_j), between goods and services (ρ), between industries within goods and services (ε_r), and the macro elasticity of substitution between college and non-college labor, σ_{macro}. In all cases we take prevalent values from the literature for the baseline analyses and consider intervals of values that cover many available estimates for robustness.

Our baseline value for ξ_j is 3.5 in all six-digit industries, which is equivalent to a trade elasticity of $\xi_j - 1 = 2.5$. In robustness checks we consider values of ξ_j between 1.9 and 5.1 and also allow ξ_j to vary across three-digit IO industries according to the estimates from Broda and Weinstein (2006). Our baseline value is near the median elasticity of 3.7 reported in Broda and Weinstein (2006) for ten-digit industries, and of 3.4–3.7 in Soderbery (2015) using the same Broda-Weinstein method but for eight-digit industries and for different years of data, as well as near the mean of 3.6 in Ossa (2015). The range of 1.9 to 5.1 corresponds to the estimates from Soderbery (2015)'s LIML procedure and Simonovska and Waugh (2014), respectively. This interval also covers typical values of the elasticity of substitution between domestic and foreign varieties in Feenstra et al. (2018).

We set the elasticity of substitution between goods and services in consumption to $\rho = 0.6$, obtained from Cravino and Sotelo (2019). For robustness we consider the range between 0.2 and 0.85, as in Comin et al. (2021) (see also Cravino and Sotelo (2019)).

The elasticities of substitution between industries within each goods and services, ε_r, are more difficult to obtain (Dawkins et al. 2001; Costinot and Rodríguez-Clare 2015). As they are expected to lie between ρ and ξ_j, we set $\varepsilon_r = 2$ in the baseline analyses and consider values between 0.6 and 3.5. A recent paper by Redding and Weinstein (2017) estimated the elasticities of substitution between 6- and 4-digit NAICS industries to be 1.47 and 1.34, respectively. The estimate by Hottman and Monarch (2020) using 4-digit HS industries is 2.78. The range of elasticities we use covers all of these values.

Finally, for the calibration across education groups we set the macro elasticity of labor substitution to $\sigma_{\text{macro}} = 1.41$ from Katz and Murphy (1992). We check robustness to the range of $[1.41, 1.8]$, with the upper bound corresponding to the estimates from Acemoglu (2002) and Acemoglu and Autor (2011).

S.2.7 Counterfactuals Based on Observed Changes in Trade Costs

While our main analysis considered hypothetical trade shocks that are uniform across industries, here we consider the distributional effects of other counterfactuals based on shocks observed in the data. We calibrate the effects of three shocks to importing costs: the introduction of Trump tariffs in 2018 (on solar panels, washing machines, steel and aluminum products, and Chinese products), the observed change in tariffs between 1992 and 2007, and the observed change in "import charges" (defined as transportation and insurance costs) in the same period. We view tariffs as iceberg trade costs, ignoring tariff revenue.

The formulas in our model section capture the effects of a shock to importing trade costs of the same magnitude $d\log\tau$ for all imports from a set of countries c. A slight modification is necessary to capture a shock that varies across industries in proportion to some variable r_j, i.e., $d\log\tau_j = r_j d\log\tau$. Indeed, in this case the industry import price index (before IO adjustments) equals $(IP_{jc}r_j/IP_j)\, d\log\tau$ instead of $(IP_{jc}/IP_j)\, d\log\tau$. Thus, simply replacing IP_{jc} with $IP_{jc}r_j$ allows us to estimate the counterfactual changes of the equilibrium in the first order approximation. We describe below how c and r_j are defined for each of the three shocks we consider. In all three cases, we first measure the shock at the level of HS codes and then average it at the level of the corresponding IO code using the HS-NAICS concordance from Pierce and Schott (2012).

The first shock is the set of tariffs introduced by the Trump administration in 2018. We combine three sets of tariffs:

1. *Solar panels and washing machines.* Actual tariffs on solar panels and large residential washing machines have a complicated structure: their rates vary over time, they are combined with quotas, and certain exceptions are provided, as described in Presidential Proclamations 9693 and 9694 of January 23, 2018. We approximate these rates by using the base rates (30% for solar panels and 20% for washers) applied to the main HS codes described in the Proclamations and to all U.S. trading partners.

2. *Steel and aluminum products.* Tariff duties on imports of steel and aluminum by trading partners are given in Section 232 of the Trade Expansion Act of 1962. The tariff increases were proposed on March 1 and amended on May 31, 2018. We identify the steel and aluminum products that were affected by these tariff increases using the published lists of HS codes. We apply a 25% tariff on steel products, excluding imports from Argentina, Australia, Brazil and South Korea, and a 10% tariff on aluminum products excluding Argentina and Australia.

3. *China tariffs.* Tariffs on products imported from China were introduced according to Section 301 of the Trade Act of 1974. They were released by the Office of the U.S. Trade Representative in three tranches with different lists of products. The first two were finalized on June 15 and August 7, 2018, taxing approximately $34bln and $16bln (in terms of 2018 imports), respectively, with a rate of 25%. The third one, finalized on September 17, introduced a tariff of 10% on approximately $200bln of imports.

The other two shocks we consider are the observed changes in (i) tariffs and (ii) import charges (transportation and insurance costs) between 1992 and 2007. We obtain data on both types of changes from the Census Bureau trade statistics made available by Schott (2008). For each IO industry and year, we measure the rate of tariffs t_j (or import charges c_j) as the share of total tariff duties (or total transportation/insurance costs) in total imports for personal consumption. For each industry j, the shocks are given by the change in $\log(1 + t_j)$ and $\log(1 + c_j)$ between 1992 and 2007.

The results are shown in Figure S11 and Panel B of Figure S12.

S.2.8 Census Data for Skill Intensity and Exports

To measure the relationship between skill intensity and exporting at the plant level (Table S6), we use Census microdata. We focus on the manufacturing sector because it is the only one where the information of the worker types is available and it is the most tradable sector.

Until recently, Census surveys did not ask establishments about education of their workers, which led to a long tradition to proxy for skill intensity by the payroll or employment share of non-production workers (e.g. Berman et al. 1994; Autor et al. 1998), who are considered to be more skilled than production workers (Berman et al. 1998). The situation has changed with the arrival of the 2010 Management and Organizational Practices Survey (MOPS) survey, which is a supplement to the Annual Survey of Manufactures (ASM), covering all largest firms as well as a sample of smaller ones.

We use MOPS questions 32–35, which ask for number of managers and employees, as well as the share of managers and non-managers with a college (bachelor) degree.[78] The shares are listed in terms of discrete bins, so we use the midpoints of those bins.[79] This yields an estimate of the share college graduates in total employment, $v^{\text{Emp}}_{\text{college}|j}$. Unfortunately we do not observe wages of college- and non-college workers. Therefore, to impute the payroll share we use the economy-wide average wages of these groups from the U.S. Census Bureau (DeNavas-Walt et al. 2011). They show that the median wage of college graduates is about 80% higher than that of non-college workers (considering individuals in the labor force and 25 years or older), so we measure the payrolls share of college graduates in each establishment j as

$$v_{\text{college}|j} = \frac{1.8 \cdot v^{\text{Emp}}_{\text{college}|j}}{1.8 \cdot v^{\text{Emp}}_{\text{college}|j} + \left(1 - v^{\text{Emp}}_{\text{college}|j}\right)}.$$

[78] The questionnaire is available at https://www2.census.gov/programs-surveys/mops/technical-documentation/questionnaires/mop-2010.pdf; also see Bloom et al. (2016). We drop observations where answers to any of these questions are missing.

[79] The bins are under 20%, 21–40%, 41–60%, 61–80%, and over 80% for managers and 0%, 1–10%, 11–20%, and over 20% for non-managers (we assign 25% to the last category).

It is very strongly correlated with $v^{\text{Emp}}_{\text{college}|j}$, so the details of imputation are not consequential. We then distribute each firm's total payroll to the two education groups according to these shares to compute the payroll-weighted average export shares by group in Table S6.

Besides the MOPS sample, we use the 2010 ASM and the full 2007 CMF, which report payroll to production and non-production workers directly. We match all of them to the Customs microdata (LFTTD) to measure export shares. Like Bernard et al. (2018), we do not use the CMF and ASM questions about plant exports, which are less reliable than direct observation of trade transactions. For firms with multiple establishments, we attribute firm exports proportionately to the value of establishment sales (shipments). We drop firms where exports exceed twice the total value of manufacturing sales, as those are likely to result from measurement error or other firm establishments which are not part of the sample (e.g. the non-manufacturing ones). We compute the export share of an establishment relative to the value of shipments.

References

Acemoglu, D. 2002. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature* 40:7–72.

Acemoglu, D., and D. H. Autor. 2011. "Skills, tasks and technologies: Implications for employment and earnings." In *Handbook of Labor Economics,* 4:1043–1171.

Aguiar, M., and M. Bils. 2015. "Has Consumption Inequality Mirrored Income Inequality?" *American Economic Review* 105:2725–2756.

Antoniades, A., and N. Zaniboni. 2016. "Exchange Rate Pass-Through into Retail Prices." *International Economic Review* 57:1425–1447.

Arkolakis, C., and M. Morlacco. 2017. "Variable Demand Elasticity, Markups, and Pass-Through." *Mimeo.*

Auer, R., A. Burstein, and S. Lein. 2021. "Exchange Rates and Prices: Evidence from the 2015 Swiss Franc Appreciation." *American Economic Review* 111:652–686.

Autor, D. H., L. F. Katz, and A. Krueger. 1998. "Computing inequality: have computers changed the labor market?" *The Quarterly Journal of Economics:* 1169–1213.

Bems, R., and J. Di Giovanni. 2016. “Income-Induced Expenditure Switching.” *American Economic Review* 106:3898–3931.

Berman, E., J. Bound, and Z. Griliches. 1994. “Changes in the Demand for Skilled Labor Within U. S. Manufacturing: Evidence From the Annual Survey of Manufactures.” *The Quarterly Journal of Economics* 109:367–397.

Berman, E., J. Bound, and S. Machin. 1998. “Implications of skill-biased technological change: international evidence.” *The Quarterly Journal of Economics* 113:1245–1279.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott. 2018. “Global Firms.” *Journal of Economic Literature* 56:565–619.

Bernard, A. B., J. B. Jensen, and P. K. Schott. 2009. “Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods.” In *Producer Dynamics: New Evidence from Micro Data,* 513–552.

Bloom, N., E. Brynjolfsson, L. Foster, and R. Jarmin. 2016. “What Drives Differences in Management?”

Broda, C., and D. E. Weinstein. 2006. “Globalization and the Gains From Variety.” *The Quarterly Journal of Economics* 121:541–585.

———. 2010. “Product Creation and Destruction: Evidence and Price Implications.” *American Economic Review* 100:691–723.

Buera, F. J., J. P. Kaboski, and R. Rogerson. 2018. “Skill Biased Structural Change.” *Mimeo.*

Caron, J., T. Fally, and J. R. Markusen. 2014. “International trade puzzles: A solution linking production and preferences.” *Quarterly Journal of Economics* 129:1501–1552.

Chaney, T. 2008. “Distorted Gravity: The Intensive and Extensive Margins of International Trade.” *American Economic Review* 98:1707–1721.

Comin, D., D. Lashkari, and M. Mestieri. 2021. “Structural Change with Long-run Income and Price Effects.” *Econometrica* 89:311–374.

Costinot, A., and A. Rodríguez-Clare. 2015. “Trade Theory with Numbers: Quantifying the Consequences of Globalization.” In *Handbook of International Economics,* 4:197–261.

Cravino, J., and S. Sotelo. 2019. “Trade Induced Structural Change and the Skill Premium.” *American Economic Journal: Macro* 11:289–326.

Dawkins, C., T. N. Srinivasan, and J. Whalley. 2001. “Calibration.” *Handbook of Econometrics* 5:3653–3703.

DeNavas-Walt, C., B. D. Proctor, and J. C. Smith. 2011. “Income, Poverty, and Health Insurance Coverage in the United States: 2010.” *U.S. Census Bureau Current Population Reports* (Washington, D.C.).

DeSalvo, B., F. F. Limehouse, and S. D. Klimek. 2016. “Documenting the Business Register and Related Economic Business Data.”

Eaton, J., and S. Kortum. 2002. “Technology, geography, and trade.” *Econometrica* 70:1741–1779.

Fajgelbaum, P., and A. K. Khandelwal. 2016. “Measuring the Unequal Gains from Trade.” *Quarterly Journal of Economics* 131:1113–1180.

Feenstra, R., P. Luck, M. Obsteld, and K. Russ. 2018. “In Search of the Armington Elasticity.” *Review of Economics and Statistics* 100:135–150.

Gabaix, X. 2011. “The Granular Origins of Aggregate Fluctuations.” *Econometrica* 79:733–772.

Garner, T. I., R. Mcclelland, and W. Passero. 2009. “Strengths and Weaknesses of the Consumer Expenditure Survey from a BLS Perspective.” *Mimeo.*

Horowitz, K. J., and M. A. Planting. 2009. *Concepts and Methods of the U.S. Input-Output Accounts.* Washington, D.C.: Bureau of Economic Analysis, Department of Commerce.

Hottman, C., and R. Monarch. 2020. “A Matter of Taste: Estimating Import Price Inflation across US Income Groups.” *Journal of International Economics* 127:103382.

Jaimovich, N., S. Rebelo, and A. Wong. 2015. “Trading Down and the Business Cycle.” *NBER Working Papers.*

Jaravel, X. 2019. “The Unequal Gains from Product Innovations: Evidences from the US Retail Sector.” *Quarterly Journal of Economics* 134:715–783.

Kamal, F., and R. Monarch. 2016. “Identifying Foreign Suppliers in U.S. Merchandise Import Transactions.” *CES Working Paper.*

Kaplan, G., and S. Schulhofer-Wohl. 2017. “Inflation at the household level.” *Journal of Monetary Economics* 91:19–38.

Katz, L. F., and K. M. Murphy. 1992. “Changes in Relative Wages, 1963-1987 - Supply and Demand Factors.” *Quarterly Journal of Economics* 107:35–78.

Kucheryavyy, K., G. Lyn, and A. Rodríguez-clare. 2020. "Grounded by Gravity: A Well-Behaved Trade Model with Industry-Level Economies of Scale." *Mimeo.*

Lebow, D. E., and J. B. Rudd. 2003. "Measurement Error in the Consumer Price Index: Where Do We Stand?" *Journal of Economic Literature* 41:159–201.

Levinson, A., and J. O'Brien. 2017. "Environmental Engel Curves." *Mimeo.*

Melitz, M. J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71:1695–1725.

Ossa, R. 2015. "Why trade matters after all." *Journal of International Economics* 97:266–277.

Pierce, J. R., and P. K. Schott. 2012. "A Concordance Between Ten-Digit U.S. Harmonized System Codes and SIC/NAICS Product Classes and Industries." *Journal of Economic and Social Measurement* 37:61–96.

Redding, S. J., and D. E. Weinstein. 2017. "Aggregating from Micro to Macro Patterns of Trade." *NBER Working Paper.*

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek. 2015. *Integrated Public Use Microdata Series: Version 6.0.* Minneapolis.

Schott, P. K. 2008. "The relative sophistication of Chinese exports." *Economic Policy* 23:6–49.

Simonovska, I., and M. E. Waugh. 2014. "The elasticity of trade: Estimates and evidence." *Journal of International Economics* 92:34–50.

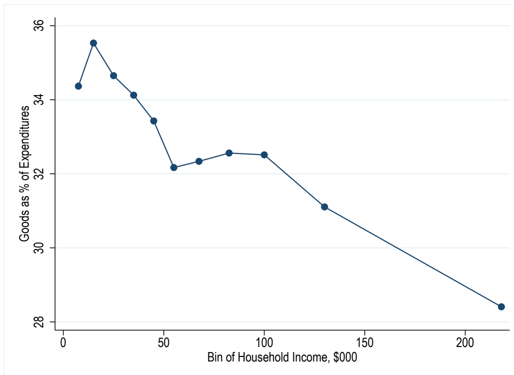
Soderbery, A. 2015. "Estimating import supply and demand elasticities : Analysis and implications." *Journal of International Economics* 96:1–17.

Tintelnot, F., A. K. Kikkawa, M. Mogstad, and E. Dhyne. 2020. "Trade and Domestic Production Networks." *Mimeo.*

Wasi, N., and A. Flaaen. 2015. "The Stata Journal." *Stata Journal* 15:672–697.

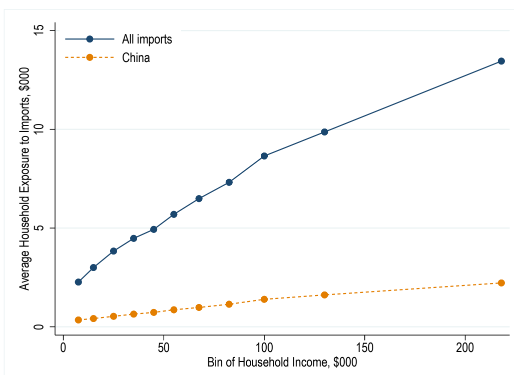
Supplementary Figures and Tables

Figure S1: Share of Spending on Goods across the Income Distribution



Notes: This binned scatterplot shows the relationship between household income and the share of (direct) expenditure on goods using industry-level CEX-IO data.

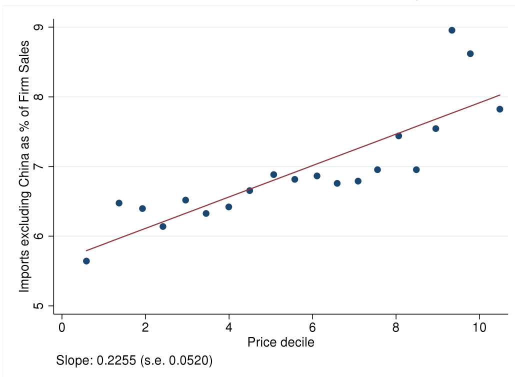
Figure S2: Average Import Expenditure in $1,000 by Income Bin



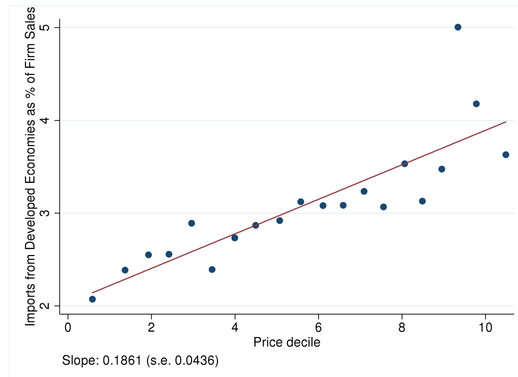
Notes: This binned scatterplot groups CEX panelists into 11 bins by household income before tax. The average value of total (direct and indirect) imports for each bin is reported, in $1,000, based on the industry-level CEX-IO data. The dollar value corresponds to the shares reported in Panel A of Figure 1, rescaled by the average of total expenditures for households within each bin.

Figure S3: The Role of Product Quality in Import Share Heterogeneity, Nielsen-Census Sample

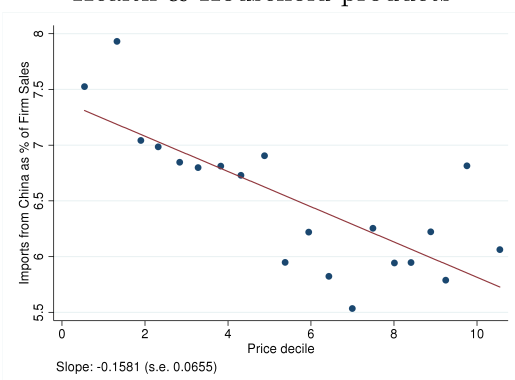
A: Prices and imports excluding China



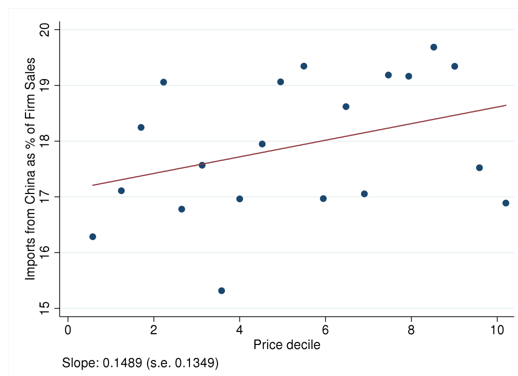
B: Prices and imports from developed economies



C: Prices and imports from China, Health & Household products



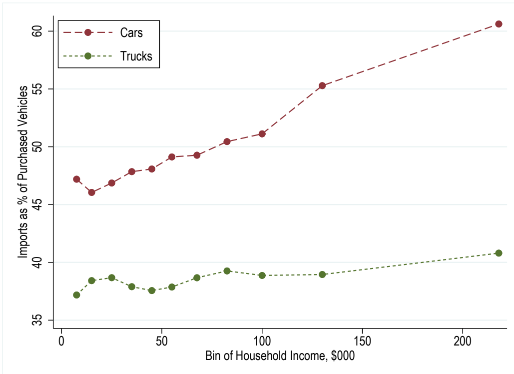
D: Prices and imports from China, General Merchandize



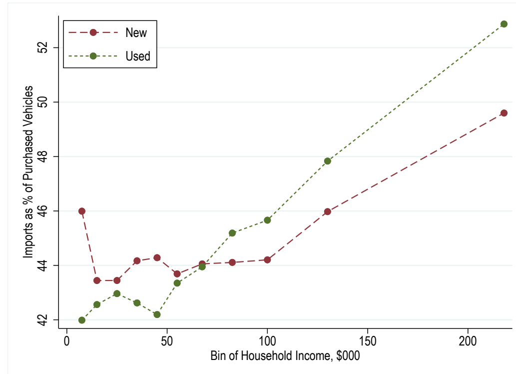
Notes: These binned scatterplots report import shares by decile of barcode prices within product modules for consumer packaged goods. Import shares are computed at the firm level using the Nielsen-Census sample. Product modules which include barcodes with quantity measured in different units (e.g. ounces vs. counts) are decomposed by measurement unit. Firms are weighted by the square-root of their Nielsen sales, and weights are decomposed across barcodes of the same firm proportionally to sales. Fixed effects of modules by year are absorbed.

Figure S4: Imports Shares on Motor Vehicles by Subsamples

A: Cars vs. light trucks

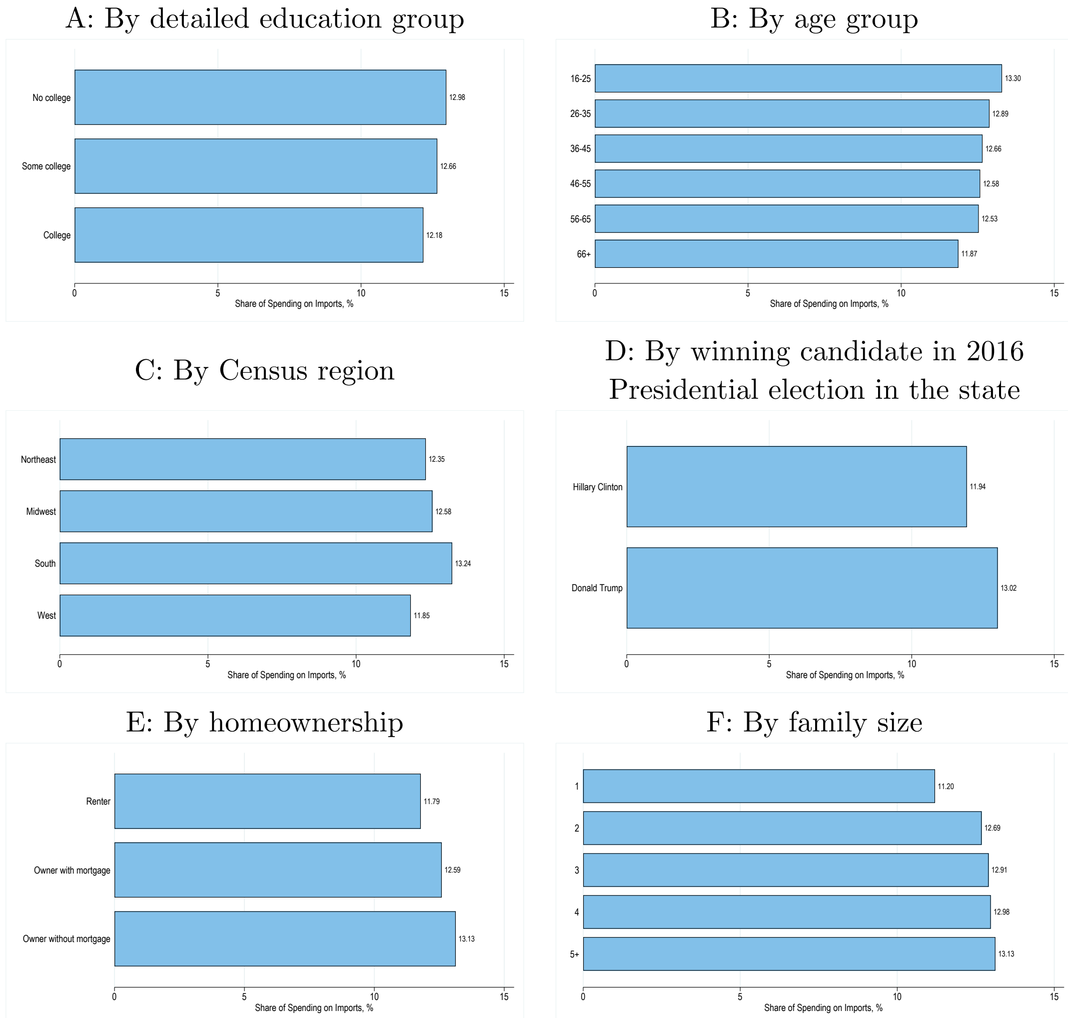


B: New vs. used vehicles



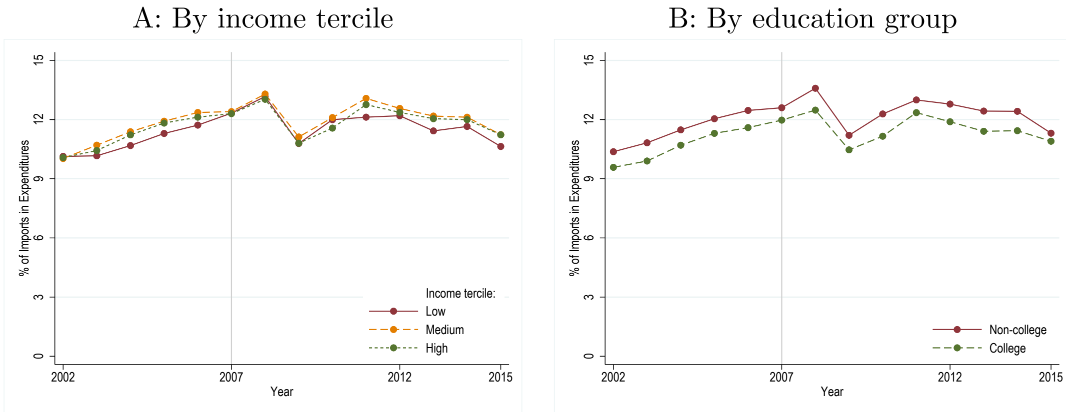
Notes: Panel A reports average import shares for purchases of cars and light trucks separately Panel B instead splits the sample by whether the vehicle was purchased new or used, based on the CEX-Ward's data. Each vehicle in the CEX is assigned a probability of being imported, based on the average import share of the car brand in the Ward's data.

Figure S5: Import Shares across Other Household Groups



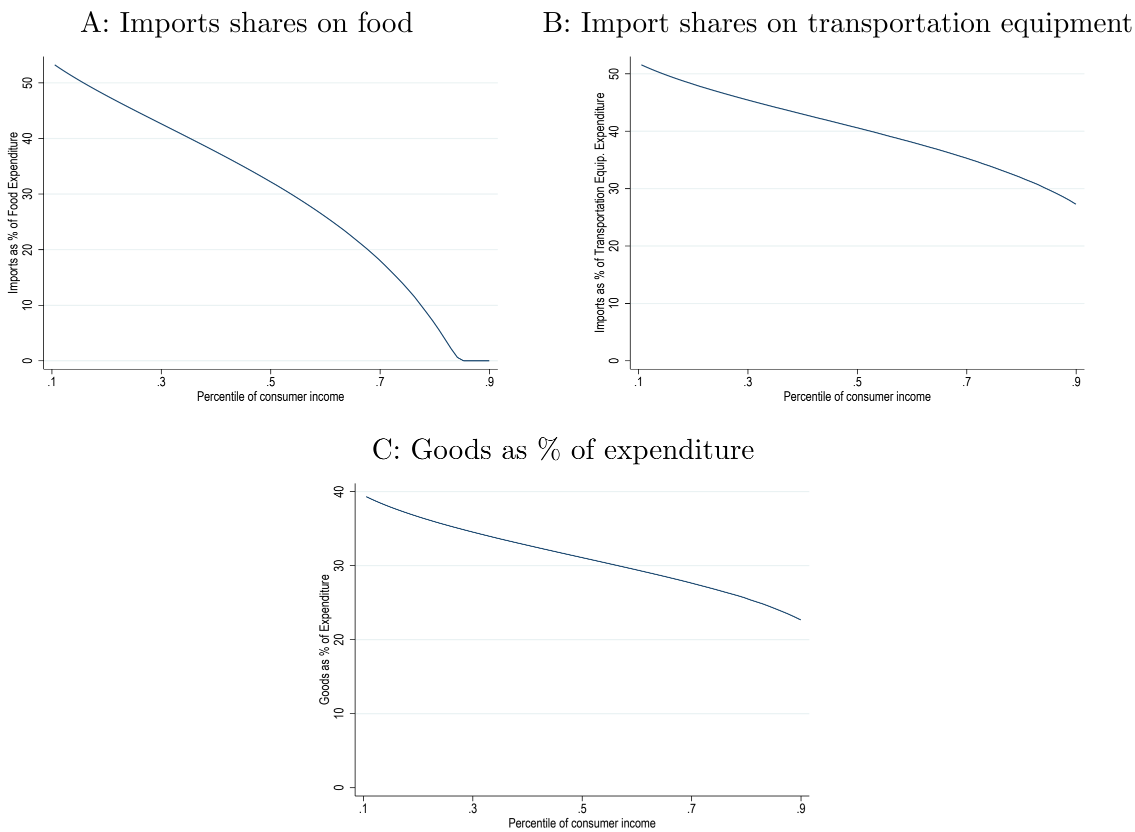
Notes: This figure shows fraction of spending on imports across groups of households, using industry-level CEX-IO data. Indirect spending on imports via imported intermediate inputs is taken into account.

Figure S6: Import Shares by Income and Education Groups over Time



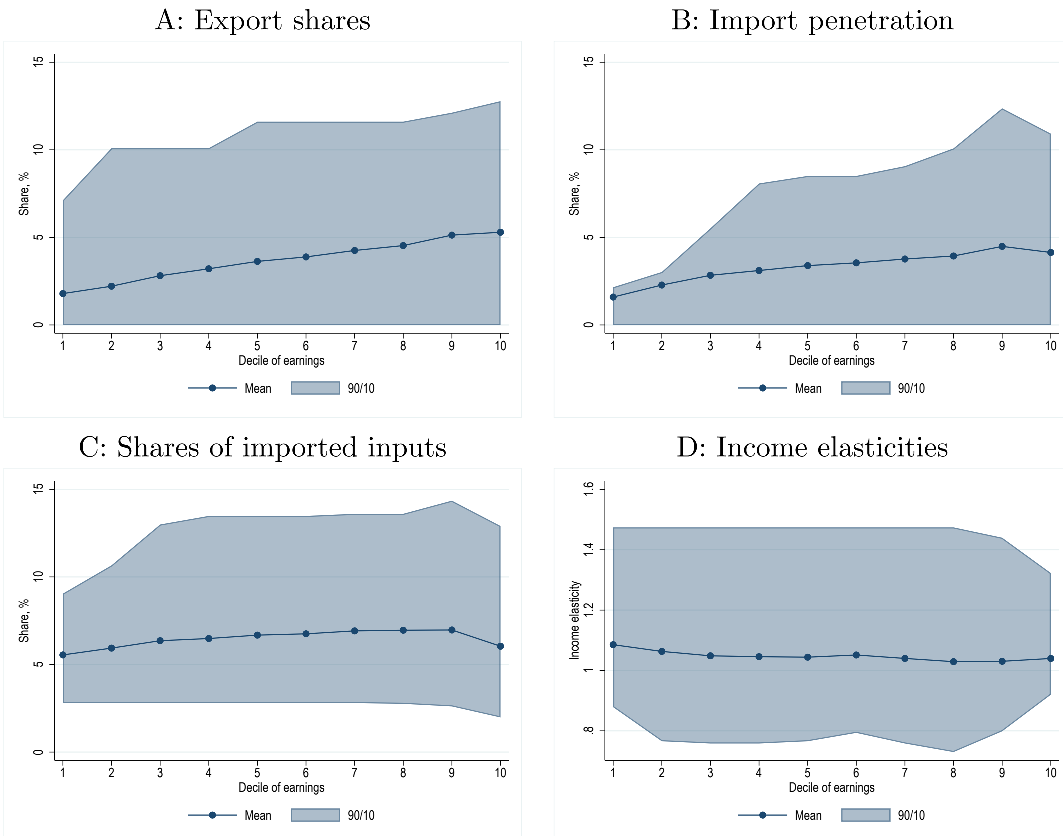
Notes: This figure shows the total fraction of imports in expenditures by demographic group (year-specific terciles of income before tax in Panel A and college education in Panel B) for 2002–2015. For each year, it combines the CEX Integrated Survey with the BEA Summary IO Tables after redefinitions. The methodology is analogous to that of Appendix S.2.1, except that IO industries are more aggregated. We use 73 three-digit commodities from the IO table and separate Non-comparable Imports from the Rest-of-the-World Adjustment. We drop used goods, rest-of-the-world adjustment, and government industries from the final calculation, which results in 71 industries, including 54 final industries matched to the CEX.

Figure S7: Additional Predictions of Fajgelbaum and Khandelwal (2016)



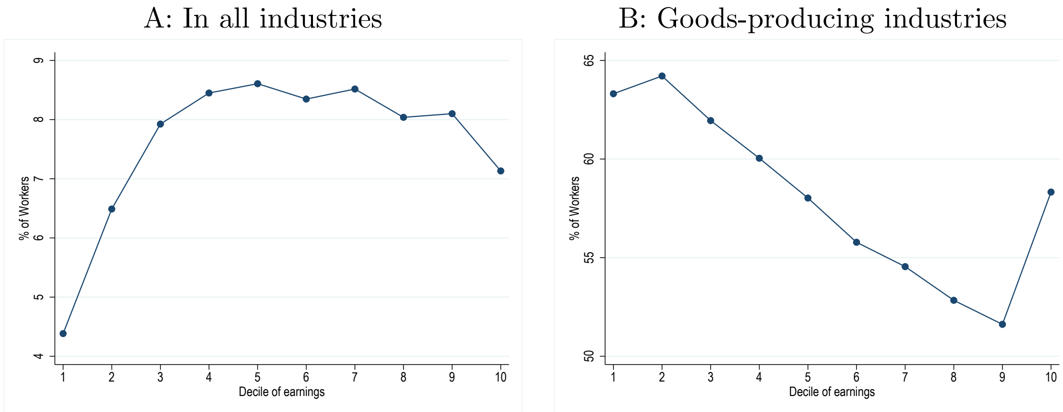
Notes: This figure presents estimates of expenditure shares across the income distribution as in Fajgelbaum and Khandelwal (2016) for the U.S., using international trade data to estimate the parameters of the AIDS demand system. Panels A and B show the expenditure shares on imported varieties within spending on food (ISIC code 3, Panel A) and transportation equipment (ISIC code 15, Panel B). Panel C shows the fraction of goods (ISIC codes 1–16) in total expenditure.

Figure S8: Raw Worker-Level Exposure to Trade



Notes: This figure plots "raw" exposure of workers to several margins of international trade, across and within deciles of initial earnings. Each worker's exposure is given by the corresponding industry variable, and no IO or other adjustments are applied. Each panel reports the average, the 10th percentile, and the 90th percentile across workers in each earnings bin.

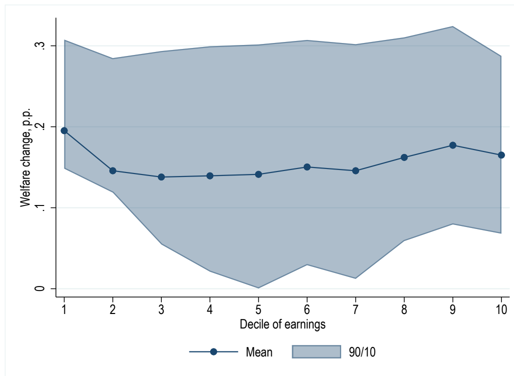
Figure S9: The Share of Losers from a Fall in Trade Costs across the Income Distribution



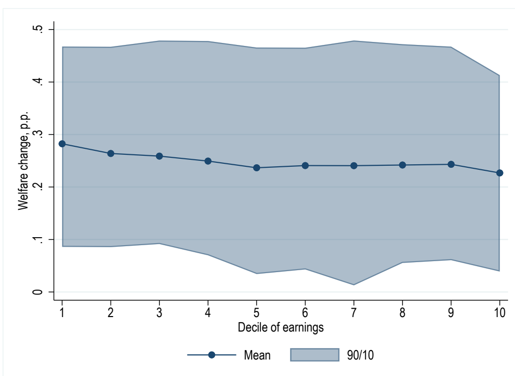
Notes: This figure reports the share of workers with negative equivalent variation in each decile, considering a uniform 10% fall in trade costs. The equivalent variation is computed using Proposition 2. Panel A considers all sectors, while Panel B focuses on goods-producing industries only.

Figure S10: Worker-level Welfare Effects of Trade Liberalizations with Specific Trading Partners

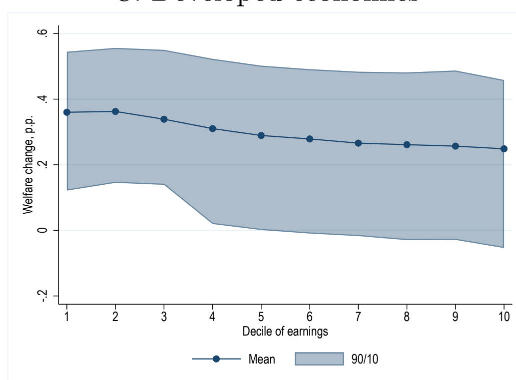
A: China



B: NAFTA



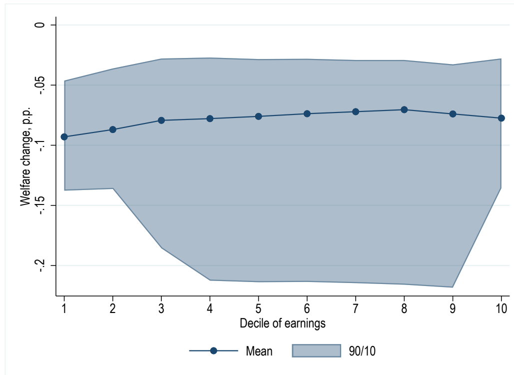
C: Developed economies



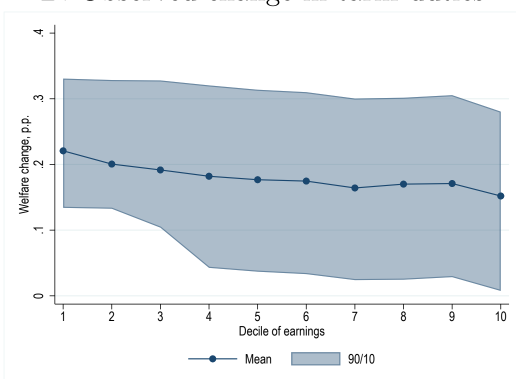
Notes: For the worker-level calibration of Section 5.3, this figure plots the welfare effects of a 10% fall in trade costs for goods imported from specific trading partners: China, NAFTA (Mexico and Canada), and 34 developed economies (OECD members, excluding NAFTA countries, plus Taiwan and Singapore). Each panel reports the average, the 10th percentile, and the 90th percentile across workers in each bin of initial worker earnings.

Figure S11: Worker-level Welfare Effects of Observed Changes in Trade Costs

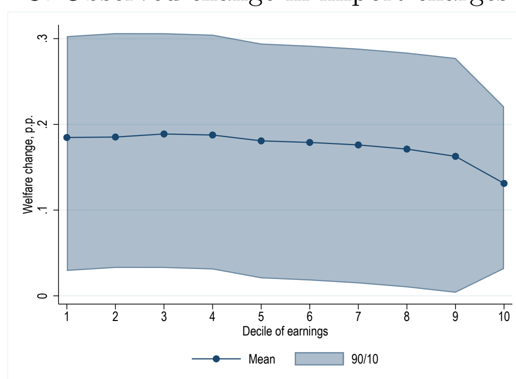
A: 2018 Trump import tariffs



B: Observed change in tariff duties

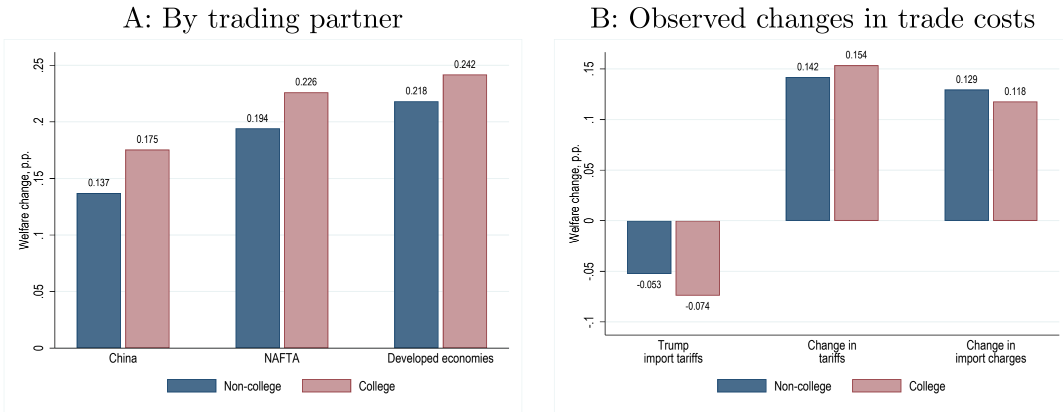


C: Observed change in import charges



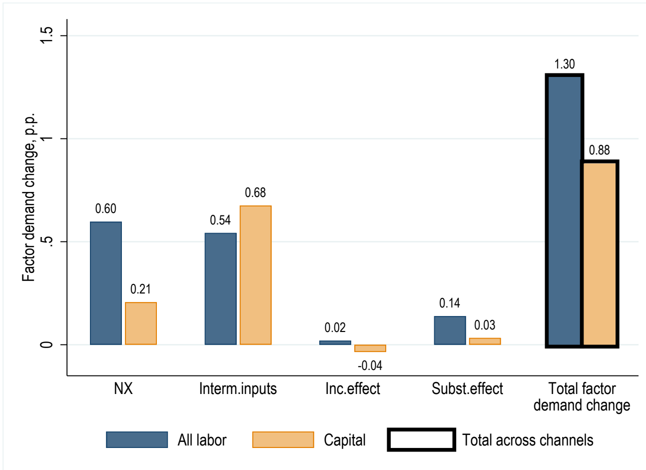
Notes: For the worker-level calibration of Section 5.3, this figure plots the welfare effects from observed shocks in the costs of importing goods. Panel A considers the introduction of trade tariffs by the Trump administration in 2018, Panel B studies observed changes in total U.S. tariff duties between 1992 and 2007, and Panel C examines the impact of changes in import charges (i.e., total transportation and insurance costs) between 1992 and 2007. Appendix S.2.7 describes the methodology.

Figure S12: Welfare Effects of Non-Uniform Trade Shocks across Education Groups



Notes: This figure plots the welfare effects from non-uniform trade shocks across education groups for the calibration of Section 5.4. Panel A considers a 10% fall in trade costs of importing goods from specific countries (China, Mexico and Canada, and 34 developed economies), while Panel B studies the effects of observed trade shocks: the introduction of import tariffs in 2018 by the Trump administration, changes in U.S. tariff duties between 1992 and 2007, and changes in import charges (i.e., total transportation and insurance costs) between 1992 and 2007. Panel A follows Proposition 2, while Appendix S.2.7 describes the methodology for Panel B.

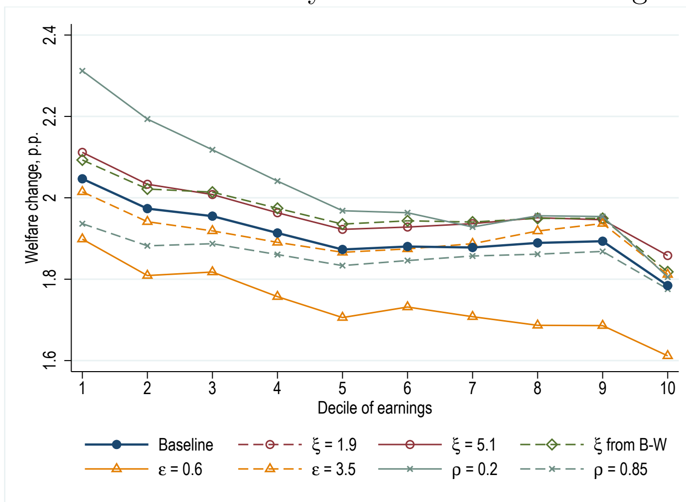
Figure S13: Changes in Factor Demand for Capital Owners and Workers for a Uniform Fall in Trade Costs



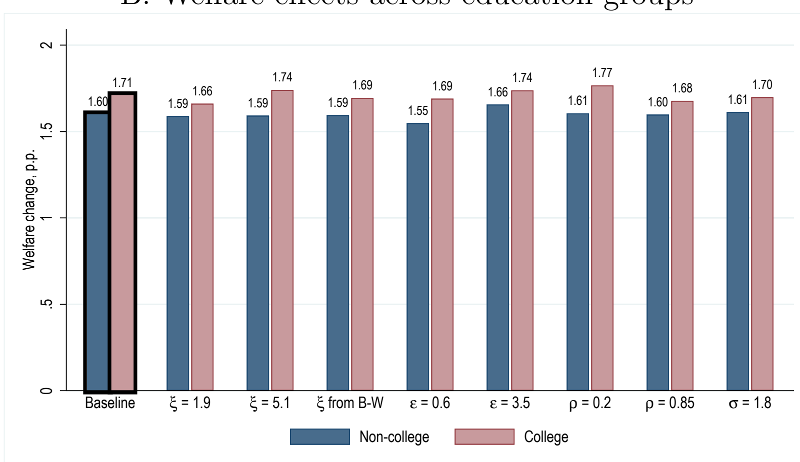
Notes: This figure reports the partial equilibrium change in factor demand for labor and capital for a uniform 10% fall in trade costs, decomposing the change into the several channels as in Proposition 2. The composition of industries in payments to capital owners is obtained from the "Gross operating surplus" row in the IO Table, similarly to how "Compensation of employees" is used for labor (Appendix S.2.4).

Figure S14: Robustness to Choice of Elasticities

A: Welfare effects by decile of initial earnings



B: Welfare effects across education groups



Notes: This figure reports the welfare effects of a 10% uniform fall in trade costs by worker groups, under different assumptions about the relevant elasticities of substitution. Panel A considers the worker-level calibration from Section 5.3, while Panel B focuses on education groups, as in Section 5.4. The baseline from Figures 6 and 7, reproduced here, uses the following elasticities of substitution in demand: across countries of origin within industries, $\xi_j = 3.5$; across industries within manufacturing or services, $\varepsilon_r = 2$; between manufacturing and services, $\rho = 0.6$. Panel B further uses the macro elasticity of substitution between workers with and without a college degree, $\sigma = 1.41$ (this elasticity is not relevant for Panel A). The figure then consider ranges of $\xi_j \in [1.9, 3.5]$, $\varepsilon_r \in [0.6, 3.5]$, $\rho \in [0.2, 0.85]$, and $\sigma \in [1.41, 1.8]$, capturing the values found in the literature (see Section 5.2). We also allow ξ_j to vary across 3-digit IO industries according to the estimates from Broda and Weinstein (2006), labeled "B-W" in the figure.

Figure S15: Examples of Products

Domestic products

A: Plates "Corelle"

UPC 071160 015449
World Kitchen, LLC

B: Plates "MainStays"

UPC 018643 157371
Merrick Engineering, Inc.

Imported products

C: Bed sheets "MainStays", Made in China

UPC 844178 030335
Jiangsu Royal Home USA, Inc.

D: Plates "Better Homes", Made in China

UPC 855602006 567
First Design Global, Inc.

E: Conditioner "Equate Beauty", Made in Canada

UPC 681131 124836
Wal-Mart Stores, Inc.

Notes: These products were photographed in a Wal-Mart store on September 16, 2017. Each barcode (UPC) is split by a space into the firm prefix in the GS1 database and the part which identifies the product within a firm. The country of origin (U.S., China, Canada) is from the product label, whereas the firm information is from the GS1 record corresponding to the barcode prefix.

Table S1: Summary Statistics

	CEX-IO (1)	Nielsen-Census (2)	CEX-Ward's (3)
Coverage	All goods and services	Consumer packaged goods	Motor vehicles
Product space	170 final industries	12,700 firm-years	45 brands
Spending share on imports, %	12.58	11.10	44.40
→ China	1.93	4.15	0.00
→ NAFTA	2.65	1.91	25.90
→ 34 developed economies	3.21	3.10	18.51
Source of expenditures by consumer group	CEX, 2006–2008	Nielsen, 2006–08 and 2011-13	CEX, 2009–15
Source of import shares	BEA IO Table, 2007	Economic Census, LFTTD, 2007, 2012	Ward's, 2009–15

Notes: This table reports summary statistics for the three datasets on consumption and imports used in the paper. Column 1 describes the industry-level data, Column 2 the micro data for consumer packaged goods, and Column 3 the micro data for motor vehicles. The sample size in Column 2 is rounded to the nearest 100 to preserve confidentiality.

Table S2: Classification of Subsectors

Goods	Services
Apparel and leather and allied products	Accommodation and food services
Chemical products	Arts, entertainment, and recreation
Computer and electronic products	Construction*
Electrical equipment, appliances, and components	Educational services
	Finance and insurance
Fabricated metal products	Government
Farms	Health care and social assistance
Food and beverage and tobacco products	Information
Forestry, fishing, and related activities*	Other services, except government
Furniture and related products	Professional, scientific, and technical services
Machinery	Real Estate, rental and leasing
Mining, except oil and gas*	Retail trade*
Miscellaneous manufacturing	Transportation and warehousing
Motor vehicles, bodies and trailers, and parts	Utilities
Nonmetallic mineral products	Wholesale trade*
Oil and gas extraction*	
Other transportation equipment	
Paper products	
Petroleum and coal products	
Plastics and rubber products	
Primary metals*	
Printing and related support activities*	
Support activities for mining*	
Textile mills and textile product mills	
Wood products*	

* Subsectors with zero final personal consumption (in the IO table or in the CEX, or both).
Notes: This table lists subsectors within the goods-producing and service sectors according to the detailed 2007 BEA input-output table. Goods-producing services include agriculture, manufacturing, and mining. Subsectors are defined by the 3-digit input-output codes for goods and 2-digit NAICS codes for services (except Management and Administrative Services, which are included in the Professional, Scientific, and Technical Services).

Table S3: List of Motor Vehicle Brands

Brand code	Brand	N	Brand code	Brand	N	Brand code	Brand	N
FOR	Ford	15,566	KIA	KIA	1,551	ISU	Isuzu	250
CHE	Chevrolet	14,576	LEX	Lexus	1,396	SAA	Saab	197
TOY	Toyota	11,972	MEC	Mercury	1,372	POR	Porsche	181
HON	Honda	8,721	BMW	BMW	1,257	MIN	Mini	175
DOD	Dodge	6,417	SAT	Saturn	1,241	LAN	Land Rover	145
NIS	Nissan	5,466	MRB	Mercedes-Benz	1,168	JAG	Jaguar	140
JEE	Jeep	3,177	ACU	Acura	1,145	ZEN	Zenn	96
GMC	GMC	2,771	CAD	Cadillac	1,129	HUM	Hummer	72
CHR	Chrysler	2,489	MIT	Mitsubishi	930	DAW	Daewoo	33
PON	Pontiac	2,318	LIN	Lincoln	796	FIA	Fiat	25
HYU	Hyundai	2,310	VOV	Volvo	709	SMA	Smart	21
BUI	Buick	2,261	INF	Infiniti	566	MGA	MG	16
MAZ	Mazda	1,858	AUD	Audi	444	TES	Tesla	11
VOK	Volkswagen	1,731	SUZ	Suzuki	347	INT	Intl. Harvester	10
SUB	Subaru	1,674	SCI	Scion	314	AUS	Austin-Healey	4

Notes: This table lists 45 brands in the CEX-Ward's sample on motor vehicles (cars and light trucks) and reports the total number of purchases in the CEX.

Table S4: Import Shares by Education Group

	Levels		College minus non-college	
	College, %	Non-college, %	p.p.	% of average
	(1)	(2)	(3)	(4)
A: Industry data, direct + indirect imports (CEX-IO)				
All countries	12.20	12.84	-0.65	-5.13
China	2.00	1.89	+0.11	+5.95
NAFTA	2.50	2.75	-0.25	-9.24
Developed economies	3.06	3.31	-0.25	-7.90
B: Consumer packaged goods, direct + indirect imports (Nielsen-Census)				
All countries	11.50	10.91	+0.59	+5.35
China	4.02	4.20	-0.18	-4.37
NAFTA	1.97	1.88	+0.09	+4.61
Developed economies	3.47	2.93	+0.55	+17.63
C: Motor vehicles, direct imports only (CEX-Ward's)				
All countries	47.73	42.66	+5.08	+11.43
NAFTA	23.07	27.37	-4.30	-16.60
Developed economies	24.66	15.29	+9.37	+50.65

Notes: This table reports the fraction of imports in expenditure for households with and without a college degree, using in turn the CEX-IO sample, the Nielsen-Census sample, and the CEX-Ward's sample. The difference between the import shares of the two education groups are reported in Column 3 in levels, and in Column 4 as a fraction of average import shares.

Table S5: Import Shares by Education Group and Firm Activity: Manufacturing, Wholesale, and Retail (Nielsen-Census Sample)

	Total imports, all products			Imports from China, Health & Household		
	MFG (1)	WH (2)	RT (3)	MFG (4)	WH (5)	RT (6)
All consumers, %	4.37	5.82	0.30	1.98	3.99	0.28
College minus non-college, p.p.	-0.09	0.62	-0.01	-0.11	-0.21	-0.01
→ Within IO industries	-0.05	0.47	-0.00	-0.10	-0.19	-0.01
N firm-years	12,700	12,700	12,700	3,700	3,700	3,700

Notes: This table estimates the average and differential fraction of imports in spending, decomposed by the main activity of the firm that registered the product: manufacturing (MFG), wholesale (WH), or retail (RT). Other activities are not shown. Each firm is assigned the main activity based on the total payroll of establishments in the corresponding NAICS sectors. Each block of three columns is based on the same data: we decompose import spending into components, without amending the sample.

Table S6: Skill-Bias of Exporters in Census Microdata

	Skill group definition:			
	College graduates	Non-production workers		
	MOPS 2010	CMF 2007	ASM 2010	MOPS 2010
	(1)	(2)	(3)	(4)
Average export share, %	22.84	14.70	19.47	22.84
Differential export share, skilled minus unskilled, p.p.:				
Overall	+5.26	+4.50	+4.52	+5.36
→ Between industries	+4.49	+4.09	+4.51	+5.20
→ Within industries	+0.77	+0.41	+0.01	+0.16
N establishments	33,400	294,200	50,500	33,400

Notes: This table shows the payroll-weighted average export shares (exports as % of sales) for three samples of manufacturing establishments: the 2010 MOPS (Columns 1 and 4), the 2007 Census of Manufactures (Column 2) and the 2010 Annual Survey of Manufactures (Column 3); see Appendix S.2.8 for data description. The table also shows the differential exposure for skilled and unskilled workers and decomposes it into "between" and "within" components across six-digit NAICS industries. Skilled workers are defined as college graduates in column 1 and non-production workers in the other columns.

Table S7: Summary Statistics, Nielsen-Census Sample

	All products	By Product Class: Food	Health & Household	General Merchandize
Spending share of imports, %	11.10	6.92	14.58	27.96
→ Imports from China	4.15	0.88	6.51	17.91
→ Imports from NAFTA	1.91	1.67	2.19	2.74
→ Imports from Developed Economies	3.10	2.42	4.24	4.90
% of Product Class in Total Sales	100.00	67.29	20.24	12.48
N firms	8,200	5,700	2,400	2,000
N firm-years	12,700	9,000	3,700	2,800
N firm-module-years	131,000	88,600	29,800	12,500

Notes: This table reports statistics on imports based on the merged Nielsen-Census sample, within consumer packaged goods overall and for three product classes: Food, Alcohol, and Tobacco ("Food"), Health and Beauty Products and Household Supplies ("Health and household"), and General Merchandize. Imports are measured at the firm level and the summary statistics are computed using the square-root of firms' Nielsen sales as weights. The reported percentage of each product class uses the same weighting scheme. The numbers of observations are rounded to the nearest 100 to preserve confidentiality.

Table S8: Nielsen-Census Matching Rules

	Non-missing exact match	Exact and [fuzzy] match
Rule 1	Zip-9	House, Name, Address, PO Box, Unit, Bldg
Rule 2		House; [Name, Address, PO Box, Unit, Bldg]
Rule 3	Zip-5, House	Name, Address, PO Box, Unit, Bldg
Rule 4		[Name, Address, PO Box, Unit, Bldg]
Rule 5	Zip-5	Name
Rule 6	City	Name, State
Rule 7	State	Name, Entity

Notes: This table lists the rules used to match names and addresses in the Nielsen and Census samples. Each rule requires an exact match and non-missing values of the variables listed in the first column, as well as an exact or probabilistic (fuzzy) match on the variables from the second columns (missing values are allowed). Variables where fuzzy match is allowed are listed in brackets. For fuzzy matching, a 75% threshold is chosen for the match quality score assigned by the `reclink2` package from Wasi and Flaaen (2015).

Table S9: Nielsen-Census Match Statistics

A: Nielsen firms

	2007		2012	
	Firms	% of Sales	Firms	% of Sales
All Nielsen	26,900	100.00	28,600	100.00
Nielsen with size filter	11,000	99.77	12,100	99.82
Matched to SSEL, same year	7,600	83.19	8,900	87.29
Matched to SSEL, any year	8,200	90.76	9,300	91.86
Matched to Economic Census	7,200	88.68	7,800	88.57
Passed consistency filter	6,100	83.02	6,600	83.61

B: Census firms in Food, Alcohol, and Tobacco

	All years	
	Firms	% of Sales
All Census	51,500	100.00
Matched to Nielsen	8,900	78.96
Matched to Nielsen with size filter	5,200	75.57
Passed consistency filter	4,800	58.73

Notes: This table reports the number of firms and the percentage of total sales remaining after each step of the merging process between the Nielsen and Census samples, explained in detail in Appendix S.2.2. Panel A measures these statistics relative to the full Nielsen sample (for 2007 and 2012 Economic Censuses separately), while Panel B measures them relatively to the set of Census firms active in the Food, Alcohol, and Tobacco Manufacturing industries (NAICS codes 311 and 312). The last line of each panel corresponds to the final merged sample, for all firms in Panel A and for those in food, alcohol, and tobacco in Panel B. The numbers of firms are rounded to the nearest 100 to preserve confidentiality.

Table S10: Distribution of Match Types, Nielsen-Census Sample

	% of Matched Firms (1)	% of Sales (2)	% of $\sqrt{}$ Sales (3)
Multi-establishment firms			
Rule 1	10.30	19.72	17.88
Rule 2	4.12	18.99	10.76
Rule 3	5.21	19.86	12.77
Rule 4	3.87	18.54	9.82
Rule 5	2.54	4.12	4.46
Rule 6	1.72	6.80	4.75
Rule 7	1.65	5.11	4.34
Total multi-establishment	29.42	93.14	64.79
Single-Establishment Firms			
Rule 1	33.87	3.42	17.23
Rule 2	10.27	0.87	4.87
Rules 3–7	26.44	2.57	13.12
Total single-establishment	70.58	6.86	35.21

Notes: This table shows the fractions of the Nielsen-Census merged sample corresponding to each of the merging rules, described in Data Appendix S.2.2. Column 1 shows the raw fraction of Nielsen firms in each category, while Column 2 shows the share of total Nielsen sales, and Column 3 weights firms by the square-root of Nielsen sales.

Table S11: Distribution of NAICS Industries, Nielsen-Census Sample

NAICS Industry		% of Firms	% of Sales	% of √ Sales	% of Private Label Brands
Code	Description	(1)	(2)	(3)	(4)
	2-digit NAICS codes				
31-33	Manufacturing	49.78	61.63	57.17	1.21
42	Wholesale	39.37	16.02	29.00	7.90
44-45	Retail	4.80	18.55	8.66	93.74
—	Other	6.04	3.80	5.18	5.19
	3-digit NAICS codes				
311	Food Manufacturing	31.16	36.74	34.78	0.73
312	Beverage and Tobacco Manufacturing	5.73	6.68	6.26	0.30
322	Paper Manufacturing	0.75	4.76	1.96	1.86
325	Chemical Manufacturing	5.36	8.18	6.97	2.79
423	Durable Goods Wholesalers	8.34	2.20	5.86	5.91
424	Nondurable Goods Wholesalers	29.96	15.24	23.05	6.49
445	Food and Beverage Stores	2.24	9.82	4.97	99.10
—	Other	16.44	16.38	16.16	49.69

Notes: Columns 1–3 of this table report the fractions of the Nielsen-Census merged sample corresponding to selected 2- and 3-digit NAICS sectors. Each firm in the Economic Census is classified into the sector where its establishments have the highest total payroll. Column 1 shows the raw fraction of firms in each sector, while Column 2 shows the share of total Nielsen sales, and Column 3 weights firms by the square-root of Nielsen sales. Column 4 measures, for firms in each sector, the sales share of Nielsen barcodes that are classified as private label brands—brands that belong to the retail store. We identify them in the Nielsen data as those which contain "CTL BR" in the barcode description.

Table S12: Nielsen-Census Sample Selection

A: Nielsen Firms

	N	% of Total Sales	Median Sales, $k	% of Sales to College Grads	Mean HH Income, $k
Matched	12,700	83.50	1,904	29.14	67.63
Didn't Match	10,400	16.50	981	30.71	69.75
P-value of t-test				[0.009]	[0.008]
P-value controlling for size				[0.425]	[0.028]

B: Census Firms in Food, Alcohol, and Tobacco

	N	% of Sales	Median Sales, $k	Median Payroll, $k	Median Employment	Mean Skill Intensity
Matched	4,800	58.73	13,303	1,889	54	0.336
Didn't Match	46,600	41.27	606	113	4	0.341
P-value of t-test						[0.744]

Notes: This table compares firms in the matched Nielsen-Census sample to other firms in Nielsen (Panel A) and in the Economic Census (Panel B) which did not find a match, in terms of size, consumer, and producer characteristics. The universe of firms in Panel A is all Nielsen firms that passed the size filter, while in Panel B it is all firms in the Economic Census active in Food, Alcohol, and Tobacco Manufacturing. P-values for t-tests for equality of means between the matched and unmatched samples are shown in brackets. The last row of Panel A performs such t-test controlling for a quadratic polynomial in log firm sales. The numbers of firms are rounded to the nearest 100 and medians are computed as geometric means of the 45 and 55 percentiles to protect confidentiality.