Appendix A generalizes the utility function in the original paper, Appendix B describes the data, Appendix C presents robustness checks, and Appendix D presents Monte Carlo simulations.

APPENDIX A: AN ALTERNATIVE FORM FOR THE UTILITY FUNCTION

Utility function (1) in the main paper captures the empirical finding that poor households spend most of their income on basic goods, while rich ones spend it on luxuries. The same parameter $\sigma_\tau$, however, controls both the elasticity of substitution across goods and the income elasticity of demand—two objects that need not be linked in reality. A more general form for preferences relaxes this link:

\[
\sum_{\tau=1}^{S} \left\{ \alpha_\tau \frac{\sigma_\tau}{\gamma_\tau(\sigma_\tau - 1)} \left[ \int_{0}^{1} q(j_\tau)^{\sigma_\tau-1/\sigma_\tau} dj_\tau \right]^{\gamma_\tau} \right\}.
\]

Denote by $\lambda$ the consumer’s Lagrange multiplier and by $P_\tau$ the CES price index for goods of type $\tau = 1, \ldots, S$. I consider two (not exhaustive, but instructive) cases.

**CASE 1**—$\gamma_\tau = \sigma_\tau / (\sigma_\tau - 1)$ for All $\tau$: The first-order conditions imply that for all $\tau$, $\lambda = \frac{\alpha_\tau}{P_\tau}$ if $q(j_\tau) > 0$ for some $j_\tau$. Since these equalities do not hold simultaneously for arbitrary prices, the consumer only demands goods of the type with the highest value for $\frac{\alpha_\tau}{P_\tau}$. Importantly, the Lagrange multiplier and hence consumer demand do not depend on income: preferences are homothetic.

**CASE 2**—$\gamma_\tau \neq \sigma_\tau / (\sigma_\tau - 1)$ for All $\tau$: Spending on any two types of goods, 1 and 2, satisfies

\[
\frac{x_1}{x_2} = \lambda^{\varphi_1 - \varphi_2} \left[ (\alpha_1)^{\varphi_1} P_1^{\phi_1} / (\alpha_2)^{\varphi_2} P_2^{\phi_2} \right],
\]

where $\varphi_\tau = -\sigma_\tau + \frac{\sigma_\tau(1-\sigma_\gamma)(\gamma_\gamma-1)}{\sigma_\tau + \gamma_\gamma - \sigma_\gamma \gamma_\gamma}$ and $\phi_\tau = \frac{(1-\sigma_\gamma)\gamma_\gamma}{\sigma_\gamma + \gamma_\gamma - \sigma_\gamma \gamma_\gamma}$ for $\tau = 1, 2$. As in equation (2), the term in brackets determines the level of $x_1/x_2$, and $(\lambda^{\varphi_1 - \varphi_2})$ determines how it changes with consumer income. Note, however, that the parameters added to equation (S1) relative to the original utility function are not
identified. Consider the case with two types, $A$ and $B$. For any set of prices $P$, and parameters $\{\sigma_A, \sigma_B, \gamma_A, \gamma_B\}$, the parameter $\alpha_A$ can be judiciously chosen to match any level of the ratio $x_A/x_B$. The rate of change of $x_A/x_B$, in turn, is governed by the exponent of the Lagrange multiplier, $(\varphi_A - \varphi_B)$. Parameters $\sigma_A, \sigma_B, \gamma_A$, and $\gamma_B$, therefore, all play the same role, and only one of them is sufficient to pin down $(\varphi_A - \varphi_B)$; the rest can be normalized. In the paper, $\gamma_A$ and $\gamma_B$ are set to 1.

In sum, utility function in the paper assumes that the type of good with a high $\sigma$, has a high income elasticity of demand and a high elasticity of substitution across goods. Case 2 shows that this assumption is not necessary for any of the results. Without changing predicted trade flows, a different normalization of the parameters may imply that the type of good with a high income elasticity of demand has a low elasticity of substitution across goods.

APPENDIX B: DATA

I use data on bilateral merchandise trade flows in 2000 from the U.N. Comtrade data base (United Nations (2008)). In compiling the data, I give precedence to trade flows reported by the importing country. If the importer’s report is not available, I use the exporter’s. I keep in the sample only countries with matching data on population and GDP from the World Bank (2008).¹ Countries also report total trade flows to the World Bank. These flows should be (weakly) larger than the sum of trade flows in the U.N. Comtrade data base, because the U.N. Comtrade data base does not contain all countries. Accordingly, I exclude countries whose total trade flows in the U.N. Comtrade data base are more than 20% larger than in the World Bank data. Seventeen countries and 0.3% of world trade are excluded using this criterion.²

The resulting data comprise 162 countries and 95% of world trade in 2000. Of these countries, 145 directly report trade to the United Nations. All trade flows to and from reporting countries are observed, and trade flows between the remaining 17 countries (marked with a dagger in Table S.I) are missing. Hence, of all possible importer–exporter pairs, 25,810 (= 162² − 17² − 145) are observed and 272 (= 17² − 17) are missing.

¹Neither the United Nations nor the World Bank officially reports statistics for Taiwan, but in practice, the U.N. country classification “Other Asia, not elsewhere specified” (code 490) refers to Taiwan. Data on Taiwan’s population and income, in turn, are taken directly from the Taiwanese government’s website http://eng.stat.gov.tw.
²These countries are Brunei Darussalam, Comoros, Djibouti, Gabon, Georgia, Guinea-Bissau, Guatemala, Honduras, Kiribati, Moldova, Mauritania, Panama, Sierra Leone, Timor-Leste, St. Vincent and Grenadine, Vanuatu, and Samoa.
## TABLE S.I
### LIST OF COUNTRIES\(^{a}\)

<table>
<thead>
<tr>
<th>Country</th>
<th>Country</th>
<th>Country</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>Denmark</td>
<td>Lebanon</td>
<td>Serbia and Montenegro</td>
</tr>
<tr>
<td>Algeria</td>
<td>Dominica</td>
<td>Lesotho</td>
<td>Seychelles^†</td>
</tr>
<tr>
<td>Angola^†</td>
<td>Dominican Republic</td>
<td>Libya^†</td>
<td>Singapore</td>
</tr>
<tr>
<td>Antigua and Barbuda</td>
<td>Ecuador</td>
<td>Lithuania</td>
<td>Slovak Republic</td>
</tr>
<tr>
<td>Argentina</td>
<td>Egypt, Arab Rep.</td>
<td>Luxembourg</td>
<td>Slovenia</td>
</tr>
<tr>
<td>Armenia</td>
<td>El Salvador</td>
<td>Macao</td>
<td>Solomon Islands^†</td>
</tr>
<tr>
<td>Australia^*</td>
<td>Equatorial Guinea^†</td>
<td>Macedonia</td>
<td>South Africa</td>
</tr>
<tr>
<td>Austria^*</td>
<td>Eritrea</td>
<td>Madagascar</td>
<td>Spain^*</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>Estonia</td>
<td>Malawi</td>
<td>Sri Lanka^†</td>
</tr>
<tr>
<td>Bahamas, The</td>
<td>Ethiopia</td>
<td>Malaysia</td>
<td>St. Kitts and Nevis</td>
</tr>
<tr>
<td>Bahrain</td>
<td>Fiji</td>
<td>Maldives</td>
<td>St. Lucia</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>Finland^*</td>
<td>Mali</td>
<td>Sudan</td>
</tr>
<tr>
<td>Barbados</td>
<td>France^*</td>
<td>Malta</td>
<td>Suriname</td>
</tr>
<tr>
<td>Belarus</td>
<td>French Polynesia</td>
<td>Mauritius</td>
<td>Swaziland</td>
</tr>
<tr>
<td>Belgium^*</td>
<td>Gambia, The</td>
<td>Mexico</td>
<td>Sweden^*</td>
</tr>
<tr>
<td>Belize</td>
<td>Germany^*</td>
<td>Mongolia</td>
<td>Switzerland</td>
</tr>
<tr>
<td>Benin</td>
<td>Ghana</td>
<td>Morocco</td>
<td>Syrian Arab Republic</td>
</tr>
<tr>
<td>Bhutan^†</td>
<td>Greece^*</td>
<td>Mozambique</td>
<td>Taiwan^†</td>
</tr>
<tr>
<td>Bolivia</td>
<td>Grenada</td>
<td>Namibia</td>
<td>Tajikistan</td>
</tr>
<tr>
<td>Bosnia and Herzegovina^†</td>
<td>Guinea</td>
<td>Nepal</td>
<td>Tanzania</td>
</tr>
<tr>
<td>Botswana</td>
<td>Guyana</td>
<td>Netherlands^*</td>
<td>Thailand</td>
</tr>
<tr>
<td>Brazil</td>
<td>Haiti^†</td>
<td>New Caledonia</td>
<td>Togo</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>Hong Kong</td>
<td>New Zealand^*</td>
<td>Tonga</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>Hungary</td>
<td>Nicaragua</td>
<td>Trinidad and Tobago</td>
</tr>
<tr>
<td>Burundi</td>
<td>Iceland</td>
<td>Niger</td>
<td>Tunisia</td>
</tr>
<tr>
<td>Cambodia</td>
<td>India</td>
<td>Nigeria</td>
<td>Turkey</td>
</tr>
<tr>
<td>Cameroon</td>
<td>Indonesia</td>
<td>Norway^*</td>
<td>Turkmenistan</td>
</tr>
<tr>
<td>Canada^*</td>
<td>Iran, Islamic Rep.</td>
<td>Oman</td>
<td>Uganda</td>
</tr>
<tr>
<td>Cape Verde</td>
<td>Ireland</td>
<td>Pakistan</td>
<td>Ukraine</td>
</tr>
<tr>
<td>Central African Republic</td>
<td>Israel</td>
<td>Papua New Guinea</td>
<td>United Arab Emirates</td>
</tr>
<tr>
<td>Chad^†</td>
<td>Italy^*</td>
<td>Paraguay</td>
<td>United Kingdom^*</td>
</tr>
<tr>
<td>Chile</td>
<td>Jamaica</td>
<td>Peru</td>
<td>United States^*</td>
</tr>
<tr>
<td>China</td>
<td>Japan^*</td>
<td>Philippines</td>
<td>Uruguay</td>
</tr>
<tr>
<td>Colombia</td>
<td>Jordan</td>
<td>Poland</td>
<td>Uzbekistan^†</td>
</tr>
<tr>
<td>Congo, Dem. Rep.^†</td>
<td>Kazakhstan</td>
<td>Portugal^*</td>
<td>Venezuela</td>
</tr>
<tr>
<td>Congo, Rep.^†</td>
<td>Kenya</td>
<td>Qatar</td>
<td>Vietnam</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>Korea, Rep.</td>
<td>Romania</td>
<td>Yemen, Rep.^†</td>
</tr>
<tr>
<td>Cote d’Ivoire</td>
<td>Kuwait</td>
<td>Russian Federation</td>
<td>Zambia</td>
</tr>
<tr>
<td>Croatia</td>
<td>Kyrgyz Republic</td>
<td>Rwanda^†</td>
<td>Zimbabwe</td>
</tr>
<tr>
<td>Cyprus</td>
<td>Lao PDR^†</td>
<td>Saudi Arabia</td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td>Latvia</td>
<td>Senegal</td>
<td></td>
</tr>
</tbody>
</table>

^aThe dagger (†) denotes countries that do not report trade to the United Nations; the asterisk (∗) denotes OECD countries.

These data used in the estimation contain total trade for 145 countries. To construct Figures 3(a), 4(a), and 5(a), I complement them with data from the World Bank (2008) on the total value of merchandise trade flows for 16 of the
17 nonreporting countries (Taiwan is missing). For the 145 reporting countries, the correlation between trade shares in the World Bank and in the U.N. Comtrade data is 0.95. This use of extraneous data is a small out-of-sample check on the model, and it does not change any of the empirical results.

The U.N. Comtrade data are available until 2005. I use the year 2000 for two reasons. First, there are more and more reliable data for 2000. Twenty-four countries that report trade for 2000 have not yet reported for 2005, and the inconsistencies between the U.N. Comtrade and the World Bank data are larger in 2005 than in 2000. Second, several countries trade close to or more than 100% of their GDP in 2005. The simple model of Section 2 does not account for trade flows beyond 100% of a country’s GDP; only the model with intermediate inputs in Section 5.2 does. But as discussed there, intermediate inputs do not qualitatively change the analysis: they make the model computationally more demanding and they compromise the identification of parameters. Still, it is worth noting that all the stylized facts of the data exploited in Section 3 also hold in the 2005 data.

All main results are robust to the use of data from 1999 or 2001, instead of 2000, or to the use of data from Feenstra, Lipsey, Deng, Ma, and Mo (2005) or from the International Monetary Fund (2008), instead of the U.N. Comtrade data base.

APPENDIX C: ROBUSTNESS

C.1. Weighting

Santos Silva and Tenreyro (2006) discussed the weighting of observations in trade. Trade flows among large countries presumably have large variances because their values are large, and trade flows among small, typically poor, countries may have large measurement errors. To account for size in the empirical Section 3, I simply divide trade flows by the product of the importer and exporter total income, \( z_{ni} = \frac{X_{ni}}{X_nX_i} \). Efficiency gains from estimating an optimal weighting matrix with NLLS are, to my knowledge, not established because numerical errors from the first stage enter the estimated matrix. In addition, here, it would require ad hoc assumptions on the structure of the matrix, since error terms are heteroskedastic and clustered.

I experiment with several alternative weights. For all weights, the message of the empirical results remains: The new model is closer to the data than the EK model because it reconciles the large volume of trade among rich countries with the small volume among small, poor countries. Normalizing trade flows to \( z_{ni} = \log(1 + X_{ni}) \) does not change the results. The results change only in

---

3 Even in linear estimations, Altonji and Segal (1996) used Monte Carlo simulations to show that the gains from estimating an optimal weighting matrix are not clear in small samples.

4 Gravity-type regressions typically involve taking the logarithm of trade flows. I take the logarithm of 1 plus trade flows because 7034 of the 25,810 observations are zeroes.
specifications that put more weight on trade flows among large countries. For example, if \( z_{ni} = \frac{X_{ni}}{(X_nX_i)^{1/2}} \), the EK model overestimates trade of small countries, instead of underestimating trade of rich countries, as in Section 3. The average trade share of the 30 smallest countries is 92% in the EK model, 58% in the new model, and 38% in the data. If even more weight is put on large countries (e.g., \( z_{ni} = X_{ni} \)), then the optimization algorithm captures only trade flows among large rich countries; parameter estimates approach those of the OECD subsample and the difference between the two models narrows.

C.2. Normalization of Parameters \( \theta_A \) and \( \sigma_A \)

In the paper, I estimate the model by fixing \( \theta_A = 8.28 \) and \( \sigma_A = 5 \) in specification 1 (equation (13) of the paper) and \( \theta_A = 8.28 \) in specification 2 (equation (15) of the paper). Table S.II shows the parameter estimates of both specifications when \( \theta_A \) equals 3.60 and 12.86, the other estimates in Eaton and Kor-

### TABLE S.II

**ESTIMATES OF THE NEW MODEL WITH DIFFERENT VALUES FOR \( \theta_A \) AND \( \sigma_A \)**

<table>
<thead>
<tr>
<th>( \theta_A )</th>
<th>Specification 1</th>
<th>Specification 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_A )</td>
<td>8.28</td>
<td>8.28</td>
</tr>
<tr>
<td>3.60</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.40</td>
<td>1.37</td>
</tr>
<tr>
<td>(0.05)</td>
<td>(0.03)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>1.61</td>
<td>1.26</td>
<td></td>
</tr>
<tr>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>1.38</td>
<td>1.20</td>
<td></td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>0.40</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>0.18</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>( \gamma_3 )</td>
<td>−0.01</td>
<td>−0.01</td>
</tr>
<tr>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>−0.01</td>
<td>−0.01</td>
<td></td>
</tr>
<tr>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

| Border | 0.98 | 0.99 |
| (0.07) | (0.08) | (0.08) |
| 1.01 | 0.99 |
| (0.07) | (0.07) | (0.05) |
| 0.95 | 0.93 |

| Language | 0.94 | 0.93 |
| (0.05) | (0.06) | (0.06) |
| 0.91 | 0.94 |
| (0.05) | (0.05) | (0.04) |
| 0.90 | 0.93 |

| Trade agreement | 1.20 | 1.23 |
| (0.12) | (0.12) | (0.14) |
| 1.24 | 1.22 |
| (0.14) | (0.11) | (0.09) |
| 1.24 | 1.21 |

| \( \theta_B \) | 13.73 | 14.67 |
| (1.02) | (0.80) | (0.71) |
| 9.49 | 19.65 |
| (1.01) | (1.01) | (1.39) |
| 15.02 | 26.01 |

| \( (\alpha_A)^{1/\sigma_A} \) | 0.68 | 0.76 |
| (0.04) | (0.02) | (0.03) |
| 0.77 | 0.82 |
| (0.03) | (0.03) | (0.01) |

| \( \sigma_B \) | 1.00 | 1.91 |
| (0.03) | (0.07) | (0.05) |
| 1.22 | 1.25 |
| (0.03) | (0.03) | (0.03) |

| \( R^2 \) | 42% | 42% |
| (0.05) | (0.05) | (0.05) |
| 42% | 42% |
| (0.05) | (0.05) | (0.05) |
| 67% | 67% |

---

\( a\)Standard errors given in parentheses are clustered by importer and exporter. The number of observations is 25,810 in all cases.
Consistent with Section 3.1.2, an increase in $\theta_A$ decreases parameters $\gamma_1$ and $\gamma_2$ and increases $\theta_B$. Estimates of $Y$ and $\theta_B$ are again similar in both specifications, suggesting that the model is correctly specified. The table also presents results from specification 1 with $\sigma_A = 2.0$ and 8.0, arbitrary values satisfying $1 < \sigma_A < (\theta_A + 1)$. An increase in $\sigma_A$ increases $\sigma_B$. In all cases of Table S.II, predicted flows do not change, so neither the $R^2$ nor the stylized facts of Section 3 change.

Direct measures of trade costs can help pin down parameter $\theta_A$. For example, Anderson and Van Wincoop (2004) estimated that trade costs across OECD countries are equivalent to an ad valorem tax of 74%. In the model, trade costs across OECD countries average 73% when $\theta_A = 12.86$. This result, however, is limited because trade costs have large measurement errors and vary a lot across goods and countries.

**APPENDIX D: MONTE CARLO SIMULATIONS**

To check for identification in the first empirical specification (equation (13) in the paper), I conduct Monte Carlo simulations. I make a random draw for each of the estimated parameters, {$Y, \alpha_A, \sigma_B, \theta_B$}, simulate data with the deterministic model of Section 2, and then run the optimization algorithm on simulated data.\(^5\) I repeat this procedure 50 times and compare the parameter estimates with the original parameter draws. The parameters that distinguish the new model from ŘK, $\alpha_A$, $\sigma_B$, and $\theta_B$, are identified with a high degree of precision: in 98% of the simulations, the estimated parameter is within a 5% distance from its original draw. Identification of the iceberg cost parameters $Y$ is weaker because they are correlated. Still, in 84% of simulations the estimated parameters are within a 5% distance from the original draw.

**REFERENCES**


\(^5\)For each parameter, I randomize over a uniform distribution with support $\gamma_1 \in [0.5, 2]$, $\gamma_2 \in [0.05, 0.5]$, $\gamma_3 \in [-0.1, 0]$, {$\gamma_{border}, \gamma_{language}, \gamma_{trade \ agreement}$} $\in [0.5, 1.2]$, $(\alpha_A)^{1/\sigma_A} \in [0.5, 0.9]$, $\sigma_B \in [1.0, 8.]$, and $\theta_B \in [5, 20]$. 


Manuscript received January, 2009; final revision received June, 2010.