Quality, Variable Markups, and Welfare: A Quantitative General Equilibrium Analysis of Export Prices*

Haichao Fan† Fudan

Yao Amber Li‡ HKUST

Sichuang Xu§ HKUST

Stephen R. Yeaple¶ PSU, NBER and CESifo

This version: February 16, 2019

Abstract

Modern trade models attribute the dispersion of international prices to physical and man-made barriers to trade, to the pricing-to-market by heterogeneous producers and to differences in the quality of output offered by firms. This paper presents a general equilibrium model that incorporates all three of these mechanisms. Our model allows us to confront Chinese firm-level data on the prices charged and revenues earned across markets. We show that all three mechanisms are necessary to fit the distribution of prices and revenues across firms and markets. Accounting for endogenous quality heterogeneity across markets and firms is shown to be critical for the welfare implications of trade and for the response of prices to trade and tariff shocks.

JEL classification: F12, F14

Keywords: quality, variable markups, export price, welfare, "Washington Apples" effect, non-homothetic preferences, specific trade costs, heterogeneous firms

*We are very grateful to the editor, Andrés Rodríguez-Clare, and two anonymous referees for their insightful comments. We thank Costas Arkolakis, Sam Kortum, Andreas Moxnes, and the participants at the 2017 HKUST Conference on International Economics (June, 2017) for helpful discussion. We gratefully acknowledge the financial support from the National Science Foundation grant SES-1360209, the Natural Science Foundation of China (No.71603155), the Research Grants Council of Hong Kong, China (General Research Funds and Early Career Scheme GRF/ECS Project No.646112), and the self-supporting project of Institute of World Economy at Fudan University. All errors are our own.

†Fan: Institute of World Economy, School of Economics, Fudan University, Shanghai, China. Email: fan_haichao@fudan.edu.cn.

‡Li: Department of Economics and Faculty Associate of the Institute for Emerging Market Studies (IEMS), Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR-PRC. Email: yaoli@ust.hk. Research Affiliate of the China Research and Policy Group at University of Western Ontario.

§Xu: Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR-PRC. Email: sxuaf@connect.ust.hk.

¶Yeaple: Corresponding author. Department of Economics, Pennsylvania State University. Email: sry3@psu.edu. Research Associate at National Bureau of Economic Research and Research Affiliate at Ifo Institute.

1 Introduction

The literature on quantitative general equilibrium models has blossomed in recent years. The popularity of these models is driven by their simplicity, by their ease of calibration, and by their flexibility to be adapted for the analysis of the impact of a wide variety of policies. Further, as shown by Eaton, Kortum and Kramarz (2011) these models can successfully confront firm-level microdata on the distribution of sales within and across countries. One feature of the microdata that has received less attention in the development of quantitative general equilibrium models is the joint distribution of firm-level prices and sales within and across countries. As has been shown in existing descriptive work (e.g. Manova and Zhang, 2012), firms from a given source country charge very different prices across countries.

In this paper we develop a simple quantitative general equilibrium model with heterogeneous firms that has been designed to confront the joint distribution of firm-level prices and sales. Variation in prices within-firm, across-country stem from the interaction between trade costs that vary between countries, firms' decisions to price-to-market, and firms' endogenous provision of goods of different quality to different countries. Our model includes all three of these features. With respect to trade cost, we explicitly allow for both standard iceberg (ad-valorem) trade costs and specific (fixed per unit) trade costs. This is natural because both types of trade costs are likely to be a feature of the constraints facing exporters in the real world and because the interaction between the two types of trade costs has been shown to affect the quality decision of firms (Hummels and Skiba, 2004).

We also allow firms to choose the quality of goods that they provide to each market that they serve. We assume that the marginal cost of production is increasing in output quality and decreasing in firm productivity. Because specific trade costs are not increasing in the quality of goods sold, firms can lower their cost of serving markets with high specific trade costs by upgrading quality, and the incentive to do this is rising in a firm's productivity because these firms sell the largest number of units. Hence, our specification delivers a "Washington-Apples" effect that varies in strength across both countries and firms and so provides a mechanism to fit the joint distribution of prices and revenues.[1]

With respect to pricing-to-market, we follow Jung, Simonovska and Weinberger (2019) by assuming that CES-like preferences that have been generalized to allow for an endogenous "choke price". Firms in our model first minimize quality-adjusted marginal costs and then set quality-adjusted prices to maximize profits in each market that they serve. While the correlation between quality-adjusted prices and quality-adjusted revenue will be negative due to the optimal markup choices of the firm, the correlation between observed (unadjusted) prices and (unadjusted) revenues will be positive as in the data.

Our paper has novel implications for the estimation of gravity equations. A large class

[1] Our formulation adapts Feenstra and Romalis (2014) to be more in line with the initial formulation in Hummels and Skiba (2004). Feenstra and Romalis (2014) do not adapt their mechanism to confront firm-level data.

of models generates gravity equations in which the elasticity of trade flows with respect to trade costs reveals key structural parameters (Arkolakis, Costinot and Rodríguez-Clare, 2012; Arkolakis et al., 2019). Our model also generates a gravity equation in which the appropriate measure of trade costs is the geometric average of specific and iceberg trade costs where the weights reflect the elasticity of marginal cost with respect to quality. In standard models a common way to estimate the trade elasticity using tariffs, which are generally ad-valorem, as a measure of trade costs (e.g. Head and Mayer, 2014). In our framework with specific trade costs, this calibration strategy necessarily leads to an underestimate of the key macro elasticity.

We calibrate our model to aggregate trade flows (gravity) and to the joint distribution of firm-country level price and sales from Chinese customs data. By selectively shutting down model mechanisms and recalibrating, we show how various mechanisms help fit the joint distribution of prices. Only our model with endogenous quality and pricing to market can accommodate the positive correlation between firm-level prices and sales in the data. Moreover, in attempting to fit the positive correlation between firm-country level prices and sales revenues, special cases of our model that lack either pricing-to-market or an endogenous quality mechanism generate unreasonable estimates for key parameters.

Our model also contributes to our understanding of the response of prices to trade cost shocks. Much recent work analyzes the markup responses of firms to changes in trade policy (e.g. De Loecker et al., 2016; Jung, Simonovska and Weinberger, 2019). In our setting, shocks to trade costs affect firm-level prices through multiple mechanisms. On the one hand, firms respond to any shock to quality-adjusted marginal costs by changing their markups. On the other hand, firms also adjust the quality of their output and this induces a price response as quality-adjusted marginal costs change.

To illustrate the potential for quality adjustment to be confused for adjustment in markups, we consider a comparative static exercise in which we alternatively shock specific and iceberg trade costs to each Chinese trading partner by enough to lower trade by 5 percent. These shocks have equivalent welfare effects but generate very different price responses. Because increases in specific trade costs induce firms to raise their quality, they lead to exaggerated price increases, whereas shocks to ad valorem trade costs induce firms to lower the quality of the goods they provide and so lead to small changes in prices. Hence, the model demonstrates the need to know the nature of trade shocks before making predictions over the associated price changes.[2]

Finally, we show that in the case of generalized CES preferences that it is possible to derive closed form, sufficient-statistic-type expressions for the gains from trade. We show that conditional on the trade elasticity, generalized CES preferences imply larger gains from trade than CES preferences. This is due to the excessive love of variety implied by generalized CES preferences. Intuitively, with generalized CES preferences, consumers obtain positive levels of

[2]On a related note, variation in prices across countries are occasionally used to measure trade costs. If bilateral trade costs vary in their mixture of specific and ad valorem costs, much of the observed differences in prices would be due to quality upgrading rather than absolute levels of trade costs.

utility simply for having the option to consume a good. Hence, goods available at the choke price lead to strictly positive levels of utility as is the case in models with fixed export costs and standard CES preferences.

Our paper contributes to two strands of the literature that seek to understand the causes and implications of international prices. First, our focus on endogenous quality puts our paper into a literature that includes the recent paper by Feenstra and Romalis (2014) who provide a monopolistic competition model that has been designed to estimate the quality of goods traded and sold domestically with the intention of purging price indices of quality variation across countries.[3] As the authors are working with country-level data, they do not develop their model to confront the firm-level joint distribution of prices and sales, which is the focus of our paper.[4]

Second, our paper also contributes to the literature featuring variable markups. These papers include Jung, Simonovska and Weinberger (2019) and Atkeson and Burstein (2008). As in Jung, Simonovska and Weinberger (2019), we consider non-homothetic preferences and a market structure that gives rise to variable markups across firms. Relative to their paper, we also consider vertically differentiated products, quality upgrading opportunities, and specific trade costs that give rise to the "Washington Apples" effect. Our framework, therefore, allows for much of the variation across countries and firms to be attributed not to variation in market power but to variation in quality of output. Allowing for quality upgrading helps to make the model with variable markups more consistent with the well-known pattern in the data that the most successful exporters tend to charge the highest prices (e.g. Manova and Zhang, 2012; Harrigan, Ma and Shlychkov, 2015). Moreover, our framework highlights the differential effect of specific and ad valorem trade costs on the international distribution of prices.

Our paper is also related to the recent work by Hottman, Redding and Weinstein (2016) who allow for both market power and quality heterogeneity to drive price dispersion across local prices in the United States. They find that a very substantial portion of heterogeneity in market shares can be attributed to quality heterogeneity but with firms' strategic pricing decisions also playing a non-trivial role. By considering a more parsimonious setting, we can conduct an analysis of the role of markup and quality dispersions to an international setting.

The remainder of this paper is organized into six sections. In Section 2, we develop a series of stylized facts concerning the international pricing behavior of Chinese firms that we will use to calibrate our model. In Section 3, we present a simple, quantitative general equilibrium model that is able to rationalize these stylized facts and which can be quantified with features of our data. In addition to characterizing the equilibria, we derive an expression for the welfare gains

[3] A paper that analytically allows for firm heterogeneity, product quality differentiation, and variable markups in an international context is Antoniades (2015). The paper shares the qualitative predictions of ours but makes no effort to confront the model with data. Moreover, its use of linear quadratic preferences would make it difficult to square with first-order features of the data.

[4] The literature on quality differences across countries is very rich. Earlier contributions include Schott (2004), Kugler and Verhoogen (2009, 2012), Khandelwal (2010), Baldwin and Harrigan (2011), Manova and Zhang (2012), Johnson (2012), Bas and Strauss-Kahn (2015), Harrigan, Ma and Shlychkov (2015), and Fan, Li and Yeaple (2015, 2018).

from shocks to the international trading environment and compare the welfare implications of our model to three other models, each of which contains only a subset of the parameters of the benchmark model. In Section 4, we describe how we solve, calibrate and simulate our benchmark model and the three alternative models. In Section 5, we assess the model's fit to the data, and contrast the model's fit relative to other models that lack one or more of the features of our model. In Section 6, we discuss the model's quantitative implications for the gains from trade. Again, we contrast the benchmark model's predicted gains from trade relative to alternative models that lack either the "Washington Apples" quality mechanism or the variable markups mechanism or both. In Section 7, we illustrate how specific and ad valorem trade shocks that have identical effects on welfare and on trade volumes have very different effects on prices. This is important as it shows how micro-econometric models that neglect specific trade costs may be misspecified. Finally, in Section 8, we provide concluding comments.

2 Stylized Facts

2.1 Data

To document the stylized facts regarding export prices across destinations and across firms within the same destination, we use two micro-level databases and one aggregate-level cross-country database. Specifically, these are (1) the transaction-level export data from China's General Administration of Customs; (2) the annual survey of industrial firms from the National Bureau of Statistics of China (NBSC); (3) the CEPII Gravity database that provides destination countries' characteristics such as population, GDP per capita, and distance to China. We use data for the year 2004 to be consistent with the calibration exercise later.[5]

The China's Customs database records each export and import transaction for the universe of Chinese firms at the HS8 product level, including values, quantities, products, source and destination countries, firm contacts (e.g., company name, telephone, zip code, and contact person), enterprise types (e.g., state owned, domestic private, foreign invested, or joint venture), and customs regimes (e.g., ordinary trade, or processing trade). We aggregate each transaction-level data to various levels, including firm-HS6-destination country, firm-HS6, or HS6-country for further analysis.[6] We compute unit values (i.e., export values divided by export quantities) as a proxy for export prices and focus on ordinary trade exporters.[7]

To characterize firms' attributes such as TFP, employment, capital intensity, and wage,

[5] To calibrate our model, we construct bilateral trade shares following the method in Ossa (2014) based on GTAP 9 Data Base for the year 2004 (see Section 4 for more details).

[6] Our main results use 2004 data. To check the sensitivity, we also experimented using the data in other years between 2000-2006 and obtained similar results. We use Chinese HS6-level data because the product codes are consistent over time while Chinese HS8 product classifications change over time.

[7] Processing traders have very little control over the prices that they receive for their goods and are often the affiliates of foreign firms who directly control the prices in transactions. This is the key reason that processing traders are excluded from this analysis.

we use the NBSC firm-level data from the annual surveys of Chinese industrial firms. This database contains detailed firm-level production, accounting and firm identification information for all state-owned enterprises (SOEs) and non-state-owned enterprises with annual sales of at least 5 million *Renminbi* (RMB, Chinese currency). We use merged data of both the Customs data and the NBSC firm survey data when firms' characteristics are needed.[8]

2.2 Empirical Regularities

In this subsection, we report three stylized facts concerning export prices across destinations and across firms within destination as well as the number of firms that export to each destination. Note that the existing literature has documented many of these facts separately, but it is useful to show that they hold in the Chinese data. Moreover, it is these facts that we seek to be able to explain within a single model and that we will use to calibrate this model.

Table 1: Export Prices across Destination

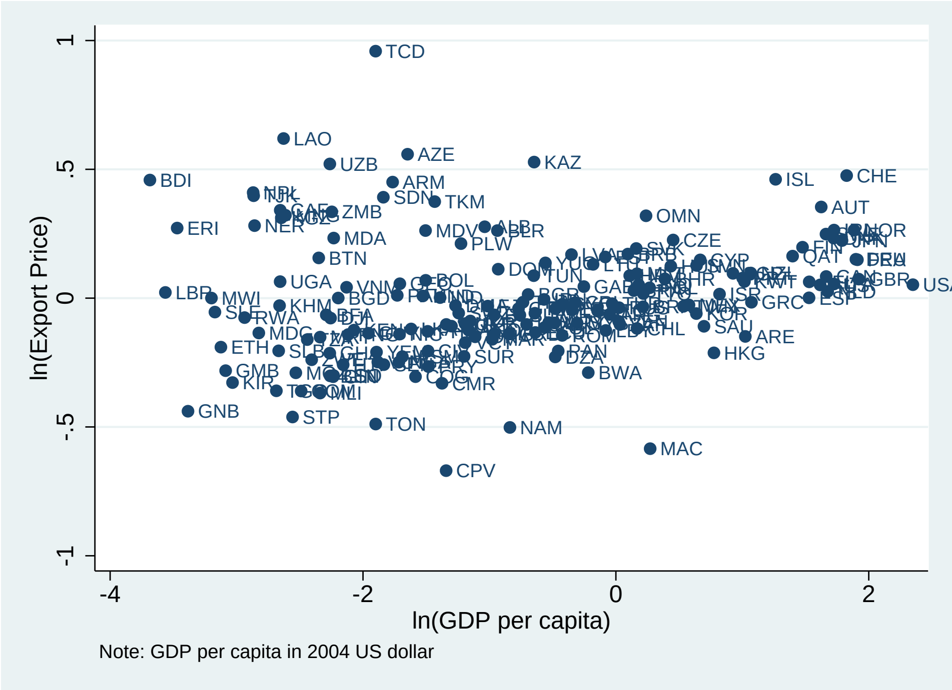
	Dependent Variable: ln(*price*)			
	$\ln(p_{fhc})$		$\ln(p_{hc})$	
	(1)	(2)	(3)	(4)
GDP per capita (current in US dollar)	0.024***	0.026***	0.042***	0.045***
	(0.005)	(0.005)	(0.010)	(0.009)
Country-level Other Control	no	yes	no	yes
Firm-Product Fixed Effect	yes	yes	no	no
Product Fixed Effect	no	no	yes	yes
Observations	1,441,468	1,441,468	173,055	173,055
R-squared	0.946	0.946	0.831	0.831

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors corrected for clustering at the destination country level in parentheses. Robust standard errors corrected for clustering at the destination country level in parentheses. The dependent variable in specifications (1)-(2) is the (log) price at the firm-HS6-country level, and in specifications (3)-(4) is the (log) price at the HS6-country level. Country-level other controls include population and distance. All regressions include a constant term.

Fact 1: Export prices across destinations.— Based on the whole customs data in 2004, Table 1 reports the regression results using (log) export prices as the dependent variable and destination country's GDP per capita as main explanatory variable, controlling for destination's population and distance to China. Columns 1-2 and 3-4 use the prices at the firm-HS6-country level and the HS6-country level, respectively. The coefficients on GDP per capita in all specifications are positive and statistically significant, indicating that export prices increase in destination's income (e.g., Manova and Zhang, 2012). To better control for country-level

[8]Due to some mis-reporting, we follow Cai and Liu (2009) and use General Accepted Accounting Principles to delete the unsatisfactory observations in the NBSC database. See Fan, Li and Yeaple (2015) for more detailed description of data and the merging process.

Figure 1: Export prices increase with destination income



Notes: Export prices for ordinary trade from China's Customs data in 2004. Prices (in logarithm) are drawn by regressing HS6-country level export prices on HS6 product fixed effects as well as controlling for destinations' population and distance and then plotting the mean residuals for each destination.

characteristics, we include the destination country's GDP per capita in columns 1 and 3, while in columns 2 and 4, we further control for population and distance. Comparing odd columns with even columns, we find that adding population and distance would not affect our results qualitatively.

In Figure 1, we plot the mean residuals of each destination from regressing log export prices on product fixed effects and log destination GDP per capita as well as destination's population and distance. The data reveal a positive relationship between export prices and destination income. We summarize the following fact:

Stylized fact 1. *Firms set higher export prices for the same product in richer destinations.*

Fact 2: Export prices across firm.— To present export prices across firm, we use the merged data of the customs and the NBSC in 2004 in Table 2 and report the results obtained by regressing export prices on firm productivity, and other firm-level controls, such as employment, capital intensity, and the wage it pays. The measure of firm productivity is revenue based TFP, estimated by the augmented Olley-Pakes' (Olley and Pakes, 1996) approach by allowing a firm's trade status and the WTO shock in the TFP realization, as in Amiti and Konings (2007).[9] In columns 1-2, we use firm-HS6-country level price and include product-country fixed effect; in columns 3-4, we use firm-HS6 price and include HS6 product fixed effect. We do not control

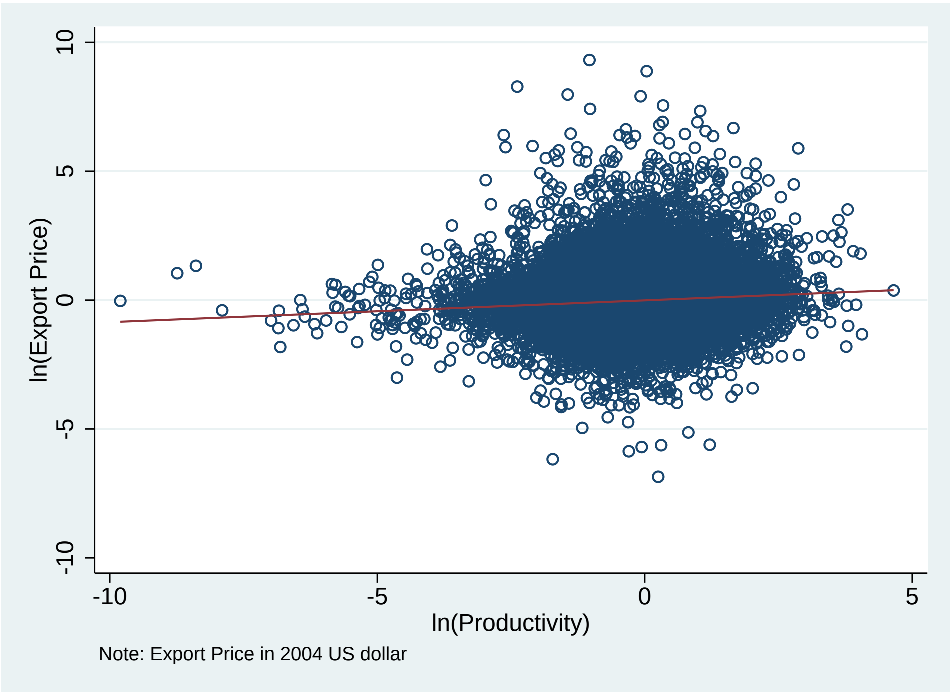
[9] Revenue TFP is computed by the same approach as in Fan, Li and Yeaple (2015, 2018) which contain detailed description of TFP estimation methods.

Table 2: Export Prices across Firm

	Dependent Variable: ln(*price*)			
	$\ln(p_{fhc})$		$\ln(p_{fh})$	
	(1)	(2)	(3)	(4)
ln(TFP)	0.095***	0.050***	0.094***	0.050***
	(0.009)	(0.009)	(0.009)	(0.011)
Firm-level Other Control	no	yes	no	yes
Product-country Fixed Effect	yes	yes	no	no
Product Fixed Effect	no	no	yes	yes
Observations	504,813	504,627	185,689	185,607
R-squared	0.775	0.779	0.638	0.644

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors corrected for clustering at the firm level in parentheses. The dependent variable in specifications (1)-(2) is the (log) price at the firm-HS6-country level, and in specifications (3)-(4) is the (log) price at the firm-HS6 level. Firm-level other controls include employment, capital-labor ratio, and wage. All regressions include a constant term.

Figure 2: Export prices increase with firm productivity



Notes: Export prices for ordinary trade from China's Customs data in 2004. Prices (in logarithm) are drawn by regressing firm-HS6 level export prices on HS6 product fixed effects and then plotting the mean residuals for each firm.

for employment, capital intensity and wage in columns 1 and 3, while in columns 2 and 4 we add those firm-level controls to show the robustness of our regression results. The coefficient on firm's TFP are all significantly positive, which is consistent with the quality-and-trade literature that high-productivity firms charge higher prices (e.g., Fan, Li and Yeaple, 2015). Figure 2 also plots export prices against firm's TFP by regressing firm-HS6 level export prices on HS6 product fixed effects and then plotting the mean residuals for each firm. Table 2 and Figure 2 yield the following fact:

Stylized fact 2. *Higher-productivity firms set higher export prices for the same product within the same market.*

Table 3: Firm Mass across Destination

	Dependent Variable: $\ln(FirmNumber)$			
	$\ln(N_{hc})$		$\ln(N_c)$	
	(1)	(2)	(3)	(4)
GDP per capita (current in US dollar)	0.236***	0.296***	0.687***	0.767***
	(0.042)	(0.020)	(0.070)	(0.042)
Country-level other Control	no	yes	no	yes
Product Fixed Effect	yes	yes	no	no
Observations	173,422	173,422	173	173
R-squared	0.322	0.528	0.292	0.808

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Robust standard errors corrected for clustering at the destination country level in parentheses. The dependent variable in specifications (1)-(2) is the (log) firm number at the HS6-country level, and in specifications (3)-(4) is the (log) firm number at the destination country level. Country-level other controls include population and distance. All regressions include a constant term.

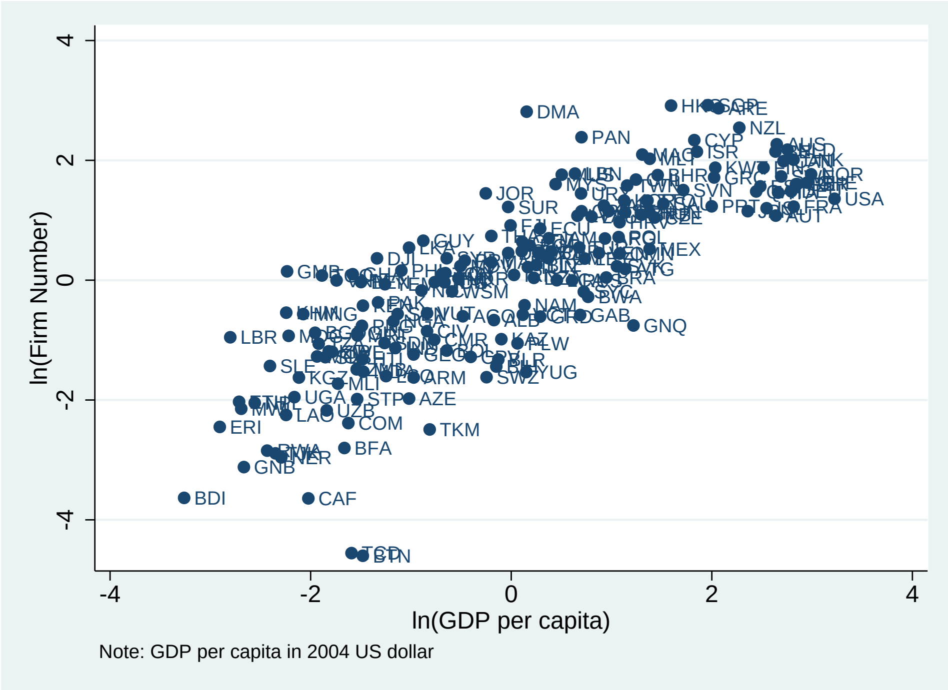
Fact 3: Extensive Margin of Firm Entry across Destinations.— We now turn to the number of exporting firms in different destinations. Table 3 reports the results of regressing the logarithm of the number of firms that export to each HS6-country (in columns 1-2) and each country (in columns 3-4) on destination country's GDP per capita, including product fixed effects in columns 1-2 and further controlling for destination's population and distance to China in columns 2 and 4. The significantly positive coefficients on the log of GDP per capita suggest that more firms export to richer destinations. Figure 3 further supports the following finding by plotting (log) firm number at each destination against destination's income:

Stylized fact 3. *More firms export to high-income destinations.*

3 Model

In this section, we introduce and solve our model. We first introduce the demand side of the model and solve for the optimal markup as a function of a firm's quality of output and marginal cost of production. We then endogenize quality choice and characterize a firm's decision to enter into a given market as a function of its heterogeneous cost draws. Third, we solve for

Figure 3: Firm Mass increases with destination income



Notes: Destination-level firm number (in logarithm) are drawn against destination's (log) GDP per capita by controlling for destinations' population and distance.

the implied aggregate variables and close the model with labor market clearing/trade balance. Finally, we derive a formula for the aggregate gains from trade and show how the model can be used to conduct comparative static exercises à la Dekle, Eaton and Kortum (2008). In addition, we compare the welfare implications of our model to three other models that each features a gravity equation and an extensive margin.

3.1 Tastes and Endowments

Consider a world populated by J countries, indexed by i and j with country j endowed with L_j units of labor. The preferences of the representative consumer in each country are identical but are non-homothetic leading to different marginal valuations of quality and access to variety. Specifically, we extend the preference system considered by Jung, Simonovska and Weinberger (2019) augmented such that varieties vary in their perceived quality. We denote the source country by i and the destination country by j. Consumers in country j have access to a set of goods Ω_j, which is potentially different across countries. Specifically, the representative consumer has preferences of:

$$U_j = \left[\sum_i \int_{\omega \in \Omega_{ij}} \left(q_{ij}(\omega) x_{ij}^c(\omega) + \overline{x} \right)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}} \tag{1}$$

where $\sigma > 1$ is the elasticity of substitution, $x_{ij}^c(\omega)$ is the quantity of variety ω from country i consumed by the representative consumer in country j, $q_{ij}(\omega)$ is it's quality, and $\overline{x} > 0$ is a constant.

Utility maximization imples that the demand curve for variety ω is given by:

$$x_{ij}(\omega) = x_{ij}^c(\omega)L_j = \frac{L_j}{q_{ij}(\omega)}\left[\frac{y_j + \bar{x}P_j}{P_{j\sigma}^{1-\sigma}}\left(\frac{p_{ij}(\omega)}{q_{ij}(\omega)}\right)^{-\sigma} - \bar{x}\right] \tag{2}$$

where $p_{ij}(\omega)$ is the price of output from country i to country j, $P_j = \sum_i \int_{\omega\in\Omega_{ij}} p_{ij}(\omega)/q_{ij}(\omega)\, d\omega$ and $P_{j\sigma} = \left\{\sum_i \int_{\omega\in\Omega_{ij}} \left(p_{ij}(\omega)/q_{ij}(\omega)\right)^{1-\sigma} d\omega\right\}^{\frac{1}{1-\sigma}}$ denote aggregate price statistics, y_j is the representative consumer's income, reflecting GDP per capita in the destination country, and N_j is the mass of varieties consumed in country j (see Appendix A for detailed derivation).

To simplify our discussion and to keep our notation compact, we define the quality-adjusted price charged by firm ω from country i selling in market j to be $\tilde{p}_{ij}(\omega) = p_{ij}(\omega)/q_{ij}(\omega)$, and we define the country j "choke" price level to be $\tilde{p}_j^* = \left(\frac{y_j+\bar{x}P_j}{\bar{x}P_{j\sigma}^{1-\sigma}}\right)^{\frac{1}{\sigma}}$. Everything else equal, high nominal per-capita incomes and higher prices imply higher choke prices facing individual firms.

We thus can write quantity, sales, and profit for a given variety exported from i to j as follows,

$$x_{ij}(\omega) = \frac{\bar{x}L_j}{q_{ij}(\omega)}\left[\left(\frac{\tilde{p}_{ij}(\omega)}{\tilde{p}_j^*}\right)^{-\sigma} - 1\right] \tag{3}$$

$$r_{ij}(\omega) = \bar{x}L_j\tilde{p}_{ij}(\omega)\left[\left(\frac{\tilde{p}_{ij}(\omega)}{\tilde{p}_j^*}\right)^{-\sigma} - 1\right] \tag{4}$$

$$\pi_{ij}(\omega) = \bar{x}L_j\left[\tilde{p}_{ij}(\omega) - \tilde{c}_{ij}(\omega)\right]\left[\left(\frac{\tilde{p}_{ij}(\omega)}{\tilde{p}_j^*}\right)^{-\sigma} - 1\right] \tag{5}$$

where $\tilde{c}_{ij}(\omega) = c_{ij}(\omega)/q_{ij}(\omega)$ is the quality-adjusted marginal cost and $c_{ij}(\omega)$ is the marginal cost of production. Given the quality-adjusted marginal cost, firms maximize their profits.

Taking as given the pricing behavior of all other firms, the monopolistically competitive producer of variety ω chooses its quality-adjusted price of the good. The first-order condition for profit maximization implicitly yields the optimal price $\tilde{p}_{ij}(\omega)$ which satisfies:

$$\sigma\frac{\tilde{c}_{ij}(\omega)}{\tilde{p}_j^*} = \left(\frac{\tilde{p}_{ij}(\omega)}{\tilde{p}_j^*}\right)^{\sigma+1} + (\sigma-1)\frac{\tilde{p}_{ij}(\omega)}{\tilde{p}_j^*}. \tag{6}$$

Note that the optimal prices and optimal profits depend only on the quality-adjusted marginal cost of production. In the next subsection, we endogenize a firm's choice of its quality-adjusted marginal cost of production.

3.2 Quality and Production

Firms are heterogeneous in productivity φ. Following Feenstra and Romalis (2014), for a firm from country i with productivity φ requires l of labor produce one unit of output with quality q according to the production function:

$$l = \frac{q^\eta}{\varphi},$$

where $\eta > 1$ is a measure of the scope for quality differentiation. In addition, a firm from country i that wishes to sell its product in country j must incur two types of variable shipping costs. The first, $\tau_{ij} \geq 1$, is the standard iceberg-type shipping cost which requires τ_{ij} units to be shipped for one unit to arrive. The second, T_{ij}, is a per-unit shipping cost (a specific trade cost). For simplicity, we assume that specific trade costs are in terms of country i labor.

For a firm from country i of productivity φ that has received country j's idiosyncratic cost shock ε, the marginal cost of supply one unit of quality q_{ij} to country j is

$$c_{ij}(\varphi, \varepsilon) = \left(T_{ij} w_i + \frac{w_i \tau_{ij}}{\varphi} q_{ij}^\eta \right) \varepsilon$$

where τ_{ij} is ad valorem trade cost and T_{ij} is a specific transportation cost from country i to country j.

Hence, the quality adjusted marginal cost of production is given by

$$\frac{c_{ij}(\varphi, \varepsilon)}{q_{ij}} = \frac{\left(T_{ij} w_i + \frac{w_i \tau_{ij}}{\varphi} q_{ij}^\eta \right) \varepsilon}{q_{ij}}. \tag{7}$$

As will be obvious in a moment when solving for optimal quality choice by firm this formulation has several desirable features. First, it will exhibit the "Washington Apples" effect: higher specific trade costs will induce firms to upgrade their quality. Second, it will be consistent with the well documented fact that more productive firms charge higher prices (e.g. Kugler and Verhoogen (2009), Manova and Zhang (2012)). Third, it will prove to be highly tractable, allowing us to avoid the tractability issues that have prevented quality and variable markups analysis in the past.

From the first-order condition associated with equation (7), the optimal level of quality for a firm with productivity φ is

$$q_{ij}(\varphi, \varepsilon) = \left(\frac{T_{ij} \varphi}{(\eta - 1) \tau_{ij}} \right)^{\frac{1}{\eta}} \tag{8}$$

and hence the quality adjusted marginal cost of supplying market j from i could be rewritten:

$$\tilde{c}_{ij}(\varphi, \varepsilon) = \frac{c_{ij}(\varphi, \varepsilon)}{q_{ij}(\varphi, \varepsilon)} = \left(\frac{\eta}{\eta - 1} T_{ij} w_i \right)^{\frac{\eta - 1}{\eta}} \left(\frac{\varphi}{\eta w_i \tau_{ij}} \right)^{-\frac{1}{\eta}} \varepsilon. \tag{9}$$

It is immediate from this expression that more productive firms produce higher quality goods but actually face lower quality-adjusted costs. Also the quality-adjusted cost is an increasing geometric average of both types of shipping costs with the weights driven by η. As η goes to one, specific trade costs matter not at all and our model becomes the model given by Jung, Simonovska and Weinberger (2019). As η goes to infinity, however, firm productivity becomes complete irrelevant and the weight of the specific trade cost goes to one. As a result, the more costly it is to upgrade quality (higher η) the less quality-adjusted marginal cost is decreasing in firm productivity. Hence, specific trade costs hit the most productive firms more heavily than the less productive.

Equation (3) implies that consumer does not have positive demand for goods with sufficiently high quality-adjusted prices. The quality adjusted price $\tilde{p}_{ij}$ can not exceeds the choke price, $\tilde{p}_j^*$. At the cutoff, equations (3) and (6) imply:

$$\tilde{p}_{ij}^* (\varphi, \varepsilon) = \tilde{c}_{ij}^* (\varphi, \varepsilon) = \tilde{p}_j^* \tag{10}$$

where $\tilde{p}_{ij}^* (\varphi, \varepsilon)$ and $\tilde{c}_{ij}^* (\varphi, \varepsilon)$ are the quality adjusted price and the quality adjusted marginal cost at the entry threshold, $\varphi_{ij}^* (\varepsilon)$. Hence, the previous equation, together with equation (9), imply that the productivity cutoff $\varphi_{ij}^* (\varepsilon)$ to sell goods from country i to country j satisfies:

$$\varphi_{ij}^* (\varepsilon) = \varphi_{ij}^* \varepsilon^{\eta-1} = \frac{\eta^\eta}{(\eta-1)^{\eta-1}} T_{ij}^{\eta-1} \tau_{ij} w_i^\eta \left(\tilde{p}_j^*\right)^{-\eta} \varepsilon^\eta, \tag{11}$$

where

$$\varphi_{ij}^* = \frac{\eta^\eta}{(\eta-1)^{\eta-1}} T_{ij}^{\eta-1} \tau_{ij} w_i^\eta \left(\tilde{p}_j^*\right)^{-\eta} \tag{12}$$

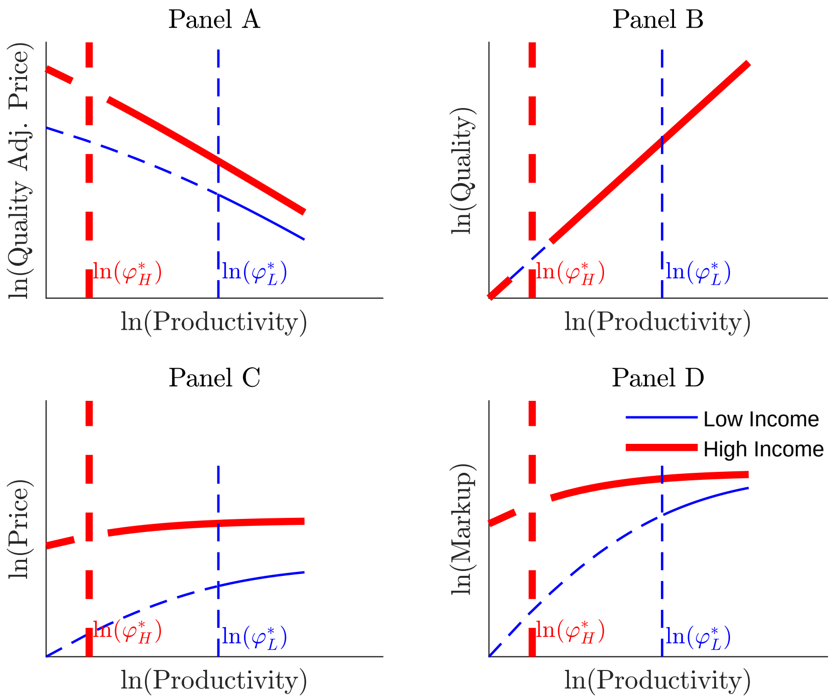
is the deterministic part of the productivity cutoff that is common across firms.

Figure 4 illustrates that the relationship of the quality-adjusted export price, export price, export quality and export markup with firm's productivity within and across countries.[10] The blue solid line represents this relationship in the low-income destination country; the red, thicker line denotes it in the high-income destination country. In Panel C of Figure 4, we depict the positive relationship between price and productivity. Since markups over marginal cost vary systematically with market characteristics, both the quality-adjusted export price, and absolute export price are higher in higher-income country. This is due to the higher markups that can be charged in richer markets.[11] If firms set a constant markups over marginal costs, then there would be no correlation between price and productivity since per-unit costs do not depend on firm productivity. Hence, the variable markups generate the positive relationship between

[10] Note that Figure 4 is an illustration based on simulation because we do not have explicit expression for price and markup as function of productivity under CES, but we can derive explicit expressions under log utility function (see Appendix B).

[11] It is straightforward to show that when there is a portion of the cost of the specific trade cost incurred in the destination country, then richer countries would also be purchasing higher quality goods than poor countries.

Figure 4: Illustration of Model Mechanism



price and productivity. This positive relationship depends on the values of η.

In Panel D of Figure 4, we depict the positive relationship between markup and productivity within and across countries. Suppose the log case (i.e., $\sigma = 1$), the markup could be explicitly expressed as $\frac{\varphi}{\varphi^*_{ij}(\varepsilon)}$. As depicted in Panel D, the markup for a firm with the same productivity in high-income destination market should be higher since export productivity cutoff φ^*_{ij} is lower in high-income market.[12]

It is worthy of comparing our model with the three classes of models in the literature to show the importance of the interaction between endogenous quality and variable markups in reconciling all three aforementioned empirical facts in Section 2 simultaneously. The first branch of models features constant markups, firm heterogeneity and product quality differentiation (e.g., Johnson, 2012). These models can predict positive correlation between price and sales within a market across firms, but would not be able to explain the facts across countries that firms set higher export prices in higher-income destinations and that more firms export to higher-income destinations. The second class of model features variable markups and firm heterogeneity but without endogenous quality (e.g., Jung, Simonovska and Weinberger, 2019). Those models perform well in explaining cross-country pricing-to-market patterns, but would not explain the stylized fact that higher-productivity firms set higher export quality and higher

[12]Conditional on the same market, the distribution of markups should be the same because the term $\left(\frac{\varphi}{\varphi^*_{ij}(\varepsilon)}\right)^{\frac{1}{2\eta}}$ would follow a Pareto distribution with shape parameter equal to $2\eta\theta$. Hence, we compare the different markup across countries for the same firm instead of depicting the market distribution within each market.

export prices for the same product within the same market.

Finally, the model of Antoniades (2015) also features variable markups, firm heterogeneity and product quality differentiation based on Melitz-Ottaviano's quadratic preferences (Melitz and Ottaviano, 2008). We differ from Antoniades (2015) in two ways. First, his model feathers quality sorting with linear demand and so yields ambiguous predictions over individual firms' prices across destinations and over the average price charged across destinations that vary in their income (Manova and Zhang, 2012). In contrast, our model unambiguously predicts that firms set higher export prices in higher-income destinations as documented by stylized fact 1.[13] Second, our model is a highly tractable, quantitative general equilibrium model that incorporates endogenous quality, variable markups, and two types of trade costs – variable and specific trade costs.

3.3 Aggregation and Equilibrium

In order to analytically solve the model and to derive stark predictions at the firm and aggregate levels, we follow much of the literature and assume that firm productivities are drawn from a Pareto distribution with cdf $G_i(\varphi) = 1 - b_i\varphi^{-\theta}$ and pdf $g_i(\varphi) = \theta b_i \varphi^{-\theta-1}$, where shape parameter $\theta > 1$ and $b_i > 0$ summarizes the level of technology in country i. We assume $\varphi_{ij}^* > b_i$ for all ij so that the cutoff is active for all country pairs. The idiosyncratic cost shock ε is drawn from a log normal distribution, where $\log \varepsilon$ follows the normal distribution with zero mean and variance σ_ε^2.

We first derive the measure of the subset of entrants from i who surpass the productivity threshold $\varphi_{ij}^*(\varepsilon)$ and so serve destination j. The exporting firm mass from i to j, N_{ij}, is defined as

$$N_{ij} = J_i \int_0^\infty \Pr\left[\varphi > \varphi_{ij}^*(\varepsilon)\right] f(\varepsilon)\, d\varepsilon,$$

where J_i is the potential firm mass in country i and $f(\varepsilon)$ is the pdf distribution of ε. As shown in Appendix C, the following simple expression of this mass of entrants can be obtained

$$N_{ij} = \kappa J_i b_i \left(\varphi_{ij}^*\right)^{-\theta}, \tag{13}$$

where κ is a constant, and φ_{ij}^* is the deterministic component of the productivity cutoff given by equation (12).[14]

Note how the measure of entrants from i into market j depends on the "choke price," $\tilde{p}_j^*$ through equation (12). An increase in the choke price induces a lower deterministic productivity cutoff and this expands the measure of firms operating there. The elasticity of the measure of active firms with respect to the choke price is $\theta\eta$, and this illustrates how the "Washington Apples" effect interacts with the underlying productivity dispersion across firms. Ceteris

[13] See Manova and Zhang (2012) for a comprehensive summary of the predicted behavior of export prices under efficiency or quality sorting and the pattern observed in the trade data.

[14] $\kappa = \int_0^\infty \varepsilon^{-\theta(\eta-1)} f(\varepsilon)\, d\varepsilon = \exp\left(\frac{1}{2}\left[(1-\eta)\,\theta\sigma_\varepsilon\right]^2\right)$.

paribus, an increase in the cost of upgrading quality acts like a decrease in the dispersion in firm productivity.

We will see that all of the other aggregates in the economy are tightly linked to (13). In deriving these aggregates it is useful to define the conditional density function for the productivity of firms from i operating in j is

$$\mu_{ij}(\varphi,\varepsilon)=\begin{cases}\theta\left[\varphi_{ij}^{*}(\varepsilon)\right]^{\theta}\varphi^{-\theta-1} & \text{if } \varphi>\varphi_{ij}^{*}(\varepsilon)\\ 0 & \text{otherwise}\end{cases} \tag{14}$$

With these definitions in mind, the aggregate price statistics, P_j and $P_{j\sigma}$, can be rewritten as

$$P_j=\sum_i N_{ij}\int_0^{\infty}\int_{\varphi_{ij}^{*}(\varepsilon)}^{\infty}\tilde{p}_{ij}(\varphi,\varepsilon)\,\mu_{ij}(\varphi,\varepsilon)\,f(\varepsilon)\,d\varphi d\varepsilon,\text{ and}$$
$$P_{j\sigma}=\left\{\sum_i N_{ij}\int_0^{\infty}\int_{\varphi_{ij}^{*}(\varepsilon)}^{\infty}\tilde{p}_{ij}(\varphi,\varepsilon)^{1-\sigma}\,\mu_{ij}(\varphi,\varepsilon)\,f(\varepsilon)\,d\varphi d\varepsilon\right\}^{\frac{1}{1-\sigma}}.$$

As shown in Appendix C that contains detailed derivation for aggregate variables P_j, $P_{j\sigma}$, X_{ij} and π_i, all variation in prices due to the idiosyncratic trade cost shocks integrate out so that we may write these price statistics as

$$P_j=\beta\tilde{p}_j^{*}N_j, \tag{15}$$
$$P_{j\sigma}=\beta_{\sigma}^{\frac{1}{1-\sigma}}\tilde{p}_j^{*}N_j^{\frac{1}{1-\sigma}}, \tag{16}$$

where $N_j=\sum_i N_{ij}$ is the total mass of firms from all countries that have positive sales in country j, and β and β_σ are constants that obtain after integrating out ε from each expression (see Appendix C). Similar constants will also appear in each of the aggregate relationships displayed below.

We assume that there is free entry. Hence, in equilibrium, the expected profit of an entrant is zero and aggregate profits obtained by individual consumer are also zero. As a result, the representative consumer's income y_j reduces to the wage rate w_j since each consumer has a unit of labor endowment. Then we have $\tilde{p}_j^{*}=\left(\frac{w_j+\bar{x}P_j}{\bar{x}P_{j\sigma}^{1-\sigma}}\right)^{\frac{1}{\sigma}}$. The expression of $\tilde{p}_j^{*}$, together with equation (15) and (16), imply that the quality-adjusted choke price is

$$\tilde{p}_j^{*}=\frac{1}{\bar{x}\left[\beta_\sigma-\beta\right]}\frac{w_j}{N_j}. \tag{17}$$

Importantly, an increase in the per capita income in a country, w_j, is associated with a greater choke price, while an increase in competition, N_j, is associated with a lower quality-adjusted choke price.

Having derived expressions for the "choke price" and the price indices, it is straightforward to show that the total expenditure of country j on the goods from country i, given by

$$X_{ij} = N_{ij} \int_0^\infty \int_{\varphi_{ij}^*(\varepsilon)}^\infty r_{ij}(\varphi, \varepsilon)\, \mu_{ij}(\varphi, \varepsilon)\, f(\varepsilon)\, d\varphi d\varepsilon,$$

can be written as

$$X_{ij} = X_j \frac{N_{ij}}{N_j}, \tag{18}$$

where $X_j \equiv w_j L_j$ is total absorption. Equation (18) shows that our model shares with many commonly used models in the literature the feature that variation in trade volumes across country occur entirely along the extensive margin.

The expected profits can be calculated using

$$\pi_i = \sum_j \int_0^\infty \int_{\varphi_{ij}^*(\varepsilon)}^\infty \pi_{ij}(\varphi, \varepsilon)\, g_{ij}(\varphi)\, f(\varepsilon)\, d\varphi d\varepsilon.$$

As shown in the appendix, these expected profits can be shown to be

$$\pi_i = \frac{1}{J_i} \frac{\beta_\pi}{\beta_\sigma - \beta} \sum_j \frac{N_{ij}}{N_j} X_j \tag{19}$$

where β_π is also a constant.[15]

The household budget equation implies that total income equals to total expenditure

$$w_i L_i = \sum_j X_{ij} \tag{20}$$

Free entry, $\pi_i = w_i f$, together with (18), (19), and (20) pin down the measure of entrants:

$$J_i = \frac{\beta_\pi}{\beta_\sigma - \beta} \frac{L_i}{f}. \tag{21}$$

So, as in standard models of monopolistic competition in the Krugman tradition, the measure of entrants is proportional to country size and invariant to the trading environment. Finally, we assume trade is balanced:

$$\sum_j X_{ij} = \sum_j X_{ji}. \tag{22}$$

This concludes our characterization of the equilibrium. Note that equations (12), (13), and (18) imply the following theoretical gravity relationship:

$$\frac{\lambda_{ij}}{\lambda_{jj}} = \frac{J_i b_i \left(T_{ij}^{\eta-1} \tau_{ij} w_i^\eta\right)^{-\theta}}{J_j b_j \left(T_{jj}^{\eta-1} \tau_{jj} w_j^\eta\right)^{-\theta}}. \tag{23}$$

[15] Notice here we have that firms' total variable profit is proportional to total revenue as Arkolakis, Costinot and Rodríguez-Clare (2012).

Equation (23) will lead to an empirical gravity equation for estimation in the later calibration. In the next subsection, we show how the gains from trade and how comparative statics on shifting trade costs can be inferred from existing data and estimates of the key model parameters, η, θ, and σ.

3.4 Welfare

In this section, we show how the measurement of the gains from trade, and the welfare implications of any shock to trade costs, are related to the key parameters of the model. We first derive an Arkolakis, Costinot and Rodríguez-Clare (2012) inspired formula relating changes in the level of domestic absorption to changes in real income and then derive the Dekle, Eaton and Kortum (2008) system of equations. The latter system of equations are novel in that they allow for both iceberg-type and specific trade cost shocks to be analyzed. We present these results as propositions whose proofs can be found in the online appendix (see Appendix D). Next we present a multi-sector extension and its welfare implication. In the end of the section we compare the welfare implications of our benchmark model to three other models.

3.4.1 Gains from Trade

Combining the utility expression (1), equation (3), and equation (C.5) (in the online appendix), the measure of indirect utility can be expressed as a function of the nominal wage relative to the equilibrium choke price:

$$U_j = \beta_u \left(\frac{w_j}{\tilde{p}_j^*} \right)^{\frac{\sigma}{\sigma-1}}$$

where $\beta_u = \bar{x}^{\frac{1}{1-\sigma}} \left(\frac{\beta_\sigma}{\beta_\sigma - \beta} \right)^{\frac{\sigma}{\sigma-1}}$ is a constant. We define the share of expenditure on goods from i in j, λ_{ij}, as:

$$\lambda_{ij} = \frac{X_{ij}}{\sum_{i'} X_{i'j}} \tag{24}$$

We denote the post adjustment value of any variable x as x' and the change in its value as $\widehat{x} = \frac{x'}{x}$, that is, a hat denotes the ratio between the counterfactual and factual value. Then the change in welfare associated with any foreign shock in country j satisfies the following Proposition:

Proposition 1. *The change in welfare associated with any foreign shock in country j can be computed as:*

$$\widehat{U}_j = \left(\widehat{\lambda}_{jj} \right)^{-\frac{\sigma}{\sigma-1}\frac{1}{1+\eta\theta}} \tag{25}$$

Equation (25) shows that the key parameters for assessing welfare implications of shocks are the taste parameter σ which plays a key role in the markup given a firm's choice of quality and its productivity, θ which governs the degree of dispersion in productivity, and η which governs the cost of quality upgrading in the model.

As in Bertoletti, Etro and Simonovska (2018), we compute the global welfare gains associated with a large change in trade costs. Equation (25) in Proposition 1 is a global measure of welfare gains from trade liberalization, because it can be derived from the measure of varieties sold in a country without taking differentiation (see Appendix E.1 for detailed derivation).[16]

Were we to strip the model of its "Washington Apples" mechanism, the model would be essentially identical to Jung, Simonovska and Weinberger (2019).[17] In that case, the coefficient on the change in the domestic consumption share $\widehat{\lambda}_{jj}$ becomes $-\frac{\sigma}{\sigma-1}\frac{1}{1+\theta}$ (see Appendix G for detailed derivation). As we discuss later in the paper, a comparison of the gains from trade implied by the models with and without the "Washington Apples" mechanism depends on the details of the calibration.

In order to evaluate the changes in welfare associated with any foreign shock, we need to measure $\widehat{\lambda}_{jj}$ and calibrate the parameters (σ, η, θ). Given the value of parameters (μ, η, θ) and initial value of X_{ij} before shocks, we have the following proposition:

Proposition 2. *The percentage change in welfare associated with any change in trade costs in country j can be computed using equation (25) combined with*

$$\widehat{\lambda}_{jj} = \frac{(\widehat{w}_j)^{-\eta\theta}}{\sum_i \lambda_{ij}\left[\widehat{T}_{ij}^{\eta-1}\widehat{\tau}_{ij}\right]^{-\theta}(\widehat{w}_i)^{-\eta\theta}} \tag{26}$$

where $\widehat{w}_j$ are implicitly given by the solution:

$$\widehat{w}_i = \sum_j \frac{\lambda_{ij} w_j L_j \left(\widehat{T}_{ij}^{\eta-1}\widehat{\tau}_{ij}\right)^{-\theta}(\widehat{w}_i)^{-\eta\theta}}{w_i L_i \sum_{i'} \lambda_{i'j}\left(\widehat{T}_{i'j}^{\eta-1}\widehat{\tau}_{i'j}\right)^{-\theta}(\widehat{w}_{i'})^{-\eta\theta}}\widehat{w}_j \tag{27}$$

Equations (26) and (27) are interesting in that they show that the elasticities associated with changes in trade costs differ depending on whether they are associated with ad valorem trade costs $\widehat{\tau}_{ij}$, or specific trade costs $\widehat{T}_{ij}$. Intuitively, shocks to both types of trade costs affect the extensive margin of entry of firms in markets and so involve the Pareto parameter θ. Shocks to specific trade costs, however, have an additional effect that works through quality upgrading and so the effect of these types of shocks depend on the elasticity of the costs associated with quality upgrading, η. To see how the quality scope parameter, η, affects welfare, we can compute welfare gains by alternating the values of η and keeping other parameters constant. Note that a lower value of η means a large scope for quality differentiation, while a higher value of η refers to a small scope for quality differentiation. Obviously, when the scope for quality upgrading is large (a lower η), it is very easy to avoid the impact of specific trade costs and

[16] A global measure comes from integrating local calculations when it is clear that all the relevant components of the expression are continuous and monotonic. Thus, the welfare formula in Proposition 1 is global. We also compute the Equivalent Variation of income associated with a change in trade costs as the global measure of the gains from trade liberalization, and find that it is proportional to Proposition 1 formula (see Appendix E.2 for details).

[17] This involves fixing the quality level to unity and setting all specific trade costs to zero.

this results in larger gains from trade. As the scope for quality upgrading becomes smaller (a higher η), the gains from trade become much smaller. This effect holds regardless of variable markups.[18]

3.4.2 Gains from Trade: Multiple Sectors

The tractability of our model can be also extended to a multi-sector setup, which corresponds to the sectoral heterogeneity of quality scope that has been featured by the recent literature.[19] Given a two-layer utility function $U_j = \prod_s C_{js}^{\alpha_s}$ with subscript s indexing sector and α_s denoting the Cobb-Douglas sector share, the demand function is similar with that of the one-sector benchmark model and is given by

$$x_{ijs}^{c}(\omega) = \frac{\overline{x}_s L_j}{q_{ijs}(\omega)} \left\{ \left[\frac{\tilde{p}_{ijs}(\omega)}{\tilde{p}_{js}^{*}} \right]^{-\sigma_s} - 1 \right\}$$

where $\tilde{p}_{js}^{*} = \left(\frac{\alpha_s \left(\sum_s \overline{x}_s P_{js} + y_j \right)}{\overline{x}_s P_{j\sigma s}^{1-\sigma_s}} \right)^{\frac{1}{\sigma_s}}$ is the corresponding quality-adjusted price cut-off for the multi-sector model.[20] We have the following proposition:

Proposition 3. *The percentage change in welfare associated with any changes in trade costs in country j can be computed as:*

$$\widehat{U}_j = \prod_s \left(\widehat{\lambda}_{jjs} \right)^{-\frac{\alpha_s \sigma_s}{\sigma_s - 1} \frac{1}{1+\eta_s \theta_s}}, \tag{28}$$

where λ_{ijs} denotes the share of expenditure on goods in sector s from i in j, η_s is the quality scope parameter in sector s, and θ_s is the sector specific Pareto distribution shape parameter.[21] Since the multi-sector model is in spirit similar to the one-sector model, in all the following quantitative exercises we shall refer to one-sector model as benchmark.

3.4.3 Gains from Trade: Alternative Models

In this subsection, we compare the welfare implications of our model to three other models that each features a gravity equation and an extensive margin. Our benchmark model (denoted by "Bench") accommodates both endogenous quality (the "Washington Apples" mechanism) and variable markups. The three alternative models are as follows: (1) a model containing only variable markups but no endogenous quality so the "Washington Apples" mechanism is

[18] Please also see the related discussion in Section 6 for the quantitative implications for the welfare gains. We find that the models with quality variation always generate lower gains from trade than those without endogenous quality (regardless of variable markups).

[19] The literature points out that firms' export pricing decision crucially depends on the quality scope that varies across sectors, e.g., Manova and Zhang (2012), Johnson (2012), Fan, Li and Yeaple (2015, 2017).

[20] Here we leave detailed derivation to Appendix F.1.

[21] The detailed proof of Proposition 3 is in Appendix F.2.

removed (labeled "no q");[22] (2) a model with the endogenous quality mechanism but with constant markups (labeled "con mkp");[23] (3) the model lacking the "Washington Apples" mechanism and with constant markups (labeled "no q, con mkp").[24] The derivations of the welfare gains from trade for the alternative models can be found in the various appendixes.

Following Proposition 1, the gains from trade in our benchmark model are given by

$$GT_j^{Bench} = 1 - (\lambda_{jj})^{\frac{\sigma}{\sigma-1}\frac{1}{1+\eta\theta}} . \tag{29}$$

As shown in Appendix G, the gains from trade under variable markups but no "Washington Apples" mechanism are given by

$$GT_j^{no\ q} = 1 - (\lambda_{jj})^{\frac{\sigma}{\sigma-1}\frac{1}{1+\theta}} . \tag{30}$$

As shown in Appendix H, the gains from trade under the model without variable markup but with quality mechanism are given by

$$GT_j^{con\ mkp} = 1 - (\lambda_{jj})^{\frac{1}{\eta\theta}} . \tag{31}$$

Finally, the model with neither mechanism implies gains from trade of

$$GT_j^{no\ q,\ con\ mkp} = 1 - (\lambda_{jj})^{\frac{1}{\theta}} . \tag{32}$$

The four models for which we have derived global gains from trade formulas each feature an extensive margin. For the two models featuring variable markups this is due to a "choke" price whereas for the two that feature constant markups this is due to a fixed cost of exporting. Holding fixed the trade elasticity, the gains from trade implied by the models can be ranked across these four models.

Of the four models, two feature CES preferences and fixed export costs, and so fall into the class of models considered in Arkolakis, Costinot and Rodríguez-Clare (2012), and two feature non-CES preferences and a choke price, and so fall into the class of models considered in Arkolakis et al. (2019). Within each class, the models differ in that one features our quality mechanism while the other does not. Note that when the quality mechanism is shut off but the variable markup is operating, as shown by equation (30), the benchmark model collapses to Jung, Simonovska and Weinberger (2019). Conditional on the trade elasticity, which is $\eta\theta$ when there is quality differentiation and θ when there is not, it follows from the parameter restriction that $\sigma - 1 < \theta \leq \eta\theta$ that the model with variable markups implies greater gains from trade than the model without variable markups. This result is due to an "excessive love of variety" implied by generalized CES preferences. Intuitively, with generalized CES

[22]This means that there are no specific trade costs and quality q is set to one for all firms. See Appendix G for the details.

[23]See details in Appendix H.

[24]This means that there are no specific trade costs, quality q is set to one and $\bar{x}$ is set to zero.

preferences, consumers obtain positive levels of utility simply for having the option to consume a good. Hence, goods available at the choke price lead to strictly positive levels of utility as is the case in models with fixed export costs and standard CES preferences. Our findings do not contradict Arkolakis et al. (2019) because they assume away positive utility provided by goods sold at the choke price.[25]

Moreover, conditional on the same parameter values across models and given the parameter condition that η is larger than one, the models with quality variation should generate lower gains from trade than those without endogenous quality, regardless of variable markups (see equation (29) vs. equation (30) and equation (31) vs. equation (32)). This is verified later in Section (6) by our quantitative results of welfare comparisons across models. However, an issue that arises in the quantitative comparisons provided below is that tariff variation alone cannot identify the trade elasticity when our quality mechanism is present and so differences in welfare estimates will also arise for purely quantitative reasons.

4 Quantification

This section describes how we solve, calibrate and simulate our benchmark model and the three alternative models. The first two subsections detail how we estimate and simulate the benchmark model. The final subsection discusses how we adjust our calibration strategy to estimate the three alternative models, each of which contains only a subset of the parameters of the benchmark model.

We first estimate the parameters of the benchmark model. There are two sets of parameters. The first set $\Theta_1 = \{\eta, \theta, \sigma_\varepsilon, \sigma\}$, including the inverse of quality scope, the productivity shape, the standard deviation of specific trade cost shocks, and the elasticity of substitution. The second set $\Theta_2 = \left\{\left\{w_j, P_{j\sigma}, P_j, fJ_i, T_{ij}^{\eta-1}\tau_{ij}, b_i, N_j\right\}_{i=1}^{I}\right\}_{j=1}^{I}$ includes all endogenous macro variables.[26] We show that our model specification enables us to identify Θ_1 without information about Θ_2. Therefore, we can first identify Θ_1, and then recover macro level parameters in Θ_2 through the structural equations implied by the model. We then simulate the model based on parameter estimations. Finally, we generate pseudo-Chinese exporters that is comparable with

[25]For models without quality mechanism, we also find that the model with variable markup yields larger welfare gains than the constant markup model (see detailed derivation in Appendix I.2). The differences between our results and those of Arkolakis et al. (2019) stem from the changes in firm productivity cutoff at the extensive margin due to the "excessive" love of variety feature in the generalized CES model which the footnote 12 in Arkolakis et al. (2019) was meant to rule out. Our generalized CES model is in the similar spirit to models with fixed exporting costs where discontinuity emerges. Once a firm chooses to export to a market, the increase in utility is discontinuous at zero consumption due to the existence of $\overline{x}$, so it is like jumping to a positive level of consumption even though a consumer might consume zero at the margin. This discontinuity feature under special conditions has been mentioned by Arkolakis et al. (2019), and our model falls into this special situation. Under such discontinuity, the welfare changes associated with the change in productivity cut-off is no longer infinitesimal. Thus, in our setup, the variable markup models yield higher gains from trade than the constant markup models. When the extensive margin effect is shut down, we obtain the same results as Arkolakis et al. (2019) that variable markup models yield lower gains from trade than constant markup models.

[26]In our calibration, we focus on 36 countries, i.e., $I = 36$.

the customs data and analyze the model fit by comparing the real data and model simulated data.

4.1 Parameterization

In this subsection, we first show how a gravity equation can be used to recover an important model parameter. Next, we show how the remaining parameters in the set Θ_1 can be recovered. Finally, we show how given estimates of the parameters in Θ_1 and the model's structural equations can be used to recover the parameters in Θ_2.

Gravity and the Two Trade Elasticities

The set $\Theta_1 = \{\eta, \theta, \sigma_\varepsilon, \sigma\}$ contains four key parameters of our model. We begin by discussing the estimation of θ. Following Caliendo and Parro (2015) and Arkolakis et al. (2018), we estimate θ from the coefficient on tariffs in a gravity equation. Taking the logarithm of equation (23) yields an empirical gravity equation for estimation:

$$\log\left(\frac{\lambda_{ij}}{\lambda_{jj}}\right) = \underbrace{\log\left[J_i b_i w_i^{-\theta\eta}\right]}_{S_i} - \underbrace{\log\left[J_j b_j \left(T_{jj}^{\eta-1}\tau_{jj} w_j^{-\eta}\right)^{\theta}\right]}_{S_j} - \theta\left(\eta-1\right)\log T_{ij} - \theta \log \tau_{ij}, \quad (33)$$

where S_i is the exporter fixed effect, and S_j is the importer fixed effect. We call the coefficient on $\log \tau_{ij}$ the *ad-valorem trade cost elasticity* and the coefficient on $\log T_{ij}$ the *specific trade cost elasticity*. Note that these coefficients are structural but identify different parameters. This is an important observation in that all four of the models for which we compare welfare implications below feature an ad-valorem trade cost elasticity that has a structural interpretation while only the two models with the "Washington Applies" mechanism feature the specific trade cost elasticity.

To estimate a trade elasticity, we must make auxiliary assumptions. First, we assume that both $\log T_{ij}$ and $\log \tau_{ij}$ are linear in bilateral pair geography. Second, we assume that the majority of the tariff variation observed for manufacturing goods are ad valorem, which is reasonable for manufactured goods.[27] Following Waugh (2010) and Jung, Simonovska and Weinberger (2019), we use a set of gravity variables to proxy for T_{ij} and for τ_{ij} through the following equations:

$$\begin{aligned}(\eta-1)\log\left(T_{ij}\right) &= \alpha^T + ex_i^T + \gamma_h^T d_h + \gamma_d^T \log\left(dist_{ij}\right), \\ \log \tau_{ij} &= \alpha^\tau + ex_i^\tau + \gamma_h^\tau d_h + \gamma_d^\tau \log\left(dist_{ij}\right) + \log tar_{ij},\end{aligned}$$

where α^T and α^τ are constants. As in Waugh (2010), we also add an exporter fixed effect, ex_i,

[27] Strictly speaking tariffs are not standard cost shifters like shipping costs, but we follow much of the literature in assuming that they are. For a discussion see Costinot and Rodríguez-Clare (2014) and Felbermayr, Jung and Larch (2013).

a set of three dummy variables, d_h, indicating whether (1) the trade is internal; (2) whether the two country use the same currency; (3) whether the two country use the same official language, and the logarithm of distance from country i to country j, $\log(dist_{ij})$. This yields the following estimating equation:

$$\log\left(\frac{\lambda_{ij}}{\lambda_{jj}}\right) = S_i - S_j - \theta\left(\left(\alpha^T + \alpha^\tau\right) + \left(ex_i^T + ex_i^\tau\right) + (\gamma_h^T + \gamma_h^\tau)d_h + \left(\gamma_d^T + \gamma_h^\tau\right)\log(dist_{ij})\right) - \theta\log tar_{ij} + \varepsilon_{ij} \tag{34}$$

where ε_{ij} is assumed to be Gaussian measurement error. Note how the coefficient on tariffs, the ad valorem trade cost elasticity, has a structural interpretation. It is the productivity distribution shape parameter θ. Further, also note that with an estimate of θ it becomes possible to back out from these estimates the aggregate trade cost $(T_{ij})^{\eta-1}\tau_{ij}$.

The bilateral trade share λ_{ij} is constructed following the method in Ossa (2014) by using the GTAP 9 data for the year 2004.[28] Bilateral gravity variables: $dist_{ij}$, d_h (common currency, common official language) is taken from the CEPII dataset. The tariff data is from WITS, where we compute the average tariff rate for all HS6 sectors of each destination to represent tar_{ij}.[29] We let $tar_{ij} = 1$ if trade is internal. We also let $tar_{ij} = 1$ if both i and j belongs to EU, NAFTA, ASEAN members countries. For the case of EU, we apply common external tariff by the EU for non-EU members. The summary statistics are presented in Table 4.

Table 4: Summary Statistics of Gravity Variables

Variable	Mean	Std. Dev.	Min.	Max.	N
$\log(\lambda_{ij}/\lambda_{jj})$	-5.221	1.842	-10.491	0	1296
$\log(tar_{ij})$	0.066	0.067	0	0.264	1296
$\log(dist_{ij})$	8.432	1.059	2.258	9.811	1296

The coefficients on the gravity variables and tariffs obtained by estimating equation (34) via OLS are shown in Table 5. The estimates on the standard gravity variables all of their expected sign and fall in common ranges for gravity equations (see Head and Mayer, 2014). For instance, a 10 percent increase in distance is associated with an approximately 7.65 percent reduction in the volume of trade. Most importantly, the coefficient of 6.1 on tar is sensible and is measured with high precision.[30] We now discuss the estimation of the model's other key parameters.

[28]The bilateral trade shares λ_{ij} are only constructed for our selected 36 countries. For any $i \neq j$, we first compute X_{ij} as the sum of trade flow from i to j across all GTAP sectors. We then compute X_{jj} as the total domestic output, X_j, minus its total export, $\sum_{i\neq j} X_{ji}$. We then compute $\lambda_{ij} = X_{ij}/\sum_i X_{ij}$. One important advantage of using GTAP is that we do not get missing/negative value for our constructed X_{jj}, and hence all the values for λ_{ij} are valid.

[29]2004 tariff data for Russia is not available. We use the year 2005 instead. We also try year 2002 as an alternative, the result is very similar.

[30]This number falls in the range of estimates in Arkolakis et al. (2018).

Table 5: Estimation of Gravity Equation

Dependent variable: $\log(\lambda_{ij}/\lambda_{jj})$	
$\log(tar_{ij})$	-6.097***
	(0.795)
$\log(dist_{ij})$	-0.765***
	(0.031)
Common language	0.349***
	(0.071)
Common currency	0.165*
	(0.086)
Same country Dummy	2.658***
	(0.139)
Importer Fixed Effects	YES
Exporter Fixed Effects	YES
Observations	1,296
R-squared	0.988

Notes: Standard errors in parentheses.

The Remaining Parameters of Θ_1

Our approach to estimating the remaining coefficients is very different. To identify the idiosyncratic dispersion in trade costs, σ_ε, the taste parameter σ, and the quality upgrading cost elasticity η, we make use of our estimate of θ, the model, and moments from firm-country-product data on unit values ($p_{ij}(\omega)$ in the model) and export values ($r_{ij}(\omega)$ in the model). The core of our estimation strategy involves using the first-order condition for price determination (6) and values of σ, σ_ε, and η to generate an artificial dataset that match the standard deviation of the logarithm of price charged by Chinese firms, the standard deviation of the logarithm of the corresponding sales, and the correlation of the logarithm of prices with the logarithm of sales.

We follow the simulated method of moments procedure in Eaton, Kortum and Kramarz (2011) and Jung, Simonovska and Weinberger (2019). In particular, we define $u \equiv b_c \varphi^{-\theta}$, where b_c denotes China's productivity. The cumulative distribution of u can be shown as follows

$$\Pr(U < u) = \Pr\left(b_c \varphi^{-\theta} < u\right) = \Pr\left(\varphi > \left(\frac{b_c}{u}\right)^{\frac{1}{\theta}}\right) = u.$$

The conditional productivity entry cutoff $\varphi^*_{ij}(\varepsilon)$ can also be written in terms of u,

$$u^*_{cj}(\varepsilon) = b_c \left[\frac{\eta^\eta}{(\eta-1)^{\eta-1}} T_{ij}^{\eta-1} \tau_{ij} w_i^\eta \left(\tilde{p}_j^*\right)^{-\eta} \varepsilon^\eta\right]^{-\theta}. \quad (35)$$

Equation (35) implies that a firm that has received cost shock ε will export when $u < u^*_{cj}(\varepsilon)$. Importantly, $\tilde{u} \equiv \frac{u}{u^*_{cj}(\varepsilon)}$ follows a uniform distribution from $(0, 1]$ where the highly efficient firms with $\tilde{u}$ close to zero and the marginal firms with $\tilde{u}$ close to 1. We first draw 1,000,000 realizations of $\tilde{u}$ from uniform distribution on $(0, 1]$. Each draw corresponds to a simulated exporters. For each exporter, we draw I (=36) destination specific realizations of $\tilde{\varepsilon}$s from the standard normal distribution. Note that by construction, $\tilde{u} \equiv \left(\frac{\varphi}{\varphi^*_{cj}(\varepsilon)}\right)^{-\theta}$ and $\tilde{\varepsilon} \equiv \frac{1}{\sigma_\varepsilon} \log \varepsilon$, thus the true productivity φ and the real cost draw ε can be recovered whenever necessary.

Combining equations (9), (10), and (11) with (6), yields the following expression:

$$\sigma \tilde{u}^{\frac{1}{\eta\theta}} = \left(\frac{\tilde{p}_{ij}(\tilde{u})}{\tilde{p}^*_j}\right)^{\sigma+1} + (\sigma - 1)\frac{\tilde{p}_{ij}(\tilde{u})}{\tilde{p}^*_j}. \tag{36}$$

Note that the inverse of the left hand side follows a Pareto distribution with location parameter 1 and shape parameter $\eta\theta$. We can recover $\frac{\tilde{p}_{ij}(\tilde{u})}{\tilde{p}^*_j}$ according to the previous equation for each $\tilde{u}$. To connect the implied pricing behavior in the model with the Chinese firm-product-country data, we define the following transformation:

$$p_{ij}(\tilde{u}, \tilde{\varepsilon}) \equiv \frac{\tilde{p}_{ij}(\tilde{u})}{\tilde{p}^*_j} c_{ij}(\tilde{\varepsilon}) \frac{\tilde{p}^*_j}{\tilde{c}_{ij}(\tilde{u})},$$

where $c_{ij}(\tilde{\varepsilon}) = \frac{\eta}{\eta-1} w_i T_{ij} \exp(\sigma_\varepsilon \tilde{\varepsilon})$ is the endogenous (unadjusted) marginal cost of firms. Using equations (9) and (11) and taking logarithms yields

$$\log p_{ij}(\tilde{u}, \tilde{\varepsilon}) = \log\left(\frac{\tilde{p}_{ij}(\tilde{u})}{\tilde{p}^*_j}\right) + \sigma_\varepsilon \tilde{\varepsilon} - \frac{1}{\eta\theta}\log(\tilde{u}) + \log\left(\frac{\eta}{\eta-1} T_{ij} w_i\right) \tag{37}$$

this implies that the standard deviation of log exporter price, once we subtract the destination average to eliminate the constant term (the last term on the right), will only depend on the parameter set $\Theta_1 = \{\eta, \theta, \sigma_\varepsilon, \sigma\}$, and is not destination specific.

Making similar transformations for the logarithm of the sales revenue of a firm, given by (4), we obtain:

$$\log r_{ij}(\tilde{u}) = \log\left(\frac{\tilde{p}_{ij}(\tilde{u})}{\tilde{p}^*_j}\right) + \log\left[\left(\frac{\tilde{p}_{ij}(\tilde{u})}{\tilde{p}^*_j}\right)^{-\sigma} - 1\right] + \log(\overline{x} L_j), \tag{38}$$

This expression shows that the standard deviation of country-product exports by Chinese firms, once it has been demeaned by subtracting its sector-destination mean, depends only on parameters $\eta\theta$ and σ. Notice that two types of relationships here are relevant. First, both parameters drive the standard deviation of $\log r_{ij}(\tilde{u})$, while only σ governs the dependence of $\log r_{ij}(\tilde{u})$ on $\tilde{p}_{ij}(\tilde{u})/\tilde{p}^*_j$. Moreover, we can obtain the correlation between log-sales and log-price given parameters $\eta\theta$, σ_ε, and σ. Our discussion suggests that these three moments are sufficient to jointly identify our three parameters $\eta\theta$, σ_ε, and σ via simulated Generalized Method of Moments, while our gravity estimate of θ allows us to separate η from θ.

We now summarize the estimation strategy. First, we calibrate σ to target the standard deviation of the log of export sales. To see this, notice that in equation (38), $\tilde{p}_{ij}(\tilde{u})/\tilde{p}^*_j$

is bounded from 0 to 1 (the marginal exporter to destination j takes value 1 while for the most productive firms it tends toward 0). An increase in σ makes sales more responsive to productivity and so leads to larger sales dispersion. Second, we choose σ_ε to target the standard deviation of the log of export price. Firms' marginal cost depends on the trade cost draw $\tilde{\varepsilon}$ (see equation (37)), so greater dispersion of these shocks yields greater dispersion of price. Third, the correlation between log-sale and log-price helps to identify $\eta\theta$. In a model without quality, as in Jung, Simonovska and Weinberger (2019), price and sales exhibit negative relationship because the productive firms have lower marginal cost. This negative relationship is overturned here because high productivity firms produce higher quality which allows firms to raise their prices. This mechanism can also be seen from the $\log(\tilde{u})$ term in equation (37): a lower $\tilde{u}$ implies a higher real efficiency and hence higer price and sales. The distribution of $\tilde{u}$ is governed by the value of $\eta\theta$. We now turn to our construction of the data moments.

To construct the three micro moments for the data, we use the Chinese customs' ordinary trade data at the year 2004. We aggregate the data into firm-country-HS6 level, construct our data moments for by each country-HS6 pair and choose the median among them. The parameters are jointly identified through the following minimization routine:

$$\min_{\eta\theta,\sigma_\varepsilon,\sigma} \left\{ \left[m^D - m^M(\eta\theta,\sigma_\varepsilon,\sigma)\right]' W \left[m^D - m^M(\eta\theta,\sigma_\varepsilon,\sigma)\right] \right\}$$

where m^D is the (column) vector that contains the data moments, and $m^M(\eta\theta,\sigma_\varepsilon,\sigma)$ contains the corresponding model moments. W is identity weighting matrix.

Following Jung, Simonovska and Weinberger (2019), we check the sensitivity of our quantitative results by comparing the estimates from our exactly identified benchmark to those obtained from an over-identified specification. In the over-identification specification, we target a larger set of the moments from the distribution of sales and prices (e.g., the 90-to-10, 90-to-50, and 99-to-90 percentile ratios of log sales and log prices). These additional moments are desirable given that the focus of the quantitative exercise in this paper is to match both sales and price dispersions as well as the relationship between the two.

Solving for Θ_2

The set of Θ_2 includes all endogenous macro variables. We begin by describing how we uncover wages, the measure of total entrants per market, and aggregate prices statistics.

To solve wage w_i for each country, we use the labor market clearing condition, which is given by

$$w_i L_i = \sum_j X_{ij} = \sum_j \lambda_{ij} w_j L_j.$$

Here we normalize the wage in US to be 1 so that every other countries' wages are all relative to the US. Market size L_i is proxied by total population of that country, which is from the CEPII dataset. Note that market size immediately pins down the number of entrants per country,

fJ_i, from equation (21).

To recover b_j, we use the importer fixed effect from the gravity estimation in equation (23) which is

$$S_j = \log\left[(fJ_j)\, b_j\, (w_j)^{-\eta\theta}\right],$$

where S_j is the estimated importer fixed effect.[31] The bilateral trade cost $\left(T_{ij}^{\eta-1}\tau_{ij}\right)$ can also be recovered from the gravity equation (23).[32] Finally, we solve for the mass of firms that serve country j, N_j, using equation (13), and equation (17). These two equations when combined yield

$$N_j = \frac{(\eta-1)^{\frac{\eta-1}{\eta}}}{\eta\bar{x}\left[\beta_\sigma - \beta\right]} \left(T_{ij}^{\eta-1}\tau_{ij}\right)^{-\frac{1}{\eta}} \frac{w_j}{w_i} \left(\frac{\kappa J_i b_i}{N_{ij}}\right)^{\frac{1}{\eta\theta}}.$$

Having recovered all the variables in this expression up to the constants, we can use Chinese custom data to compute the total number of firms that export from China to country j, $N_{China,j}$, except for China itself. Then N_j $(j \neq China)$ can be computed from the above equation.

4.2 Model Simulation

Given estimates for all the key parameters, we can simulate the model to assess its ability to reproduce the facts that were illuminated in Section 2. We follow the procedures below to construct the full panel of model generated exporters:

(1) For each draw of $\tilde{u}$, we construct entry hurdles $u_{cj}^*(\tilde{\varepsilon})$ for each country j using equation (35).

(2) For each $\tilde{u}$, we compute $u_{cj}^{*\max} = \max_{j\neq China}\left\{u_{cj}^*(\tilde{\varepsilon})\right\}$. This is the minimum requirement productivity for a firm to sell their product in countries other than China. We then construct $u = u_{cj}^{*\max}\tilde{u}$ using our draw of $\tilde{u}$ in step (1). Because in the model, the measure of firms that export from China to country j is $u_{cj}^{*\max}$, our artificial exporter u is assigned a sampling weight of $u_{cj}^{*\max}$.

(3) For each u, we set the export status δ_{cj} indicating whether firm u exports to j to be given by

$$\delta_{cj}(u) = \begin{cases} 1, & \text{if } u \leq u_{cj}^*(\tilde{\varepsilon}) \\ 0, & \text{otherwise} \end{cases}$$

(4) We recover firm level variables, which include productivity, price and sales. First, we obtain firm level productivity from $\varphi = \left(\frac{b_c}{u}\right)^{\frac{1}{\theta}}$. Second, we construct exporter-destination quality $q_{ij}(\varphi,\varepsilon) = \left(\frac{\varphi}{\eta-1}\frac{T_{ij}}{\tau_{ij}}\right)^{\frac{1}{\eta}}$. Note that at this juncture, we have to take a stand on the relative magnitudes and cross-country variation in T_{ij} and τ_{ij}. Motivated by the discussion in Hummels and Skiba (2004), we assume that T_{ij} specific costs account for all of the geographic

[31] In the above regression, we've added both the importer and exporter fixed effect. This induces multi-collinearity. To avoid this, we follow Levchenko and Zhang (2016) and normalize the importer fixed effect S_j for US to 0. Essentially, we choose US for the reference country, and the importer fixed effect estimates for all other countries are all relative to the reference country.

[32] Note that we set $T_{jj}^{\eta-1}\tau_{jj} = 1$ for all j.

variation in the gravity equation and τ_{ij} is driven exclusively by tariffs. Finally, we compute firm-level prices that are not adjusted for quality:

$$p_{ij}(\tilde{u}, \tilde{\varepsilon}) \equiv \frac{\tilde{p}_{ij}(\tilde{u}, \tilde{\varepsilon})}{\tilde{p}_j^*} \tilde{p}_j^* q_{ij}(\tilde{u}, \tilde{\varepsilon}),$$

where $\tilde{p}_{ij}(\tilde{u}, \tilde{\varepsilon})$ are solved through the pricing equation (36). Finally, firm sales can be constructed from equation (4).

In summary, after dropping non-exporting Chinese firms, we have constructed a dataset that contains one million exporting firms that can export to a maximum of $(I - 1)$ countries.

4.3 Estimation of the Alternative Models

The procedures to estimate and to simulate the three alternative models are similar to those for the benchmark model. Because we are only interested in how these models fit the joint distribution of firm-level prices and sales and because the parameter estimates in the set Θ_1 are those that are necessary to compute welfare gains, we confine our discussion to these parameters.

As all four models feature a structural ad-valorem trade cost elasticity, we use the coefficient from the gravity equation above to discipline the value of θ across all models. Conditional on the same θ, we then use the same set of moments, namely, the standard deviation of sales and prices as well as the correlation between sales and prices, to jointly calibrate the key parameters across models. It is important to use the same set of moment conditions to yield consistent parameter estimations across models. As pointed out by Simonovska and Waugh (2014), it could be the case that targeting the same moments in the data results in parameter estimates in different models that are different, and this would matter for welfare quantification of alternative models. Note that in models in which there is no "Washington Apples" mechanism the shocks to specific trade costs do not exist and there is no quality to adjust so that σ_ε and η are not estimated. We now turn to the estimation results and the assessment of model fit.

5 Results and Model Fit

In this section, we report the parameter estimates for the four alternative models and compare their ability to fit the data. We begin with the benchmark model by reporting the parameter estimates for Θ_1 for both the exactly identified and the over identified cases. We then report summary statistics for our estimates of the parameters in Θ_2 calculated using the exactly identified parameters in Θ_1 and compare these statistics to the data. We conclude the section by reporting the parameter estimates for each of the alternative models and demonstrate that fitting the key feature of the joint distribution of prices and revenues requires both variable markups and endogenous quality. Omitting either mechanism makes the model unable to generate the positive correlation between prices and revenues across firms and markets.

5.1 Parameter Estimates and Fit: Benchmark Model

We begin with our estimates of the key parameters of the benchmark model which are shown in the following table.

Table 6: Calibration of Θ_1

Parameter	symbol	value (Exact ID)	value (Over ID)
elasticity of substitution	σ	4.8179	5.4819
std. dev. of cost shock	σ_ε	0.6004	0.7599
inverse of quality scope	η	1.7111	1.2193
trade elasticity w.r.t. tariff	θ	6.0973	6.0973

Table 7: Data Targets and Simulation Results

moment	data	model (Exact ID)	model (Over ID)
Panel A: targeted moments			
std(log(sale))	1.3916	1.3916	1.4935
std(log(price))	0.6017	0.6017	0.7613
corr(log(sale), log(price))	0.0543	0.0543	0.0541
trade elasticity w.r.t. tariff	6.0973	6.0973	6.0973
log(sales) 90-10	4.1551	-	1.9511
log(price) 90-10	2.0297	-	3.6124
log(sales) 90-50	2.0369	-	0.9752
log(price) 90-50	1.0451	-	1.6070
log(sales) 99-90	1.3814	-	0.7954
log(price) 99-90	1.3242	-	1.4837
Panel B: non-targeted moments			
exporter domestic sales advantage	1.7152	2.0831	3.3971
firm frac. with exp. intensity (0.00, 0.10]	38.2064	27.2619	64.4882
firm frac. with exp. intensity (0.10, 0.50]	35.5425	72.5898	35.5118
firm frac. with exp. intensity (0.50, 1.00]	26.2511	0.1483	0.0000

Notes: The targeted moments are constructed from customs data, which covers the universe of all exporters and importers. The non-targeted moments are constructed from the merged sample based on customs data and Chinese Manufacturing Survey data provided by NBSC (National Bureau of Statistics of China), because we need both exporters and non-exporters in the non-targeted moments to check exporter domestic sales advantage, and we also need total sales information from the NBSC data to compute export intensity.

Table 6 lists our calibration results for the key set of parameters Θ_1. Table 6 shows that the parameter estimates obtained under both exact identification and over identification strategies are similar. As in Jung, Simonovska and Weinberger (2019), when we try to match the tails of the sales and prices distribution in the over identification case, σ increases to match the large dispersion in the firm-level data. Compared with the exact-identified case, the over-identified model slightly overpredicts the dispersion of firm sales and prices.

Table 7 further presents the data targets and the simulation results for both targeted moments (see Panel A) and non-targeted moments (see Panel B). Given the trade elasticity, our model matches the targeted moments relatively well although it underestimates the extreme skewness in firm sales and overestimates the skewness in firm prices.

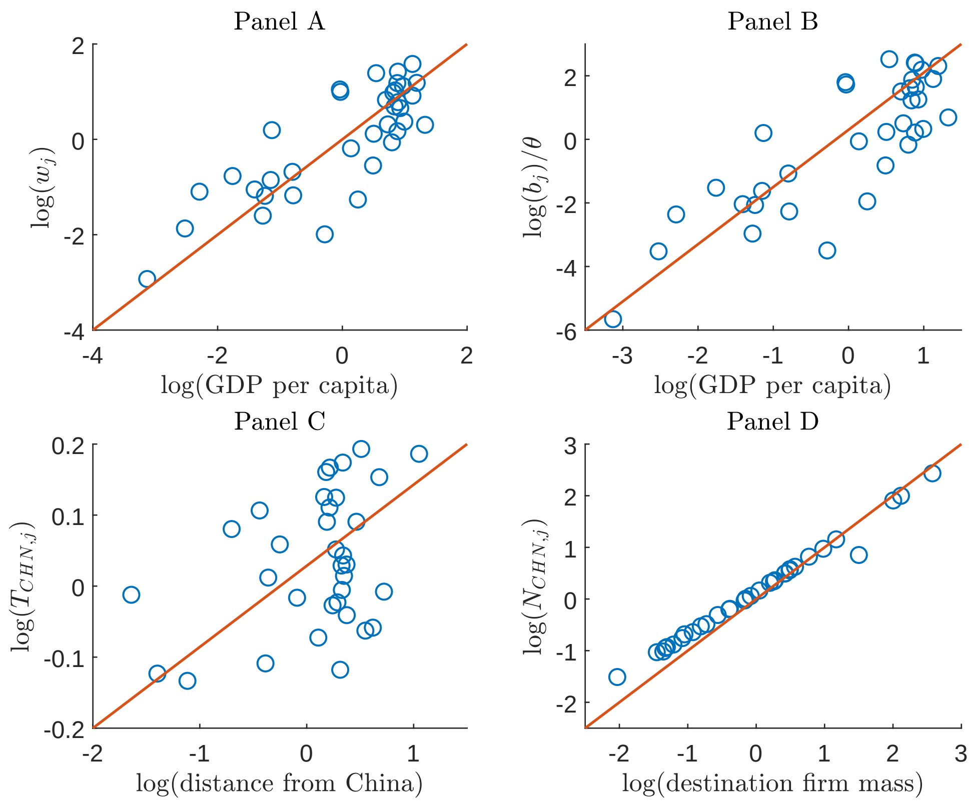
Our non-targeted moments are exporter sales advantage, measured as the ratio of domestic sales of exporters to non-exporters, and exporters' export intensity measured as the share of output that is exported. There are three measures of export intensity: the share of firms that export less than 10 percent of their total revenue, the share of firms that export between 10 and 50 percent of their output, and the share of firms that export more than 50 percent of their output. All non-targetted moments were computed using a merged sample between customs data and the NBSC manufacturing survey data. Here, we see that the overidentified specification does a better job fitting the export intensity distribution than the exactly identified model.

The markup distribution formula in our model is the same as in Jung, Simonovska and Weinberger (2019). Yet, we fit to different moments and different parameter values are obtained. Thus, our model's generated markup distributions have a relatively thin tail than those in Jung, Simonovska and Weinberger (2019). Our estimate of the elasticity of substitution implies that the upper bound for markups would be $\frac{\sigma}{\sigma-1} = 1.26$. Given that $\theta = 6$, the model's generated markups distribution has a relative thin-tail. Thus, the average markup charged by exporters in our model is lower than that of Jung, Simonovska and Weinberger (2019). More specifically, our model implied average markup is 1.0229, the log(markups) 99-50 percentile ratio is 0.0853, and the log(markups) 90-50 is 0.0517. We plot the model simulated markups and sales distribution in Figure 9 in Appendix K.

We now check the model's fit for the solution to our model. The four panels of Figure 5 demonstrate the fit of our model to data. The first panel shows that the logarithm of the wage by country relative to country averages implied by the model closely follows the logarithm of GDP per capita relative to country averages as reported in the CEPII data set, explaining over 80% of the variation in cross country incomes. In the second panel, we plot the implied productivity by country versus its GDP per capita. This too shows a very strong fit. In the third panel, we plot model generated specific trade costs against the real data of distance from China to each destination country and observe a very strong positive slope. In the last panel is the number of Chinese firms that serve a particular country predicted by the model against the actual number of entrants. Our model's predictions closely mirror the variation across countries in terms of the extensive margin.

We now turn our attention to the key object of interest in our paper, the relationship between the price charged by a firm and its sales. Figure 6 illustrates the price and sales relationship for both data and model. For the data, we first construct firm's normalized sales

Figure 5: A Check on the Solution of the Model

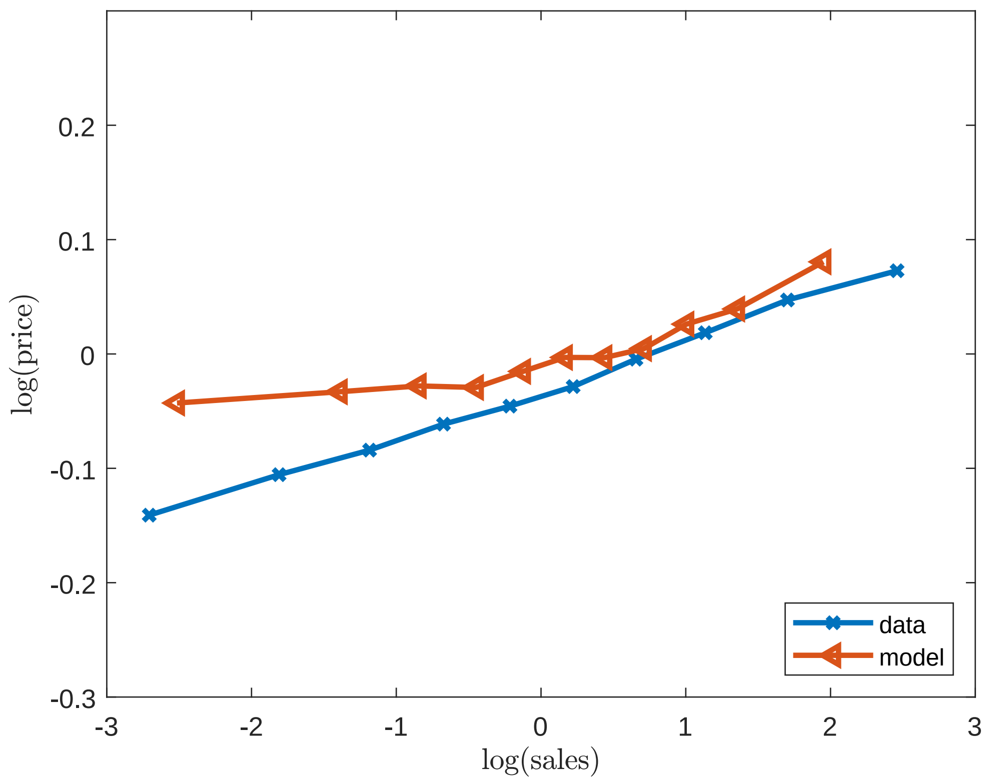


by subtracting each firm's log sales by its HS6×destination average. We apply the same treatment for the firm's price. Then, for each HS6×destination pair, we sort firms' normalized sales into 10 deciles. In this step, we require that each HS6×destination have at least 10 firms so that the 10 deciles can be properly obtained. We then compute the median of both the normalized price and sales at each decile for each HS6×destination pairs. We finally aggregate the median value for all HS6×destination pairs, leaving only one value for each sales decile. For the model, we follow a similar procedure. Thus, each dot in the figure represents deviations of log sales from their relevant industry mean relative to the deviations of log price from their relevant industry mean.[33]

Quantitatively, the model traces the data reasonably well. In the data, when log firm sales increase from -3 to +3, the logarithm of the firm price increases by 0.25, whereas in the model, it increases by about 0.15. Hence, the model explains about 60% of the positive relationship between price and sales. The increase for the model mostly comes from large firms, i.e. firms that have higher sales than average. For the small firms, the model predicts a higher price level than that of the data. The reason appears to stem from the endogenous cut-off price induced by non-homothetic preferences that limit the scope for variation among small firms.

[33] Figure 6 also suggests the positive correlation between prices and market share since market share is equal to firm sales over the total sales by all Chinese exporting firms in that destination market. Thus, the relationship between prices and market share would be the same as the relationship between prices and sales.

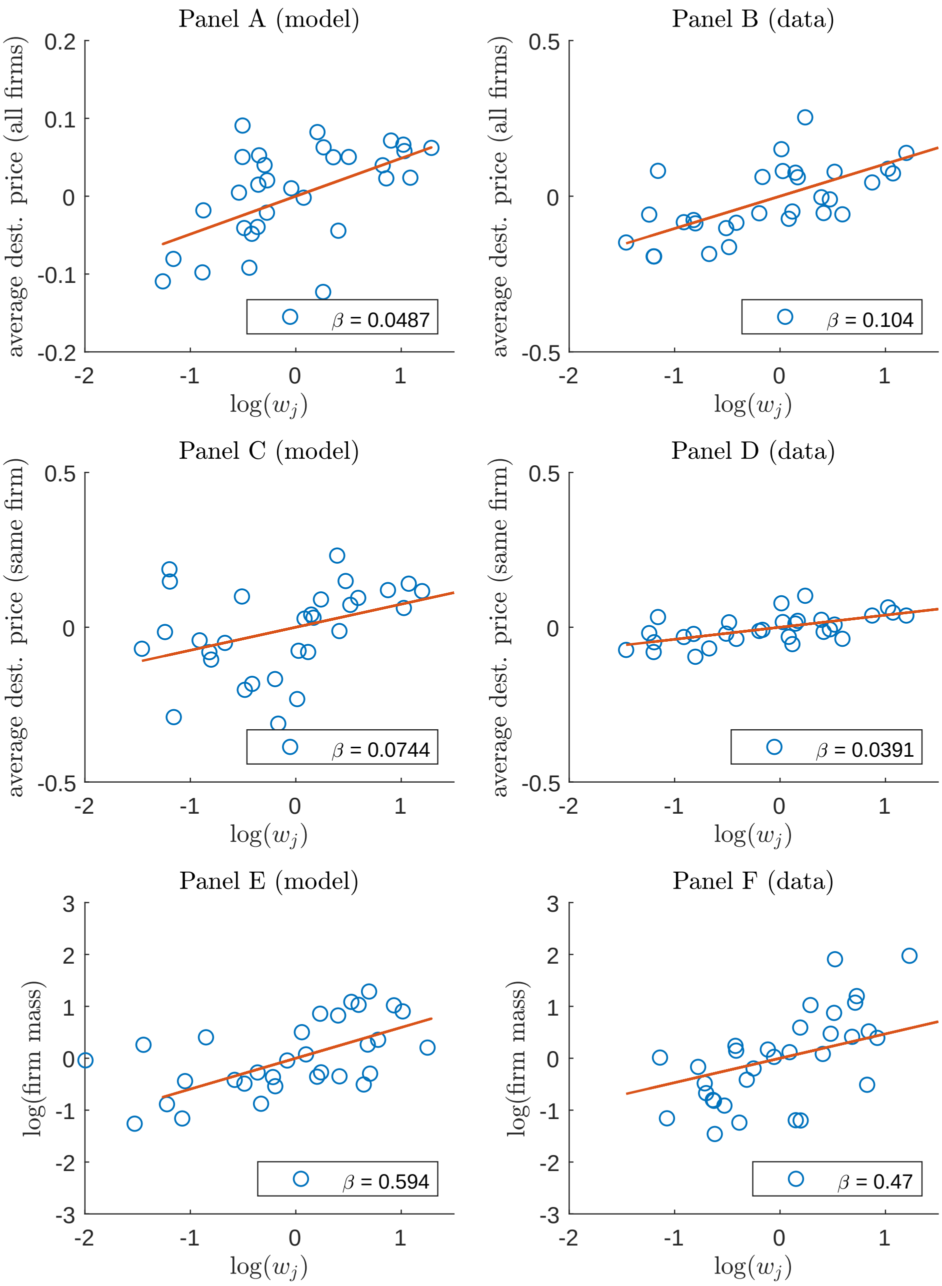
Figure 6: Model Fit: Price-Sales Relationship



Note that the positive relationship between prices and sales in Figure 6 also highlights the importance of the interaction of variable markups and endogenous quality. This is because, with endogenous quality under monopolistic competition, variable markups as in Jung, Simonovska and Weinberger (2019) are essential for our model, which aims to reconcile the price dispersion across firms and across markets, to generate positive relationship between sales and prices. If firms were to set constant markups over marginal costs, there would be no correlation between firms' sales and prices which can be seen from the marginal cost formula $c_{ij}(\tilde{\varepsilon}) = \frac{\eta}{\eta-1} w_i T_{ij} \exp(\sigma_\varepsilon \tilde{\varepsilon})$. In other words, the variable markup mechanism is crucial for our model that features both endogenous quality and pricing-to-market to deliver factual relationship of prices and sales. On the other front, there are existing studies that rely on the quality mechanism alone to generate this positive relationship, such as Johnson (2012), but these endogenous-quality models are not able to explain the facts across countries that firms set higher export prices in higher-income destinations and that more firms export to higher-income destinations. Our model is to generate exporter pricing pattern both within market and across markets in a unified general equilibrium framework.

We conclude this section by considering the model fit along dimensions not directly fit in our calibration procedure. We first consider the within and across firm variation in prices as a function of the GDP per capita of the destination country. Figure 7 shows this relationship for the model in the left-hand panels and in the data in the right hand panels. The top two

Figure 7: Model Fit: Price-Wage Relationship and Entrants-Wage Relationship



Notes: In the top two panels, we normalize each exporter's price by it's price at USA ($\log\left(p_{CHN,j}\left(\varphi,\varepsilon\right)/p_{CHN,US}\left(\varphi,\varepsilon\right)\right)$). we then calculate the average destination price as the mean of this normalized price across firms on each destination. For the bottom two panels, we calculate the average destination price as the simple average of log price for all exporters on that destination. For the model, w_j is model predicted wage rate; for the data, w_j is the 2004 destination GDP per capita in CEPII. For consistency with our empirical exercise, we control for log destination population, and log distance for both the data and the model. Since the model does not have an exact counterpart for distance, we thus use T_{ij} as a proxy.

panels are the variation across country within firms (intensive margin) and the bottom two panels are the relationships averaged across all firms (intensive and extensive margin). The model predicts a slightly stronger correlation between price and GDP per capita than the data but slightly less variation than the average across all firms. Both deviations can be understood with respect to the price-revenue relationship shown in Figure 6. Looking at only the intensive margin disproportionately picks up firms in the higher end of the productivity distribution that have high prices and high revenue, while the average price that includes the extensive margin picks up the small firms whose behavior the model has trouble fitting.

We now look more closely at the extensive margin in Figure 7. The panel E is the model prediction of the measure of entrants as a function of country per capita income while the panel F is the actual data. The model correctly predicts a positive relationship between the two, but there is slightly less variation in the model predictions than there is in the data. In addition, we also check the relationship between firm sales, prices and quality with market size (measured by the product of population and wage) and plot those positive relationships simulated by the model in Figure 10 in Appendix K.

5.2 Parameter Estimates and Model Fit: Alternative Models

We now present the parameter estimates and assess the fit of alternative models. Recall that each model has been fit to the same set of moments but differ in the mechanisms that are available to fit the data. The various parameter estimates for each calibration are shown in Table 8. By observing how the quality of the fit changes as mechanisms are removed, we can assess how large the "Washington Apples" effect is in our benchmark model. For comparison purposes, we show the parameter estimates for our benchmark model in the first column.

Table 8: Parameter Values of the Alternative Models

parameters	Bench	no q	con mkp	no q, con mkp
σ	4.818	1.210	22.682	7.086
σ_ε	0.600	-	0.602	-
η	1.711	-	3.558	-
θ	6.097	6.097	6.097	6.097

Table 9: Fit of the Alternative Models

moments	data	Bench	no q	con mkp	no q, con mkp
std(log(sale))	1.392	1.392	1.262	1.000	0.999
std(log(price))	0.602	0.602	0.084	1.000	0.164
corr(log(sale), log(price))	0.054	0.054	-0.767	-0.000	-1.000
trade elasticity w.r.t. tariff	6.097	6.097	6.097	6.097	6.097

Consider the fit of the three alternative models to the key moments. In all models, the dispersion of price and sales are matched but our benchmark model does a better job. More importantly, the positive relationship between price and sales would only be matched through our benchmark model (see Table 9). This is also related to the previous discussion in the theory part after the illustration of model mechanisms in Figure 4 when comparing our benchmark model with alternative models in the literature. Not all alternative models can reconcile the dispersion of prices and sales and the correlation between the two simultaneously.[34] Without the "Washington Apples" effect interacting with the variable markup induced by non-homothetic preferences, it is not possible to generate a positive relationship between sales and observed prices at the firm level. With the "Washington Apples" effect, it is in principle possible that a configuration of parameters would allow for a positive correlation, but given the parameters estimated to fit the full set of moments this is not the quantitative outcome. Lacking the variable markups the correlation in the "con mkp" model is essentially zero.

6 Welfare Analysis

In this section we compare the welfare implications of the four different models. As discussed earlier in the text, theory tells us that conditional on the aggregate trade elasticity, the models have different implications for the magnitude of the gains from openness. In addition to having different theoretical reasons for different welfare implications, the different parameter estimates that obtain in practice when fit to the same data make the welfare implications quantitatively different.

Table 10: Welfare Comparison

country	Bench	no q	con mkp	no q, con mkp
CAN	5.925	36.196	2.519	8.676
DEU	3.934	25.566	1.662	5.789
FRA	3.478	22.929	1.468	5.124
GBR	4.706	29.857	1.993	6.912
JPN	1.292	9.125	0.542	1.914
USA	2.130	14.647	0.895	3.148
⋮	⋮	⋮	⋮	⋮
MEDIAN	4.403	28.200	1.863	6.473

Table 10 shows the various estimates of the gains from trade by each of the models for a subset of the countries in our dataset (see Appendix J for all countries). Column 1 shows the gains from trade estimated from our benchmark model. Column (2) corresponds to the model with only variable markups but without endogenous quality ("no q") which falls within the type of models analyzed by Jung, Simonovska and Weinberger (2019) and Arkolakis et al.

[34] This is also the reason why the estimated σ is a very low value in the "no q" model and a very large value in the "con mkp" model. For example, in the "no q" model, there should be no correlation between sales and prices but we force the elasticity of substitution σ to match all three moments (std(log-sale, std(log-price), and corr(log-sale, log-price)). Thus, a low value of $\sigma = 1.21$ is obtained.

(2019) (henceforth, an ACDR-type model). Column (3) corresponds to the gains from trade in the model without variable markup but with endogenous quality ("con mkp"). The last column refers to a typical Melitz-type model with neither variable markup nor the "Washington Apples" mechanism ("no q, no var mkp") which belongs to the ACR-type model.

We begin our discussion of the results by first comparing the models with variable markups versus those with constant markups. As noted earlier, the quantitative results in Table 10 are consistent with the previous discussion of welfare formulas in equations (29)-(32): the "Bench" model (with median gains of 4.40%) yield greater welfare gains than the "con mkp" model (with median gains of 1.86%); the "no q" model tend to produce gains from trade that exceed those of the "no q, con mkp" model with median gains of 28.2% versus 6.48%.

Turning now to the welfare gains due to "Washington Apple" effects, when comparing the gains from trade under the models with endogenous quality versus those without endogenous quality (regardless of variable markups), we find that the models with quality variation always generate lower gains from trade (see Column 1 versus Column 2 and Column 3 versus Column 4 in Table (10). This can also be verified by directly comparing the welfare formulas in equations (29)-(32) conditional on the same parameter values across models and given the parameter condition that η is larger than one.

Given the recalibration of the each of the models, it is ultimately true that the particular ordering of welfare gains regarding the quality mechanism across models that obtains is quantitative. Here, the distinction between the specific trade cost elasticity and the ad-valorem trade cost elasticity is important.[35] For instance, if we were to set $\tau_{ij} = T_{ij}$, we would obtain the "true" trade elasticity with respect to trade costs is $\eta\theta$ in the models with endogenous quality. Our decision to impose that the ad-valorem trade cost elasticity is the same across models is born in part of a desire to highlight the fact that all models have such an elasticity but models with specific-trade costs have an additional trade elasticity with no analog in standard models.

We now turn to a comparative static that also highlights the complications that arise in models with ad-valorem trade costs, specific trade costs, and endogenous quality upgrading.

7 Comparative Static

In this section we show that the impact of trade cost shocks on prices depends crucially on the nature of the shock. Consider a 5% increase in trade costs between country i and j as measured by $T_{ij}^{\eta-1}\tau_{ij}$. As can be seen in proposition 2 and in the gravity equation, whether this increase was due to an increase in $T_{ij}^{\eta-1}$ or τ_{ij} or some mixture of the two has no bearing on welfare or trade volume effects of the liberalization. As shown in this section, there are very big differences in the effect of these trade liberalizations on prices. Intuitively, an increase in

[35] Note that the above results are obtained by assuming tariff to act as cost shifters and using tariff to measure trade elasticity. However, as discussed briefly in Costinot and Rodríguez-Clare (2014) and Felbermayr, Jung and Larch (2013), tariffs could also be viewed as revenue shifters which would lead to a different estimation of trade elasticity instead of viewing tariffs as iceberg trade costs.

$T_{ij}^{\eta-1}$ raises the cost of serving the market and induces quality upgrading which leads to higher prices, whereas an increase in τ_{ij} induces firms to reduce their quality. Combined with the extensive margin effect through a change in firm productivity cutoff after increases in trade costs, the overall effects on average export prices are different for two types of trade costs.

In this section we demonstrate how these shocks lead to changes in prices quantitatively and then contrast the price effects of a 5% increase in ad valorem trade cost with an equivalent increase in specific trade cost. In addition, we check the effect of two types of trade costs shock on the distributional moments of prices, sales, and markups.

Applying "hat" algebra to the choke price $\tilde{p}_j^*$ and equations (12) and (13), it is straightforward to solve $\widehat{\tilde{p}}_j^*$ and $\widehat{\varphi}_{ij}^*$ according to the following two equations:[36]

$$\widehat{\tilde{p}}_j^* = \frac{\widehat{w}_j}{\sum_i \lambda_{ij} \left(\widehat{\varphi}_{ij}^*\right)^{-\theta}}, \text{ and} \tag{39}$$

$$\widehat{\varphi}_{ij}^* = \widehat{T}_{ij}^{\eta-1} \widehat{\tau}_{ij} (\widehat{w}_i)^{\eta} \left(\widehat{\tilde{p}}_j^*\right)^{-\eta}, \tag{40}$$

where $\widehat{w}_j$ can be solved from the system of equations (27). We can obtain other macro variables in a similar way by applying the hat algebra.

Next, we re-simulate the model to generate pseudo exporters using our solved macro variables after the trade shock. We use the same firm productivity draw (φ) and cost shock draw (ε) in the benchmark simulation. This guarantees that our comparative statics are performed on the same set of firms and cost draws, and all the changes are solely driven by the change in T_{ij} or the change in τ_{ij}. Specifically, for a firm with productivity φ and cost draw ε, we construct after-shock firm price using

$$\left(p_{CHN,j}\left(\varphi,\varepsilon\right)\right)' = \left(\frac{\tilde{p}_{CHN,j}\left(\varphi,\varepsilon\right)}{\tilde{p}_j^*}\right)' \left(\tilde{p}_j^*\right)' \left(q_{CHN,j}\left(\varphi,\varepsilon\right)\right)',$$

where $\left(\tilde{p}_{CHN,j}\left(\varphi,\varepsilon\right)/\tilde{p}_j^*\right)'$ depends on $\left(\varphi/\left(\varphi_{CHN,j}^*\left(\varepsilon\right)\right)'\right)^{\frac{1}{\eta}}$ via the firm pricing equation (36) and where $\left(\varphi_{CHN,j}^*\left(\varepsilon\right)\right)' = \left(\varphi_{CHN,j}^* \widehat{\varphi}_{CHN,j}^*\right)\varepsilon^{\eta-1}$ denotes the after-shock productivity cut-off.[37] Similarly, $\left(\tilde{p}_j^*\right)' = \widehat{\tilde{p}}_j^* \tilde{p}_j^*$ is the after-shock quality adjusted choke price and $\left(q_{CHN,j}\left(\varphi,\varepsilon\right)\right)' = \left(\varepsilon T_{CHN,j}' \varphi/\left(\eta-1\right)\tau_{CHN,j}'\right)^{\frac{1}{\eta}}$ is the after-shock optimal quality choice. Finally, we compute the mean of log-price across firms for each destination.

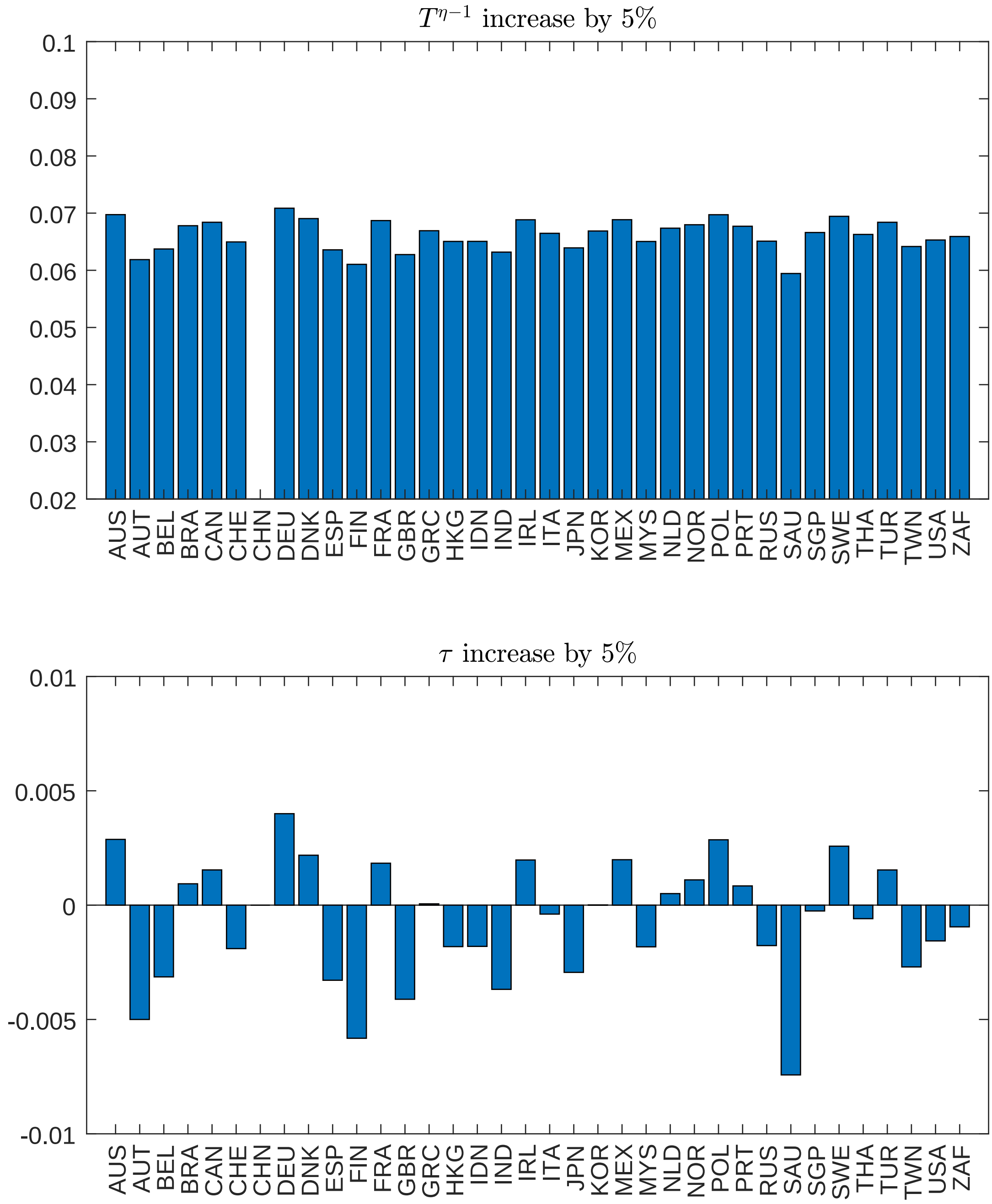
Figure 8 shows the results of our comparative static. The top panel shows the impact of $\widehat{T}_{ij}^{\eta-1} = 1.05$ for $i \neq j$ on average export prices set by our model simulated Chinese firms across countries in our data set while the bottom panel shows the results across the same set of countries for $\widehat{\tau}_{ij} = 1.05$ for $i \neq j$.

The differences in the results are both striking and intuitive. On average a 5% increase in

[36]The exact steps are omitted here to save space.

[37]Due to an increase in $\varphi_{CHN,j}^*$, some unproductive firms that use to export to destination j before the shock will not be able to export after the shock.

Figure 8: Different role of T and τ on export prices



Notes: y-axis is average destination (log) price increase after the shock.

specific trade costs induces an approximately 6.5% increase in export prices as the shock both raises the cost of serving the market and induces firms to upgrade their quality. The increase in firm productivity cutoff magnifies this latter effect so that there appears to be more than 100% pass through. For the case of a shock to ad valorem trade costs, the effect on average is very close to zero because there are competing effects of roughly equal magnitude. On the one hand, higher ad valorem trade costs induce firms to downgrade their quality and so reduce their prices. On the other hand, higher ad valorem trade costs raise the firm productivity cutoff which induces weaker firms to exit and thus increase average prices. These two effect offset each other so the overall effects of ad valorem trade costs on export prices are small.

If firms set constant markups over marginal costs, the ad valorem trade costs would not affect the price, and hence the effect on export prices is only from the changes in specific trade costs. After introducing variable markups, the ad valorem trade costs would affect both productivity cutoff and prices. However, its impact of ad valorem trade costs on prices is still smaller compared to the impact of specific trade costs on prices.

The key point to take away from this comparative static is that when trade costs are mixture of ad valorem and specific as must be so in the real world, the relationship between import prices, export volumes, and the gains from trade becomes complicated. The nature of the shock determines this relationship.

Table 11: Effects of T and τ shocks on distributions of prices, markups, and sales (% change)

	CAN	DEU	FRA	GBR	JPN	USA
Panel A: T shock						
mean(log(prices))	5.86	5.77	5.75	5.80	5.67	5.70
Panel B: τ shock						
mean(log(prices))	-1.00	-1.09	-1.11	-1.06	-1.19	-1.16
panel C: common responses to T and τ shocks						
std(log(prices))	0.02	0.01	0.01	0.02	0.02	0.02
log(prices) 99-50	-0.02	0.01	-0.04	0.03	0.02	0.03
mean(log(markups))	-1.00	-1.09	-1.11	-1.06	-1.19	-1.16
std(log(markups))	2.59	2.85	2.91	2.76	3.12	3.03
log(markups) 99-50	2.81	3.11	3.18	3.00	3.40	3.30
mean(log(sales))	-78.04	-80.06	-80.36	-78.92	-87.62	-85.28
std(log(sales))	70.57	72.18	71.12	71.04	78.74	75.33
log(sales) 99-50	20.61	21.63	21.60	21.00	23.67	22.99
corr(log(prices), log(sales))	-10.50	-21.35	-29.09	-15.61	-11.84	-18.10
corr(log(prices), log(markups))	2.04	3.73	4.86	2.67	3.03	2.89
corr(log(markups), log(sales))	-16.32	-16.21	-15.31	-15.85	-18.02	-16.59

Finally, we examine the effect of different trade costs on distributions of prices, sales, and markups in different destinations in Table 11. We focus on the same set of firms that export to the specific destination before and after the trade cost shock and find the following observations. First, due to the quality mechanism, price levels change differently depending on trade shocks from T or τ, which can be seen from the mean of log prices in panels A and B. Second, only the price levels show differential responses to T and τ shocks. The other variables – including the dispersion moments of prices, markups, and sales, the levels of markups and sales, as well as the correlations between prices, markups, and sales – display identical changes in response to either T shock or τ shock. This is because the two types of trade cost shocks have identical effect on productivity cut-off $\varphi^*_{cj}(\varepsilon)$ by construction. We report those common responses of various distributional moments to T and τ shocks in Panel C.

It is interesting to note that after the trade cost shock, the dispersion of prices alters very little, while the dispersion of sales changes substantially.[38] This is because high- versus low-productivity firms show differential responses to trade cost shocks. To demonstrate the mechanism at work, we illustrate the changes in prices and sales by a low- versus high-productivity firm that exports to destination j using Figure 11 (see Appendix K for details). The analytical result of the illustration in Figure 11 suggests that firms with different initial productivities change their export prices to a similar extent, whereas the associated changes in their sales are profoundly asymmetric across firms, with relatively less productive firms reducing their sales by more. As a result, we observe little changes in the dispersion of log(prices) but larger changes in the dispersion of log(sales) after the trade cost shock.

8 Conclusion

In this paper, we analyzed a model that contains three mechanisms that contribute to price dispersion across firms and countries. These mechanisms include firm heterogeneity in productivity, non-homothetic preferences that give rise to variable markups, and a "Washington Apples" mechanism that features specific trade costs and quality choice by producers. These three mechanisms allow our model to fit well the rich pattern of cross-country and cross-firm price variation observed in the data. Removing any one of these mechanisms made it difficult for a simpler model to fit the key aspects of the joint distribution of firm-level prices and sales.

A nice feature of our model is that incorporates specific trade costs into a quantitative framework in a simple manner. An important implication of adding specific trade costs is that there are now two distinct trade elasticities that arise. Cost shifters that act as ad-valorem trade costs imply a lower elasticity than cost shifters that act as specific-trade costs. In the absence of a way of categorizing trade costs, standard gravity equation analysis is problematic. To overcome this, we showed that the aggregate trade elasticity could still be recovered from variation in markups as in Jung, Simonovska and Weinberger (2019).

We also showed that the relationship between export prices and the gains from trade depends substantially on the nature of trade costs. Specifically, among trade cost shocks with equivalent welfare implications, shocks to specific trade costs generated outsized shifts in export prices while shocks to ad valorem trade costs had little impact on these prices. This means that in the absence of accounting for quality upgrading and for its interaction with pricing-to-market, it is hard to infer the relationship between export prices and the welfare effects of trade shocks.

[38] See, for example, for Canada, under a cost shock of a 5% increase in $T^{\eta-1}$, the changes in the distributional variables are the following: std(log(prices))=0.01, 99-to-50 percentile ratio of log(prices)=-0.02 whereas std(log(sales))=67.83, 99-to-50 percentile ratio of log(sales)=40.45.

Finally, we demonstrated that global welfare results can be derived from variable markup frameworks and that these can be compared across models conditional on a given aggregate trade elasticity. We found that generalized CES systems imply greater gains from trade than non-generalized CES preferences because they feature an "excessive" love of variety that arises from positive utility obtained from simply having access to a variety.

Going forward, we hope that research in the field of international trade will become more cognizant of the importance of modeling trade costs more flexibly. We hope that our framework will encourage more research by demonstrating the potential quantitative importance of specific trade costs and by showing that it is possible to write down relatively simple models that allow for both firm heterogeneity and non-iceberg-type variable trade costs.

References

Amiti, Mary, and Jozef Konings. 2007. "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia." *American Economic Review*, 97(5): 1611–1638.

Antoniades, Alexis. 2015. "Heterogeneous Firms, Quality, and Trade." *Journal of International Economics*, 95(2): 263–273.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review*, 102(1): 94–130.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 2019. "The Elusive Pro-Competitive Effects of Trade." *The Review of Economic Studies*, 86(1): 46–80.

Arkolakis, Costas, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple. 2018. "Innovation and Production in the Global Economy." *American Economic Review*, 108(8): 2128–73.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review*, 98(5): 1998–2031.

Baldwin, Richard, and James Harrigan. 2011. "Zeros, Quality, and Space: Trade Theory and Trade Evidence." *American Economic Journal: Microeconomics*, 3(2): 60–88.

Bas, Maria, and Vanessa Strauss-Kahn. 2015. "Input-trade liberalization, export prices and quality upgrading." *Journal of International Economics*, 95(2): 250–262.

Bertoletti, Paolo, Federico Etro, and Ina Simonovska. 2018. "International Trade with Indirect Additivity." *American Economic Journal: Microeconomics*, 10(2): 1–57.

Cai, Hongbin, and Qiao Liu. 2009. "Competition and Corporate Tax Avoidance: Evidence from Chinese Industrial Firms." *Economic Journal*, 119(537): 764–795.

Caliendo, Lorenzo, and Fernando Parro. 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *Review of Economic Studies*, 82(1): 1–44.

Costinot, Arnaud, and Andrés Rodríguez-Clare. 2014. "Chapter 4 - Trade Theory with Numbers: Quantifying the Consequences of Globalization." In *Handbook of International Economics*. Vol. 4 of *Handbook of International Economics*, , ed. Gita Gopinath, Elhanan Helpman and Kenneth Rogoff, 197 – 261. Elsevier.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2008. "Global Rebalancing with Gravity: Measuring the Burden of Adjustment." *IMF Staff Papers*, 55(3): 511–540.

De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. 2016. "Prices, Markups, and Trade Reform." *Econometrica*, 84: 445–510.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz. 2011. "An Anatomy of International Trade: Evidence From French Firms." *Econometrica*, 79(5): 1453–1498.

Fan, Haichao, Yao Amber Li, and Stephen R. Yeaple. 2015. "Trade Liberalization, Quality, and Export Prices." *Review of Economics and Statistics*, 97(5): 1033–1051.

Fan, Haichao, Yao Amber Li, and Stephen R. Yeaple. 2017. "On the Relationship Between Quality and Productivity: Evidence from China's Accession to the WTO." *Journal of International Economics (forthcoming).*

Fan, Haichao, Yao Amber Li, and Stephen R. Yeaple. 2018. "On the relationship between quality and productivity: Evidence from China's accession to the WTO." *Journal of International Economics*, 110: 28 – 49.

Feenstra, Robert C., and John Romalis. 2014. "International Prices and Endogenous Quality." *The Quarterly Journal of Economics.*

Felbermayr, Gabriel, Benjamin Jung, and Mario Larch. 2013. "Optimal tariffs, retaliation, and the welfare loss from tariff wars in the Melitz model." *Journal of International Economics*, 89(1): 13 – 25.

Harrigan, James, Xiangjun Ma, and Victor Shlychkov. 2015. "Export prices of U.S. firms." *Journal of International Economics*, 97(1): 100 – 111.

Head, Keith, and Thierry Mayer. 2014. "Gravity Equations: Workhorse,Toolkit, and Cookbook." , ed. Gita Gopinath, Elhanan Helpman and Kenneth Rogoff Vol. 4 of *Handbook of International Economics*, Chapter 0, 131–195. Elsevier.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein. 2016. "Quantifying the Sources of Firm Heterogeneity." *The Quarterly Journal of Economics*, 131(3): 1291–1364.

Hummels, David, and Alexandre Skiba. 2004. “Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture.” *Journal of Political Economy*, 112(6): 1384–1402.

Johnson, Robert C. 2012. “Trade and prices with heterogeneous firms.” *Journal of International Economics*, 86(1): 43–56.

Jung, Jae Wook, Ina Simonovska, and Ariel Weinberger. 2019. “Exporter heterogeneity and price discrimination: A quantitative view.” *Journal of International Economics*, 116: 103 – 124.

Khandelwal, Amit. 2010. “The Long and Short (of) Quality Ladders.” *Review of Economic Studies*, 77(4): 1450–1476.

Kugler, Maurice, and Eric Verhoogen. 2009. “Plants and Imported Inputs: New Facts and an Interpretation.” *The American Economic Review*, 99(2): pp. 501–507.

Kugler, Maurice, and Eric Verhoogen. 2012. “Prices, Plant Size, and Product Quality.” *Review of Economic Studies*, 79(1): 307–339.

Levchenko, Andrei A., and Jing Zhang. 2016. “The evolution of comparative advantage: Measurement and welfare implications.” *Journal of Monetary Economics*, 78(C): 96–111.

Manova, Kalina, and Zhiwei Zhang. 2012. “Export Prices Across Firms and Destinations.” *The Quarterly Journal of Economics*, 127: 379–436.

Melitz, Marc J., and Gianmarco I. P. Ottaviano. 2008. “Market Size, Trade, and Productivity.” *Review of Economic Studies*, 75(1): 295–316.

Olley, G. Steven, and Ariel Pakes. 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry.” *Econometrica*, 64(6): pp. 1263–1297.

Ossa, Ralph. 2014. “Trade Wars and Trade Talks with Data.” *American Economic Review*, 104(12): 4104–46.

Schott, Peter K. 2004. “Across-product Versus Within-product Specialization in International Trade.” *The Quarterly Journal of Economics*, 119(2): 646–677.

Simonovska, Ina, and Michael E. Waugh. 2014. “Trade Models, Trade Elasticities, and the Gains from Trade.” National Bureau of Economic Research, Inc NBER Working Papers 20495.

Waugh, Michael E. 2010. “International Trade and Income Differences.” *American Economic Review*, 100(5): 2093–2124.

The Online Appendix for "Quality, Variable Markups, and Welfare: A Quantitative General Equilibrium Analysis of Export Prices"

A Derivation of Demand Function

The utility of a consumer in country j takes the following form:

$$U_j = \left[\sum_i \int_{\omega \in \Omega_{ij}} \left(q_{ij}(\omega) x_{ij}^c(\omega) + \overline{x} \right)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}} \tag{A.1}$$

subject to the following budget constraint:

$$\sum_i \int_{\omega \in \Omega_{ij}} p_{ij}(\omega) x_{ij}^c(\omega) d\omega \leq y_j \tag{A.2}$$

So that the Lagrange function can be written as: $\mathcal{L} = \left[\sum_i \int_{\omega \in \Omega_{ij}} \left(q_{ij}(\omega) x_{ij}^c(\omega) + \overline{x} \right)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}} + \lambda \left(y_j - \sum_i \int_{\omega \in \Omega_{ij}} p_{ij}(\omega) x_{ij}^c(\omega) d\omega \right)$,where λ is the Lagrange multiplier, y_j denotes the consumer's income. Taking the first order condition with respect to $x_{ij}^c(\omega)$ yields:

$$\lambda p_{ij}(\omega) = U_j^{\frac{1}{\sigma}} \left(q_{ij}(\omega) x_{ij}^c(\omega) + \overline{x} \right)^{-\frac{1}{\sigma}} q_{ij}(\omega), \tag{A.3}$$

Following Jung, Simonovska and Weinberger (2019), we define $P_{j\sigma} = \left\{ \sum_i \int_{\omega \in \Omega_{ij}} \tilde{p}_{ij}(\omega)^{1-\sigma} d\omega \right\}^{\frac{1}{1-\sigma}}$, and $P_j = \sum_i \int_{\omega \in \Omega_{ij}} \tilde{p}_{ij}(\omega) d\omega$, where $\tilde{p}_{ij}(\omega) = p_{ij}(\omega) / q_{ij}(\omega)$ is the quality adjusted price. The first order condition (A.3) can be rewritten as:

$$q_{ij}(\omega) x_{ij}^c(\omega) + \overline{x} = U_j \left(\lambda \tilde{p}_{ij}(\omega) \right)^{-\sigma} \tag{A.4}$$

Plugging equation (A.4) into equation (A.1), we have:

$$\lambda = \frac{1}{P_{j\sigma}}$$

Then substituting the above equation into equation (A.4) yield the solution for $x_{ij}^c(\omega)$:

$$q_{ij}(\omega) x_{ij}^c(\omega) = \left[\frac{\tilde{p}_{ij}(\omega)}{P_{j\sigma}} \right]^{-\sigma} U_j - \bar{x}, \tag{A.5}$$

Plugging the previous equation (A.5) into the budget constraint, we have:

$$
\begin{aligned}
y_j &= \sum_i \int_{\omega \in \Omega_{ij}} \tilde{p}_{ij}(\omega) q_{ij}(\omega) x^c_{ij}(\omega) d\omega \\
&= \sum_i \int_{\omega \in \Omega_{ij}} \left[\frac{\tilde{p}_{ij}(\omega)}{P_{j\sigma}}\right]^{-\sigma} U_j \tilde{p}_{ij}(\omega) d\omega - \bar{x} \sum_i \int_{\omega \in \Omega_{ij}} \tilde{p}_{ij}(\omega) d\omega \\
&= U_j P_{j\sigma} - \bar{x} P_j,
\end{aligned}
$$

Hence, we have:

$$
U_j = \frac{y_j + \bar{x} P_j}{P_{j\sigma}} \tag{A.6}
$$

Combing the previous equation (A.6) with equation (A.5) implies:

$$
x_{ij}(\omega) = x^c_{ij}(\omega) L_j = \frac{L_j}{q_{ij}(\omega)} \left[\frac{y_j + \bar{x} P_j}{P^{1-\sigma}_{j\sigma}} \left(\frac{p_{ij}(\omega)}{q_{ij}(\omega)}\right)^{-\sigma} - \bar{x}\right] \tag{A.7}
$$

B Log Utility Function

The representative consumer in country j's demand satisfies:

$$
x_{ij}(\omega) = x^c_{ij}(\omega) L_j = \frac{\bar{x} L_j}{q_{ij}(\omega)} \left[\frac{\psi_j}{\tilde{p}_{ij}(\omega)} - 1\right] \tag{B.1}
$$

where $\tilde{p}_{ijs}(\omega) = \frac{p_{ij}(\omega)}{q_{ij}(\omega)}$ and $\psi_j = \frac{y_j + \bar{x} P_j}{\bar{x} N_j}$. The aggregate prices satisfies $P_j = \sum_i \int_{\omega \in \Omega_{ij}} \tilde{p}_{ij}(\omega) d\omega$. Now, sales and profit for a given variety exported from i to j are as follows,

$$
r_{ij}(\omega) = \bar{x} L_j \tilde{p}_{ij}(\omega) \left[\frac{\psi_j}{\tilde{p}_{ij}(\omega)} - 1\right] \tag{B.2}
$$

$$
\pi_{ij}(\omega) = \bar{x} L_j \left[\tilde{p}_{ij}(\omega) - \tilde{c}_{ij}(\omega)\right] \left[\frac{\psi_j}{\tilde{p}_{ij}(\omega)} - 1\right] \tag{B.3}
$$

where $\tilde{c}_{ij}(\omega) = \frac{c_{ij}(\omega)}{q_{ij}(\omega)}$ is the quality-adjusted marginal cost. Given the quality adjusted marginal cost, firms maximize their profits. This implies that the optimal price of the good satisfies:

$$
\tilde{p}_{ij}(\omega) = \sqrt{\psi_j \tilde{c}_{ij}(\omega)}
$$

We assume that the marginal cost of producing a variety of final good with quality q_{ij} by a firm with productivity φ is given by:

$$
c_{ij}(\varphi, \varepsilon) = \left(T_{ij} w_i + \frac{w_i \tau_{ij}}{\varphi} q^{\eta}_{ij}\right) \varepsilon
$$

where τ_{ij} is ad valorem trade cost and T_{ij} is a specific transportation cost from country i to country j. Maximizing the profit is equivalent to minimizing the quality-adjusted cost $\tilde{c}_{ij}(\omega)$

by the envelop theorem. Choosing the quality to minimize the quality-adjusted marginal cost implies that the optimal level of quality for a firm with productivity φ is:

$$q_{ij}(\varphi, \varepsilon) = \left(\frac{T_{ij}\varphi}{(\eta - 1)\, \tau_{ij}} \right)^{\frac{1}{\eta}} \tag{B.4}$$

and hence the quality adjusted marginal cost of production now is:

$$\tilde{c}_{ij}\left(\varphi, \varepsilon\right) = \left(\frac{\eta}{\eta - 1} T_{ij} w_i \right)^{\frac{\eta-1}{\eta}} \left(\frac{\varphi}{\eta w_i \tau_{ij}} \right)^{-\frac{1}{\eta}} \varepsilon \tag{B.5}$$

At the productivity cutoff $\varphi^*_{ij}\left(\varepsilon\right)$, we have $\tilde{p}^*_{ij}\left(\varphi, \varepsilon\right) = \tilde{c}^*_{ij}\left(\varphi, \varepsilon\right) = \psi_j$, which implies that the productivity cutoff $\varphi^*_{ij}\left(\varepsilon\right)$ takes the following form:

$$\varphi^*_{ij}\left(\varepsilon\right) = \varphi^*_{ij}\varepsilon^{\eta} = \frac{\eta^{\eta}}{(\eta - 1)^{\eta-1}} T^{\eta-1}_{ij} \tau_{ij} w^{\eta}_i \left(\psi_j\right)^{-\eta} \varepsilon^{\eta},$$

In the log utility function, price could be written as:

$$p_{ij}(\varphi, \varepsilon) = \left[\frac{\varphi}{\varphi^*_{ij}\left(\varepsilon\right)} \right]^{\frac{1}{2\eta}} \frac{\eta}{\eta - 1} T_{ij}\varepsilon.$$

Different from the CES utility function, now the markup function could be expressed explicitly as $\left[\frac{\varphi}{\varphi^*_{ij}(\varepsilon)}\right]^{\frac{1}{2\eta}}$.

C Derivation for P_j, $P_{j\sigma}$, X_{ij} and π_i

To derive the aggregate variables, we define $t_{ij} = \tilde{p}_{ij}\left(\omega\right)/p^*_j$. Following the insight of Arkolakis et al. (2019) and Jung, Simonovska and Weinberger (2019), this will make the integration not country specific. From equations (9) and (11), we have:

$$\frac{\tilde{c}_{ij}\left(\varphi, \varepsilon\right)}{\tilde{p}^*_j} = \frac{\tilde{c}_{ij}\left(\varphi, \varepsilon\right)}{\tilde{c}^*_{ij}\left(\varphi, \varepsilon\right)} = \left(\frac{\varphi}{\varphi^*_{ij}\left(\varepsilon\right)} \right)^{-\frac{1}{\eta}} \tag{C.1}$$

Combining the above equation with equation (6) we have:

$$\sigma \left(\frac{\varphi}{\varphi^*_{ij}\left(\varepsilon\right)} \right)^{-\frac{1}{\eta}} = t^{\sigma+1}_{ij} + (\sigma - 1)\, t_{ij} \tag{C.2}$$

which implies that t_{ij} is a monotonically decreasing function of φ. Note that t_{ij} will lies between $(0, 1]$ since $\varphi \in \left[\varphi^*_{ij}\left(\varepsilon\right), \infty\right)$. Totally differentiating both sides gives us:

$$d\varphi = -\eta\sigma^{\eta}\varphi^*_{ij}\left(\varepsilon\right) \frac{(\sigma + 1)\, t^{\sigma}_{ij} + (\sigma - 1)}{\left[t^{\sigma+1}_{ij} + (\sigma - 1)\, t_{ij}\right]^{1+\eta}} dt_{ij} \tag{C.3}$$

First, we derive $P_{j\sigma}$. By definition, we have:

$$\begin{aligned}
P_{j\sigma} &= \left\{\sum_i N_{ij} \int_0^\infty \int_{\varphi_{ij}^*(\varepsilon)}^\infty \tilde{p}_{ij}(\varphi,\varepsilon)^{1-\sigma} \mu_{ij}(\varphi,\varepsilon) f(\varepsilon) d\varphi d\varepsilon\right\}^{\frac{1}{1-\sigma}} \\
&= \tilde{p}_j^* \left\{\sum_i N_{ij} \int_0^\infty \left[\int_{\varphi_{ij}^*(\varepsilon)}^\infty t_{ij}^{1-\sigma} \mu_{ij}(\varphi,\varepsilon) d\varphi\right] f(\varepsilon) d\varepsilon\right\}^{\frac{1}{1-\sigma}} \qquad \text{(C.4)}
\end{aligned}$$

Plugging in the expression of conditional density $\mu_{ij}(\varphi,\varepsilon)$ into equation (C.4) and then we transform the integration variable from φ to t_{ij} by using the relationship between φ and t_{ij}, the inner integration with respect to productivity can be written as:

$$\int_{\varphi_{ij}^*(\varepsilon)}^\infty t_{ij}^{1-\sigma} \mu_{ij}(\varphi,\varepsilon) d\varphi = \frac{\eta\theta}{\sigma^{\eta\theta}} \int_0^1 t_{ij}^{1-\sigma} \left[t_{ij}^{\sigma+1} + (\sigma-1) t_{ij}\right]^{\eta\theta-1} \left[(\sigma+1) t_{ij}^{\sigma} + (\sigma-1)\right] dt_{ij}$$

which is a constant, and we denote it as β_σ. Thus,

$$P_{j\sigma} = \beta_\sigma^{\frac{1}{1-\sigma}} \tilde{p}_j^* N_j^{\frac{1}{1-\sigma}}$$

Second, we derive P_j. By definition, we have

$$\begin{aligned}
P_j &= \sum_i N_{ij} \int_0^\infty \int_{\varphi_{ij}^*(\varepsilon)}^\infty \tilde{p}_{ij}(\varphi,\varepsilon) \mu_{ij}(\varphi,\varepsilon) f(\varepsilon) d\varphi d\varepsilon \\
&= \tilde{p}_j^* \sum_i N_{ij} \int_0^\infty \left[\int_{\varphi_{ij}^*(\varepsilon)}^\infty t_{ij} \mu_{ij}(\varphi,\varepsilon) d\varphi\right] f(\varepsilon) d\varepsilon \\
&= \beta \tilde{p}_j^* N_j
\end{aligned}$$

In the last equality, we use the same variable transformation method as before where β is a constant, defined by:

$$\beta = \frac{\eta\theta}{\sigma^{\eta\theta}} \int_0^1 t_{ij} \left[t_{ij}^{\sigma+1} + (\sigma-1) t_{ij}\right]^{\eta\theta-1} \left[(\sigma+1) t_{ij}^{\sigma} + (\sigma-1)\right] dt_{ij}$$

To derive the equations (C.5) and (C.6), we plug in $\tilde{p}_j^* = \left(\frac{w_j + \bar{x} P_j}{\bar{x} P_{j\sigma}^{1-\sigma}}\right)^{\frac{1}{\sigma}}$ into $P_{j\sigma}$ and P_j, we have:

$$\begin{aligned}
P_{j\sigma} &= \beta_\sigma^{\frac{1}{1-\sigma}} \left(\frac{w_j + \bar{x} P_j}{\bar{x} P_{j\sigma}^{1-\sigma}}\right)^{\frac{1}{\sigma}} N_j^{\frac{1}{1-\sigma}} \\
P_j &= \beta \left(\frac{w_j + \bar{x} P_j}{\bar{x} P_{j\sigma}^{1-\sigma}}\right)^{\frac{1}{\sigma}} N_j,
\end{aligned}$$

which provide us with 2 equations to solve for $P_{j\sigma}$ and P_j. Solving the system yields:

$$\bar{x}P_j = \frac{\beta}{\beta_\sigma - \beta} w_j \tag{C.5}$$

$$\bar{x}P_{j\sigma} = \frac{\beta_\sigma^{\frac{1}{1-\sigma}}}{\beta_\sigma - \beta} N_j^{\frac{\sigma}{1-\sigma}} w_j \tag{C.6}$$

Next, we derive bilateral trade flow X_{ij}, which is given by:

$$\begin{aligned}
X_{ij} &= N_{ij} \int_0^\infty \left[\int_{\varphi_{ij}^*(\varepsilon)}^\infty r_{ij}(\varphi, \varepsilon) \mu_{ij}(\varphi, \varepsilon) d\varphi \right] f(\varepsilon) d\varepsilon \\
&= N_{ij} \left(\bar{x} \tilde{p}_j^* L_j \right) \int_0^\infty \left[\int_{\varphi_{ij}^*(\varepsilon)}^\infty t_{ij} \left(t_{ij}^{-\sigma} - 1 \right) \mu_{ij}(\varphi, \varepsilon) d\varphi \right] f(\varepsilon) d\varepsilon \\
&= (\beta_\sigma - \beta) \bar{x} \tilde{p}_j^* L_j N_{ij} = X_j \frac{N_{ij}}{N_j}
\end{aligned}$$

where $X_j = \sum_i X_{ij}$ is total absorption.

Finally, we derive firm's expected average profit π_i, which satisfies:

$$\begin{aligned}
\pi_i &= \frac{1}{J_i} \sum_j N_{ij} \int_0^\infty \int_{\varphi_{ij}^*(\varepsilon)}^\infty \pi_{ij}(\varphi, \varepsilon) \mu_{ij}(\varphi) f(\varepsilon) d\varphi d\varepsilon \\
&= \frac{1}{J_i} \beta_\pi \sum_j \bar{x} \tilde{p}_j^* L_j N_{ij} = \frac{1}{J_i} \frac{\beta_\pi}{\beta_\sigma - \beta} \sum_j X_{ij} \\
&= \frac{1}{J_i} \frac{\beta_\pi}{\beta_\sigma - \beta} \sum_j \frac{N_{ij}}{N_j} X_j
\end{aligned}$$

where

$$\beta_\pi = \frac{\eta\theta}{\sigma^{\eta\theta}} \int_0^1 \frac{\left(t_{ij}^{\sigma+1} - t_{ij} \right) \left(t_{ij}^{-\sigma} - 1 \right)}{\sigma} \left[t_{ij}^{\sigma+1} + (\sigma - 1) t_{ij} \right]^{\eta\theta - 1} \left[(\sigma + 1) t_{ij}^\sigma + (\sigma - 1) \right] dt_{ij}$$

D Proof of Propositions

D.1 Proof of Proposition 1

The percentage change of U_j satisfies:

$$d \ln U_j = \frac{\sigma}{\sigma - 1} \left(d \ln w_j - d \ln \tilde{p}_j^* \right) \tag{D.1}$$

Based on equations (11), (13) and (21), we can rewrite N_{ij} as:

$$N_{ij} = \frac{\kappa \beta_\pi}{f \beta_X} b_i L_i \left[\frac{\eta^\eta}{(\eta - 1)^{\eta - 1}} T_{ij}^{\eta - 1} \tau_{ij} w_i^\eta \left(\tilde{p}_j^* \right)^{-\eta} \right]^{-\theta} \tag{D.2}$$

where $\beta_X = \beta_\sigma - \beta$ is a constant. This implies that

$$\lambda_{jj} = \frac{X_{jj}}{\sum_i X_{ij}} = \frac{N_{jj}}{\sum_i N_{ij}} = \frac{b_j L_j \left(T_{jj}^{\eta-1} \tau_{jj} w_j^{\eta}\right)^{-\theta}}{\sum_i b_i L_i \left(T_{ij}^{\eta-1} \tau_{ij} w_i^{\eta}\right)^{-\theta}} \tag{D.3}$$

Consider the foreign shocks: $(b_i, L_i, T_{ij}, \tau_{ij})$ is changed to $(b_i', L_i', T_{ij}', \tau_{ij}')$ for $i \neq j$ such that $b_j = b_j', L_j = L_j', T_{jj} = T_{jj}', \tau_{jj} = \tau_{jj}'$. Totally differentiating the previous equation implies:

$$d\ln\lambda_{jj} = \sum_i \lambda_{ij} \left[\theta\eta \left(d\ln w_i - d\ln w_j\right) - d\ln \xi_{ij}\right] \tag{D.4}$$

where $d\ln\xi_{ij}$ reflects any foreign shock, which satisfies:

$$d\ln\xi_{ij} = -\theta\left(\eta - 1\right) d\ln T_{ij} - \theta d\ln\tau_{ij} + d\ln b_i + d\ln L_i$$

The expression of $\tilde{p}_j^*$, together with equation (C.5) and (C.6), imply that:

$$d\ln\tilde{p}_j^* = \frac{1}{\sigma} d\ln w_j + \frac{\sigma - 1}{\sigma} d\ln P_{j\sigma} = d\ln w_j - \sum_i \lambda_{ij} d\ln N_{ij} \tag{D.5}$$

Totally differentiating the expression of N_{ij} and substituting the percentage change of N_{ij} into the previous equation, we have:

$$\begin{aligned} d\ln\tilde{p}_j^* &= d\ln w_j - \sum_i \lambda_{ij} d\ln N_{ij} \\ &= d\ln w_j + \sum_i \lambda_{ij} \left[\theta\eta \left(d\ln w_i - d\ln\tilde{p}_j^*\right) - d\ln\xi_{ij}\right] \\ &= \frac{1}{1+\eta\theta} d\ln w_j + \frac{1}{1+\eta\theta} \sum_i \lambda_{ij} \left[\theta\eta d\ln w_i - d\ln\xi_{ij}\right] \end{aligned} \tag{D.6}$$

Hence, the percentage change in welfare satisfies:

$$\begin{aligned} d\ln U_j &= \frac{\sigma}{\sigma-1} \left(d\ln w_j - d\ln\tilde{p}_j^*\right) \\ &= -\frac{\sigma}{\sigma-1} \frac{1}{1+\eta\theta} \sum_i \lambda_{ij} \left[\theta\eta \left(d\ln w_i - d\ln w_j\right) - d\ln\xi_{ij}\right] \\ &= -\frac{\sigma}{\sigma-1} \frac{1}{1+\eta\theta} d\ln\lambda_{jj} \end{aligned} \tag{D.7}$$

Integrating the previous expression between the initial equilibrium (before the shock) and the new equilibrium (after the shock), we finally get

$$\widehat{U}_j = \left(\widehat{\lambda}_{jj}\right)^{-\frac{\sigma}{\sigma-1}\frac{1}{1+\eta\theta}} \tag{D.8}$$

It shows that the changes in welfare at country j can be inferred from changes in the share of domestic expenditure, λ_{jj}, using the parameter, $-\frac{\sigma}{\sigma-1}\frac{1}{1+\eta\theta}$.

D.2 Proof of Proposition 2

We consider an arbitrary change in trade costs from τ_{ij} to τ'_{ij} and T_{ij} to T'_{ij}. The share of expenditure on domestic goods in the initial and new equilibrium, respectively, are given by:

$$\lambda_{jj} = \frac{X_{jj}}{\sum_i X_{ij}} = \frac{b_j L_j \left(T_{jj}^{\eta-1} \tau_{jj} w_j^{\eta}\right)^{-\theta}}{\sum_i b_i L_i \left(T_{ij}^{\eta-1} \tau_{ij} w_i^{\eta}\right)^{-\theta}} \tag{D.9}$$

$$\lambda'_{jj} = \frac{b_j L_j \left(T_{jj}^{\eta-1} \tau_{jj} \left(w'_j\right)^{\eta}\right)^{-\theta}}{\sum_i b_i L_i \left(\left(T'_{ij}\right)^{\eta-1} \tau'_{ij} \left(w'_i\right)^{\eta}\right)^{-\theta}} \tag{D.10}$$

Combing the previous two equations, we obtain:

$$\widehat{\lambda}_{jj} = \frac{\left(\widehat{w}_j\right)^{-\eta\theta}}{\sum_i \lambda_{ij} \left[\left(\widehat{T}_{ij}\right)^{\eta-1} \widehat{\tau}_{ij}\right]^{-\theta} \left(\widehat{w}_i\right)^{-\eta\theta}} \tag{D.11}$$

Labor market clearing condition implies that:

$$w_i L_i = \sum_j \lambda_{ij} w_j L_j = \sum_j \frac{b_i L_i \left[T_{ij}^{\eta-1} \tau_{ij}\right]^{-\theta} w_i^{-\eta\theta}}{\sum_{i'} b_{i'} L_{i'} \left[T_{i'j}^{\eta-1} \tau_{i'j}\right]^{-\theta} w_{i'}^{-\eta\theta}} w_j L_j \tag{D.12}$$

After τ_{ij} becomes τ'_{ij} and T_{ij} becomes T'_{ij}, the previous equation becomes:

$$w'_i L_i = \sum_j \frac{b_i L_i \left[\left(T'_{ij}\right)^{\eta-1} \tau'_{ij}\right]^{-\theta} \left(w'_i\right)^{-\eta\theta}}{\sum_{i'} b_{i'} L_{i'} \left[\left(T'_{i'j}\right)^{\eta-1} \tau'_{i'j}\right]^{-\theta} \left(w'_{i'}\right)^{-\eta\theta}} w'_j L_j$$

We can rearrange the previous expression as:

$$\widehat{w}_i w_i L_i = \sum_j \frac{\lambda_{ij} \left[\widehat{T}_{ij}^{\eta-1} \widehat{\tau}_{ij}\right]^{-\theta} \left(\widehat{w}_i\right)^{-\eta\theta}}{\sum_{i'} \lambda_{i'j} \left[\widehat{T}_{i'j}^{\eta-1} \widehat{\tau}_{i'j}\right]^{-\theta} \left(\widehat{w}_{i'}\right)^{-\eta\theta}} \widehat{w}_j w_j L_j$$

which implies the equation (27).

E Global Measure of Welfare Gains

E.1 Derivation of Equation (25) in Proposition 1

The welfare measure can be written as follows:

$$U_j = \left[\sum_i \int_{\omega \in \Omega_{ij}} \left(q_{ij}(\omega) x_{ij}^c(\omega) + \bar{x}\right)^{\frac{\sigma-1}{\sigma}} d\omega\right]^{\frac{\sigma}{\sigma-1}} = \frac{w_j + \bar{x} P_j}{P_{j\sigma}} \tag{E.1}$$

which together with the expression of $\bar{x}P_j = \frac{\beta}{\beta_\sigma - \beta} w_j$ and $\bar{x}P_{j\sigma} = \frac{\beta_\sigma^{\frac{1}{1-\sigma}}}{\beta_\sigma - \beta} N_j^{\frac{\sigma}{1-\sigma}} w_j$, implies that

$$U_j = \bar{x}\beta_\sigma^{\frac{\sigma}{\sigma-1}} N_j^{\frac{\sigma}{\sigma-1}}, \tag{E.2}$$

By definition, $N_j = \sum_i N_{ij}$, we thus have the following relationship

$$\hat{N}_j = \sum_i \lambda_{ij}\hat{N}_{ij}, \tag{E.3}$$

and combining the equation (E.2), we have

$$\hat{U}_j = \left(\sum_i \lambda_{ij}\hat{N}_{ij}\right)^{\frac{\sigma}{\sigma-1}}, \tag{E.4}$$

The equation (17) implies that $\lambda_{jj} = \frac{N_{jj}}{N_j} = \frac{N_{jj}}{\sum_i N_{ij}}$, so

$$\hat{N}_j = \sum_i \lambda_{ij}\hat{N}_{ij} = \frac{\hat{N}_{jj}}{\hat{\lambda}_{jj}}, \tag{E.5}$$

substituting into the last $\hat{U}_j$ equation, we have

$$\hat{U}_j = \left(\frac{\hat{\lambda}_{jj}}{\hat{N}_{jj}}\right)^{-\frac{\sigma}{\sigma-1}}, \tag{E.6}$$

We thus have

$$\hat{N}_{jj} = \left(\hat{\varphi}^*_{jj}\right)^{-\theta} = \left(\frac{\hat{w}_j}{\hat{\tilde{p}}^*_j}\right)^{-\theta\eta} = \left(\hat{N}_j\right)^{-\theta\eta} = \left(\frac{\hat{N}_{jj}}{\hat{\lambda}_{jj}}\right)^{-\theta\eta} = \left(\hat{\lambda}_{jj}\right)^{\frac{\theta\eta}{1+\theta\eta}} \tag{E.7}$$

where the first equality stems from the equation (13), the second equality stems from the equation (12), the third equality stems from the equation (17), the fourth equality stems from the equation (E.5). The previous equation (E.6), together with the equation (E.7), implies that:

$$\hat{U}_j = \left(\frac{\hat{\lambda}_{jj}}{\hat{N}_{jj}}\right)^{-\frac{\sigma}{\sigma-1}} = \left(\frac{\hat{\lambda}_{jj}}{\left(\hat{\lambda}_{jj}\right)^{\frac{\theta\eta}{1+\theta\eta}}}\right)^{-\frac{\sigma}{\sigma-1}} = \left(\hat{\lambda}_{jj}\right)^{-\frac{\sigma}{\sigma-1}\frac{1}{1+\theta\eta}}$$

E.2 Equivalent Variation as Global Measure of Welfare

Formally, the exact welfare change in country j is computed as $e\left(\mathbf{p}_j, U'_j\right)/w_j - 1$, where $\mathbf{p}_j$ and w_j are the set of good prices and the wage in the initial equilibrium, respectively, and U'_j is the utility level in the counterfactual equilibrium. The expenditure function in country j takes the following form:

$$e_j = \sum_i \int_{\omega \in \Omega_{ij}} p_{ij}(\omega) x_{ij}^c(\omega) d\omega \tag{E.8}$$

subject to the following budget constraint:

$$\left[\sum_i \int_{\omega \in \Omega_{ij}} \left(q_{ij}(\omega) x_{ij}^c(\omega) + \overline{x} \right)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}} \geq U_j \tag{E.9}$$

Taking the first order condition with respect to $x_{ij}^c(\omega)$ yields:

$$p_{ij}(\omega) = \lambda U_j^{\frac{1}{\sigma}} \left(q_{ij}(\omega) x_{ij}^c(\omega) + \overline{x} \right)^{-\frac{1}{\sigma}} q_{ij}(\omega), \tag{E.10}$$

where λ is the Lagrange multiplier. The previous equation can be rewritten as:

$$q_{ij}(\omega) x_{ij}^c(\omega) + \overline{x} = U_j \left(\tilde{p}_{ij}(\omega) / \lambda \right)^{-\sigma} \tag{E.11}$$

where $\tilde{p}_{ij}(\omega) = p_{ij}(\omega) / q_{ij}(\omega)$ is the quality adjusted price. Plugging equation (E.11) into equation (E.9), we have:

$$\lambda = P_{j\sigma} = \left[\sum_i \int_{\omega \in \Omega_{ij}} \left(\tilde{p}_{ij}(\omega) \right)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}}$$

Then substituting the above equation into equation (E.11) yields the solution for $x_{ij}^c(\omega)$:

$$q_{ij}(\omega) x_{ij}^c(\omega) = \left[\frac{\tilde{p}_{ij}(\omega)}{P_{j\sigma}} \right]^{-\sigma} U_j - \overline{x}, \tag{E.12}$$

Plugging the previous equation (E.12) into the object function, we have:

$$\begin{aligned} e_j &= \sum_i \int_{\omega \in \Omega_{ij}} \tilde{p}_{ij}(\omega) q_{ij}(\omega) x_{ij}^c(\omega) d\omega \\ &= \sum_i \int_{\omega \in \Omega_{ij}} \left[\frac{\tilde{p}_{ij}(\omega)}{P_{j\sigma}} \right]^{-\sigma} U_j \tilde{p}_{ij}(\omega) d\omega - \bar{x} \sum_i \int_{\omega \in \Omega_{ij}} \tilde{p}_{ij}(\omega) d\omega \\ &= P_{j\sigma} U_j - \bar{x} P_j, \end{aligned}$$

Hence, the exact welfare change in country j is computed as

$$\begin{aligned} e\left(\mathbf{p}_j, U_j'\right) / w_j - 1 &= \frac{P_{j\sigma} U_j' - \bar{x} P_j - \left(P_{j\sigma} U_j - \bar{x} P_j \right)}{P_{j\sigma} U_j - \bar{x} P_j} \\ &= \frac{P_{j\sigma} U_j}{P_{j\sigma} U_j - \bar{x} P_j} \frac{U_j' - U_j}{U_j} \end{aligned}$$

where $P_{j\sigma} U_j = \frac{\beta_\sigma}{\beta_\sigma - \beta} w_j$ and $\bar{x} P_j = \frac{\beta}{\beta_\sigma - \beta} w_j$ in equilibrium. Hence, the exact welfare change in country j satisfies

$$e\left(\mathbf{p}_j, U_j'\right)/w_j - 1 = \frac{\beta_\sigma}{\beta_\sigma - \beta}\frac{U_j' - U_j}{U_j} = \frac{\beta_\sigma}{\beta_\sigma - \beta}\widehat{U}_j$$

F Multi Sector Extension

F.1 Derivation of Multi Sector Model

Household utility in country j can be written as:

$$U_j = \prod_s C_{js}^{\alpha_s}, \tag{F.1}$$

with

$$C_{js} = \left[\sum_i \int_{\omega\in\Omega_{ijs}} \left(q_{ijs}(\omega)x_{ijs}^c(\omega) + \overline{x}_s\right)^{\frac{\sigma_s-1}{\sigma_s}} d\omega\right]^{\frac{\sigma_s}{\sigma_s-1}}, \tag{F.2}$$

The representative consumer in country j's demand satisfies:

$$x_{ijs}^c(\omega) = \frac{\overline{x}_s}{q_{ijs}(\omega)}\left\{\left[\frac{\tilde{p}_{ijs}(\omega)}{\tilde{p}_{js}^*}\right]^{-\sigma_s} - 1\right\} \tag{F.3}$$

where $\tilde{p}_{ijs}(\omega) = \frac{p_{ijs}(\omega)}{q_{ijs}(\omega)}$ and $\tilde{p}_{js}^* = \left[\frac{\alpha_s\left(\sum_s \bar{x}_s P_{js} + y_j\right)}{\bar{x}_s P_{j\sigma s}^{1-\sigma_s}}\right]^{\frac{1}{\sigma_s}}$. The aggregate prices satisfy $P_{js} = \sum_i \int_{\omega\in\Omega_{ijs}} \tilde{p}_{ijs}(\omega)\, d\omega$ and $P_{j\sigma s} = \left\{\sum_i \int_{\omega\in\Omega_{ijs}} \tilde{p}_{ijs}(\omega)^{1-\sigma} d\omega\right\}^{\frac{1}{1-\sigma}}$. Now, quantity, sales, and profit for a given variety exported from i to j in sector s are as follows,

$$x_{ijs}(\omega) = \frac{\bar{x}_s L_j}{q_{ijs}(\omega)}\left[\left(\frac{\tilde{p}_{ijs}(\omega)}{p_{js}^*}\right)^{-\sigma_s} - 1\right] \tag{F.4}$$

$$r_{ijs}(\omega) = \bar{x}_s L_j \tilde{p}_{ijs}(\omega)\left[\left(\frac{\tilde{p}_{ijs}(\omega)}{p_{js}^*}\right)^{-\sigma_s} - 1\right] \tag{F.5}$$

$$\pi_{ijs}(\omega) = \bar{x}_s L_j\left[\tilde{p}_{ijs}(\omega) - \tilde{c}_{ijs}(\omega)\right]\left[\left(\frac{\tilde{p}_{ijs}(\omega)}{p_{js}^*}\right)^{-\sigma_s} - 1\right] \tag{F.6}$$

where $\tilde{c}_{ijs}(\omega) = \frac{c_{ijs}(\omega)}{q_{ijs}(\omega)}$ is the quality-adjusted marginal cost. Given the quality adjusted marginal cost, firms maximize their profits. This implies that the optimal price of the good satisfies:

$$\sigma\frac{\tilde{c}_{ijs}(\omega)}{p_{js}^*} = \left(\frac{\tilde{p}_{ijs}(\omega)}{p_{js}^*}\right)^{\sigma+1} + (\sigma - 1)\frac{\tilde{p}_{ijs}(\omega)}{p_{js}^*} \tag{F.7}$$

We assume that the marginal cost of producing a variety of final good with quality q_{ijs} by a firm with productivity φ is given by:

$$c_{ijs}(\varphi, \varepsilon) = \left(T_{ijs}w_i + \frac{w_i\tau_{ijs}q_{ijs}^{\eta_s}}{\varphi}\right)\varepsilon$$

where τ_{ijs} is ad valorem trade cost and T_{ijs} is a specific transportation cost from country i to country j in sector s. Productivity φ follows the Pareto distribution with c.d.f. $G_i(\varphi) = 1 - b_{is}\varphi^{-\theta_s}$, and ε follows the log-normally distribution with the variance σ_s in sector s. Maximizing the profit is equivalent to minimizing the quality-adjusted cost $\tilde{c}_{ijs}(\omega)$ by the envelop theorem. Choosing the quality to minimize the quality-adjusted marginal cost implies that the optimal level of quality for a firm with productivity φ is:

$$q_{ijs}(\varphi, \varepsilon) = \left(\frac{T_{ijs}\varphi}{(\eta_s - 1)\,\tau_{ijs}} \right)^{\frac{1}{\eta_s}} \tag{F.8}$$

and hence the quality adjusted marginal cost of production now is:

$$\tilde{c}_{ijs}(\varphi, \varepsilon) = \left(\frac{\eta_s}{\eta_s - 1} T_{ijs} w_i \right)^{\frac{\eta_s - 1}{\eta_s}} \left(\frac{\varphi}{\eta_s w_i \tau_{ijs}} \right)^{-\frac{1}{\eta_s}} \varepsilon \tag{F.9}$$

At the productivity cutoff $\varphi^*_{ijs}(\varepsilon)$, we have $p^*_{ijs}(\varphi, \varepsilon) = c^*_{ijs}(\varphi, \varepsilon) = p^*_{js}$, which implies that the productivity cutoff $\varphi^*_{ijs}(\varepsilon)$ takes the following form:

$$\varphi^*_{ijs}(\varepsilon) = \varphi^*_{ijs}\varepsilon^{\eta_s} = \frac{\eta_s^{\eta_s}}{(\eta_s - 1)^{\eta_s - 1}} T_{ijs}^{\eta_s - 1} \tau_{ijs} w_i^{\eta_s} \left(\tilde{p}^*_{js}\right)^{-\eta_s} \varepsilon^{\eta_s},$$

Based on the similar derivation in the one-sector model in Section 3, we know that the exporting firm mass N_{ijs}, the aggregate price P_{js} and $P_{j\sigma s}$, the trade flow X_{ijs}, the expected average profit π_{is} and the potential firm mass J_{is} in sector s satisfy:

$$N_{ijs} = \kappa_s J_{is} b_{is} \left(\varphi^*_{ijs}\right)^{-\theta_s} \tag{F.10}$$

$$\bar{x}_s P_{js} = \beta_s \tilde{p}^*_{js} N_{js} \tag{F.11}$$

$$\bar{x}_s P_{j\sigma s} = \beta_{\sigma s}^{\frac{1}{1-\sigma_s}} \tilde{p}^*_{js} N_{js}^{\frac{1}{1-\sigma_s}} \tag{F.12}$$

$$X_{ijs} = \beta_{Xs} \bar{x}_s \tilde{p}^*_{js} N_{ijs} L_j \tag{F.13}$$

$$\pi_{is} = \beta_{\pi s} \sum_j \bar{x}_s \kappa_s b_{is} \left(\varphi^*_{ijs}\right)^{-\theta_s} \tilde{p}^*_{js} L_j \tag{F.14}$$

$$J_{is} = \frac{\beta_{\pi s}}{\beta_{Xs}} \frac{\alpha_s L_i}{f_s} \tag{F.15}$$

where κ_s, β_s, $\beta_{\sigma s}$, $\beta_{\pi s}$ and β_{Xs} are constant. Now, the expression of choke price $\tilde{p}^*_{js}$, together with the equation (F.11) and (F.12), implies[39]

$$\bar{x}_s P_{js} = \gamma_s w_j \tag{F.16}$$

$$\bar{x}_s P_{j\sigma s} = \frac{\gamma_s}{\beta_s} \beta_{\sigma s}^{\frac{1}{1-\sigma_s}} N_{js}^{\frac{\sigma_s}{1-\sigma_s}} w_j \tag{F.17}$$

$$\tilde{p}_{js}^* = \frac{\gamma_s}{\beta_s} \frac{w_j}{N_{js}} \tag{F.18}$$

where γ_s are determined by $\beta_s \alpha_s \left(\sum_s \gamma_s + 1\right) = \beta_{\sigma s} \overline{x}_s^{\sigma_s} \gamma_s$.

F.2 Proof of Proposition 3

The percentage change of U_j satisfies:

$$d \ln U_j = \sum_s \frac{\alpha_s \sigma_s}{\sigma_s - 1} \left(d \ln w_j - d \ln \tilde{p}_{js}^*\right) \tag{F.19}$$

Based on equations (11), (13) and (21), we can rewrite N_{ij} as:

$$N_{ijs} = \frac{\kappa \beta_{\pi s}}{\beta_{Xs} f_s} \alpha_s b_{is} L_i \left(\frac{\eta_s^{\eta_s}}{(\eta_s - 1)^{\eta_s - 1}} T_{ijs}^{\eta_s - 1} \tau_{ijs} w_i^{\eta_s} \left(\tilde{p}_{js}^*\right)^{-\eta_s} \right)^{-\theta_s} \tag{F.20}$$

which implies that

$$\lambda_{jjs} = \frac{X_{jjs}}{\sum_i X_{ijs}} = \frac{N_{jjs}}{\sum_i N_{ijs}} = \frac{b_{js} L_j \left(T_{jjs}^{\eta-1} \tau_{jjs} w_j^{\eta}\right)^{-\theta}}{\sum_i b_{is} L_i \left(T_{ijs}^{\eta-1} \tau_{ijs} w_i^{\eta}\right)^{-\theta}} \tag{F.21}$$

Consider the foreign shocks: $(b_{is}, L_i, T_{ijs}, \tau_{ijs})$ is changed to $(b'_{is}, L'_i, T'_{ijs}, \tau'_{ijs})$ for $i \neq j$ such that $b_{js} = b'_{js}, L_j = L'_j, T_{jjs} = T'_{jjs}, \tau_{jjs} = \tau'_{jjs}$. Totally differentiating the previous equation implies:

$$d \ln \lambda_{jjs} = \sum_i \lambda_{ijs} \left[\theta \eta \left(d \ln w_i - d \ln w_j\right) - d \ln \xi_{ijs}\right] \tag{F.22}$$

[39] We can get them by first conjecturing $\overline{x}_s P_{js} = \gamma_s w_j$, where γ_s is sector level constant. Then $\sum_s \bar{x}_s P_{js} = \left(\sum_s \gamma_s\right) w_j$, which implies the price cut-off $\tilde{p}_{js}^*$ can be written as:

$$\left(\tilde{p}_{js}^*\right)^{\sigma_s} = \frac{\alpha_s \left(\sum_s \gamma_s + 1\right) w_j}{\overline{x}_s P_{j\sigma s}^{1-\sigma_s}} = \frac{\beta_s^{1-\sigma_s} \alpha_s \left(\sum_s \gamma_s + 1\right)}{\beta_{\sigma s} \overline{x}_s^{\sigma_s} \gamma_s^{1-\sigma_s}} \left(\frac{w_j}{N_{js}}\right)^{\sigma_s}$$

Hence, we have

$$\bar{x}_s P_{js} = \beta_s \left(\sigma_s, \theta_s, \eta_s\right) \tilde{p}_{js}^* N_{js} = \left[\frac{\beta_s \alpha_s \left(\sum_s \gamma_s + 1\right)}{\beta_{\sigma s} \overline{x}_s^{\sigma_s} \gamma_s^{1-\sigma_s}}\right]^{\frac{1}{\sigma_s}} w_j = \gamma_s w_j$$

Hence, γ_s is determined by

$$\beta_s \alpha_s \left(\sum_s \gamma_s + 1\right) = \beta_{\sigma s} \overline{x}_s^{\sigma_s} \gamma_s$$

Hence, we have equations (F.16), (F.17) and (F.18).

where $d\ln\xi_{ijs}$ reflects any foreign shock, which satisfies:

$$d\ln\xi_{ijs} = -\theta_s\left(\eta_s - 1\right)d\ln T_{ijs} - \theta_s d\ln\tau_{ijs} + d\ln b_{is} + d\ln L_i$$

The expression of $\tilde{p}_j^*$, together with equation (C.5) and (C.6), imply that:

$$d\ln\tilde{p}_{js}^* = \frac{1}{\sigma_s}d\ln w_j + \frac{\sigma_s - 1}{\sigma_s}d\ln P_{j\sigma s} = d\ln w_j - \sum_i \lambda_{ijs}d\ln N_{ijs} \tag{F.23}$$

Totally differentiating the expression of N_{ij} and substituting the percentage change of N_{ij} into the previous equation, we have:

$$\begin{aligned} d\ln\tilde{p}_{js}^* &= d\ln w_j - \sum_i \lambda_{ijs}d\ln N_{ijs} \\ &= d\ln w_j + \sum_i \lambda_{ijs}\left[\eta_s\theta_s\left(d\ln w_i - d\ln\tilde{p}_{js}^*\right) - d\ln\xi_{ijs}\right] \\ &= \frac{1}{1+\eta_s\theta_s}d\ln w_j + \frac{1}{1+\eta_s\theta_s}\sum_i \lambda_{ijs}\left[\eta_s\theta_s d\ln w_i - d\ln\xi_{ijs}\right] \end{aligned} \tag{F.24}$$

Hence, the percentage change in welfare satisfies:

$$\begin{aligned} d\ln U_j &= \sum_s \frac{\alpha_s\sigma_s}{\sigma_s - 1}\left(d\ln w_j - d\ln\tilde{p}_{js}^*\right) \\ &= -\sum_s \frac{\alpha_s\sigma_s}{\sigma_s - 1}\frac{1}{1+\eta_s\theta_s}\sum_i \lambda_{ijs}\left[\eta_s\theta_s\left(d\ln w_i - d\ln w_j\right) - d\ln\xi_{ijs}\right] \\ &= -\sum_s \frac{\alpha_s\sigma_s}{\sigma_s - 1}\frac{1}{1+\eta_s\theta_s}d\ln\lambda_{jjs} \end{aligned} \tag{F.25}$$

Integrating the previous expression between the initial equilibrium (before the shock) and the new equilibrium (after the shock), we finally get

$$\widehat{U}_j = \prod_s \left(\widehat{\lambda}_{jjs}\right)^{-\frac{\alpha_s\sigma_s}{\sigma_s-1}\frac{1}{1+\eta_s\theta_s}} \tag{F.26}$$

It shows that the changes in welfare at country j can be inferred from changes in the share of domestic expenditure, λ_{jjs}, using the parameter, $\frac{\alpha_s\sigma_s}{\sigma_s-1}\frac{1}{1+\eta_s\theta_s}$.

G Fixed Quality Case without T_{ij}

We prove the welfare implication of our model without q_{ij} and T_{ij}. From the demand system, we have the representative consumer in country j's demand given by:

$$x_{ij}(\omega) = L_j\left[\frac{y_j + \bar{x}P_j}{P_{j\sigma}^{1-\sigma}}p_{ij}\left(\omega\right)^{-\sigma} - \bar{x}\right] \tag{G.1}$$

where $P_j = \sum_i \int_{\omega\in\Omega_{ij}} p_{ij}(\omega)d\omega$ and $P_{j\sigma} = \left\{\sum_i \int_{\omega\in\Omega_{ij}} p_{ij}(\omega)^{1-\sigma}d\omega\right\}^{\frac{1}{1-\sigma}}$. Now, quantity, sales, and profit for a given variety exported from i to j are as follows,

$$x_{ij}(\omega) = \bar{x}L_j\left[\left(\frac{p_{ij}(\omega)}{p_j^*}\right)^{-\sigma} - 1\right] \tag{G.2}$$

$$r_{ij}(\omega) = \bar{x}L_j p_{ij}(\omega)\left[\left(\frac{p_{ij}(\omega)}{p_j^*}\right)^{-\sigma} - 1\right] \tag{G.3}$$

$$\pi_{ij}(\omega) = \bar{x}L_j\left[p_{ij}(\omega) - c_{ij}(\omega)\right]\left[\left(\frac{p_{ij}(\omega)}{p_j^*}\right)^{-\sigma} - 1\right] \tag{G.4}$$

where $p_j^* = \left(\frac{y_j+\bar{x}P_j}{\bar{x}P_{j\sigma}^{1-\sigma}}\right)^{\frac{1}{\sigma}}$ is the choke price. Given the quality adjusted marginal cost, firms maximize their profits. This implies that the optimal price of the good satisfies:

$$\sigma\frac{c_{ij}(\omega)}{p_j^*} = \left(\frac{p_{ij}(\omega)}{p_j^*}\right)^{\sigma+1} + (\sigma-1)\frac{p_{ij}(\omega)}{p_j^*} \tag{G.5}$$

For the production, we assume that the marginal cost of production is

$$c_{ij} = \frac{w_i\tau_{ij}}{\varphi}\varepsilon$$

where φ follows the Pareto distribution with c.d.f. $G_i(\varphi) = 1 - b_i\varphi^{-\theta}$ and ε is drawn from a log normal distribution. At the productivity cutoff φ_{ij}^* to sell goods from country i to country j, we have $p_{ij}^*(\varphi) = c_{ij}^*(\varphi) = p_j^*$, which implies:

$$\varphi_{ij}^* = \frac{w_i\tau_{ij}}{p_j^*}\varepsilon \tag{G.6}$$

Based on the similar derivation in Section 3, we know that the exporting firm mass N_{ij}, the aggregate price P_j and $P_{j\sigma}$, the trade flow X_{ij}, the expected average profit π_i and the potential firm mass J_i satisfy:

$$N_{ij} = \kappa' J_i b_i\left(\varphi_{ij}^*\right)^{-\theta} \tag{G.7}$$

$$\bar{x}P_j = \beta' p_j^* N_j \tag{G.8}$$

$$\bar{x}P_{j\sigma} = \beta'_\sigma p_j^* N_j^{\frac{1}{1-\sigma}} \tag{G.9}$$

$$X_{ij} = \beta'_X \bar{x} p_j^* N_{ij} L_j \tag{G.10}$$

$$\pi_i = \beta'_\pi \sum_j \bar{x}\kappa' b_i\left(\varphi_{ij}^*\right)^{-\theta} p_j^* L_j \tag{G.11}$$

$$J_i = \frac{\beta'_\pi}{\beta'_X}\frac{L_i}{f} \tag{G.12}$$

where κ', β', β'_σ, β'_X and β'_π are constant. The expression of choke price p_j^*, together with the equation (G.8) and (G.9), implies

$$\bar{x}P_j = \frac{\beta'}{\beta'_\sigma - \beta'} w_j \tag{G.13}$$

$$\bar{x}P_{j\sigma} = \frac{(\beta'_\sigma)^{\frac{1}{1-\sigma}}}{\beta'_\sigma - \beta'} N_j^{\frac{\sigma}{1-\sigma}} w_j \tag{G.14}$$

$$p_j^* = \frac{1}{\bar{x}\left(\beta'_\sigma - \beta'\right)} \frac{w_j}{N_j} \tag{G.15}$$

Now, the welfare still satisfy:

$$U_j = \beta_u \left(\frac{w_j}{p_j^*}\right)^{\frac{\sigma}{\sigma-1}}$$

where $\beta_u = \bar{x}^{\frac{1}{1-\sigma}} \left(\frac{\beta'_\sigma}{\beta'_\sigma - \beta'}\right)^{\frac{\sigma}{\sigma-1}}$ is a constant. The percentage change of U_j satisfies:

$$d\ln U_j = \frac{\sigma}{\sigma-1}\left(d\ln w_j - d\ln p_j^*\right) \tag{G.16}$$

Now, λ_{jj} satisfies:

$$\lambda_{jj} = \frac{N_{jj}}{\sum_i N_{ij}} = \frac{b_j L_j \left(\tau_{jj} w_j\right)^{-\theta}}{\sum_i b_i L_i \left(\tau_{ij} w_i\right)^{-\theta}} \tag{G.17}$$

Consider the foreign shocks: (b_i, L_i, τ_{ij}) is changed to (b'_i, L'_i, τ'_{ij}) for $i \neq j$ such that $b_j = b'_j, L_j = L'_j, T_{jj} = T'_{jj}, \tau_{jj} = \tau'_{jj}$. Totally differentiating the previous equation implies:

$$d\ln \lambda_{jj} = \sum_i \lambda_{ij}\left[\theta\left(d\ln w_i - d\ln w_j\right) - d\ln \xi_{ij}\right] \tag{G.18}$$

where $d\ln \xi_{ij}$ reflects any foreign shock, which satisfies:

$$d\ln \xi_{ij} = -\theta d\ln \tau_{ij} + d\ln b_i + d\ln L_i$$

The expression of p_j^* imply that:

$$d\ln p_j^* = d\ln w_j - \sum_i \lambda_{ij} d\ln N_{ij} \tag{G.19}$$

Totally differentiating the expression of N_{ij} and substituting the percentage change of N_{ij} into the previous equation, we have:

$$\begin{aligned} d\ln p_j^* &= d\ln w_j + \sum_i \lambda_{ij}\left[\theta\left(d\ln w_i - d\ln p_j^*\right) - d\ln \xi_{ij}\right] \\ &= \frac{1}{1+\theta} d\ln w_j + \frac{1}{1+\theta}\sum_i \lambda_{ij}\left[\theta d\ln w_i - d\ln \xi_{ij}\right] \end{aligned} \tag{G.20}$$

Hence, the percentage change in welfare satisfies:

$$d\ln U_j = \frac{\sigma}{\sigma-1}\left(d\ln w_j - d\ln p_j^*\right)$$
$$= -\frac{\sigma}{\sigma-1}\frac{1}{1+\theta}d\ln\lambda_{jj}$$

Integrating the previous expression between the initial equilibrium (before the shock) and the new equilibrium (after the shock), we finally get

$$\widehat{U}_j = \left(\widehat{\lambda}_{jj}\right)^{-\frac{\sigma}{\sigma-1}\frac{1}{1+\theta}} \tag{G.21}$$

It shows that the changes in welfare at country j can be inferred from changes in the share of domestic expenditure, λ_{jj}, using the parameter, $-\frac{\sigma}{\sigma-1}\frac{1}{1+\theta}$.

H No Variable Markup Case with $\bar{x} = 0$

We prove the welfare implication of our model with a constant markup. From the demand system, we have the representative consumer in country j's demand given by:

$$x_{ij}(\omega) = \frac{w_j L_j}{q_{ij}(\omega) P_{j\sigma}^{1-\sigma}}\left(\frac{p_{ij}(\omega)}{q_{ij}(\omega)}\right)^{-\sigma} \tag{H.1}$$

where $P_{j\sigma} = \left\{\sum_i \int_{\omega\in\Omega_{ij}} \tilde{p}_{ij}(\omega)^{1-\sigma} d\omega\right\}^{\frac{1}{1-\sigma}}$. To make our derivation compact, we define $\tilde{p}_{ij}(\omega) = p_{ij}(\omega)/q_{ij}(\omega)$. We thus can write quantity, sales, and profit for a given variety exported from i to j as follows,

$$x_{ij}(\omega) = \frac{w_j L_j}{q_{ij}(\omega)}\frac{\tilde{p}_{ij}(\omega)^{-\sigma}}{P_{j\sigma}^{1-\sigma}} \tag{H.2}$$

$$r_{ij}(\omega) = w_j L_j \frac{\tilde{p}_{ij}(\omega)^{1-\sigma}}{P_{j\sigma}^{1-\sigma}} \tag{H.3}$$

$$\pi_{ij}(\omega) = w_j L_j \left[\tilde{p}_{ij}(\omega) - \tilde{c}_{ij}(\omega)\right]\frac{\tilde{p}_{ij}(\omega)^{-\sigma}}{P_{j\sigma}^{1-\sigma}} \tag{H.4}$$

where $\tilde{c}_{ij}(\omega) = c_{ij}(\omega)/q_{ij}(\omega)$ is the quality adjusted marginal cost, where $c_{ij}(\omega)$ is the marginal cost of production. Given the quality adjusted marginal cost, firms maximize their profits. This implies that the optimal price of the good satisfies:

$$\tilde{p}_{ij}(\omega) = \frac{\sigma-1}{\sigma}\tilde{c}_{ij}(\omega) \tag{H.5}$$

In a similar spirit as in Feenstra and Romalis (2014), the marginal cost of producing a variety of final good with quality q_{ij} by a firm with productivity φ is:

$$c_{ij}(\varphi,\varepsilon) = \left(T_{ij}w_i + \frac{w_i\tau_{ij}q_{ij}^{\eta}}{\varphi}\right)\varepsilon$$

where φ follows the Pareto distribution with c.d.f. $G_i(\varphi) = 1 - b_i\varphi^{-\theta}$ and ε is drawn from a log normal distribution with zero mean and variance σ_{ε}^2. From the first-order condition associated with the previous marginal cost equation, the optimal level of quality for a firm with productivity φ is:

$$q_{ij}(\varphi,\varepsilon) = \left[\frac{T_{ij}\varphi}{(\eta-1)\,\tau_{ij}}\right]^{\frac{1}{\eta}} \tag{H.6}$$

and hence the quality adjusted marginal cost of production, the quality adjusted marginal cost and the export profit could be rewritten as:

$$\tilde{c}_{ij}(\varphi,\varepsilon) = \frac{c_{ij}(\varphi,\varepsilon)}{q_{ij}(\varphi,\varepsilon)} = \left(\frac{\eta}{\eta-1}T_{ij}w_i\right)^{\frac{\eta-1}{\eta}}\left(\frac{\varphi}{\eta w_i\tau_{ij}}\right)^{-\frac{1}{\eta}}\varepsilon \tag{H.7}$$

$$\tilde{p}_{ij}(\omega) = \frac{\sigma-1}{\sigma}\left(\frac{\eta}{\eta-1}T_{ij}w_i\right)^{\frac{\eta-1}{\eta}}\left(\frac{\varphi}{\eta w_i\tau_{ij}}\right)^{-\frac{1}{\eta}}\varepsilon \tag{H.8}$$

$$\pi_{ij}(\omega) = \frac{1}{\sigma}w_jL_j\frac{\tilde{p}_{ij}(\omega)^{1-\sigma}}{P_{j\sigma}^{1-\sigma}} \tag{H.9}$$

There is also an export fixed cost $f_{ij}w_i$, which need to pay before the exporting. As a result, only a fraction of firms will export and export produtivity cutoff satisfies:

$$\varphi_{ij}^{*} = \left[\frac{\sigma-1}{\sigma}\left(\frac{\eta}{\eta-1}T_{ij}w_i\right)^{\frac{\eta-1}{\eta}}(\eta w_i\tau_{ij})^{\frac{1}{\eta}}\varepsilon\left(\frac{\sigma w_i f_{ij}P_{j\sigma}^{1-\sigma}}{w_jL_j}\right)^{\frac{1}{\sigma-1}}\right]^{\eta} \tag{H.10}$$

With these definitions in mind, the aggregate price statistics, $P_{j\sigma}$, can be rewritten as:

$$P_{j\sigma} = \left\{\frac{\eta\theta\kappa}{\eta\theta-(\sigma-1)}\sum_i b_iJ_i\left(\frac{\sigma-1}{\sigma}\left(\frac{\eta}{\eta-1}T_{ij}w_i\right)^{\frac{\eta-1}{\eta}}(\eta w_i\tau_{ij})^{\frac{1}{\eta}}\right)^{-\theta\eta}\left(\frac{\sigma w_if_{ij}}{w_jL_j}\right)^{\frac{\sigma-1-\theta\eta}{\sigma-1}}\right\}^{-\frac{1}{\theta\eta}} \tag{H.11}$$

where κ is a constant. The bilateral trade flow, X_{ij}, would satisfy:

$$X_{ij} = N_{ij}\int_0^{\infty}\int_{\varphi_{ij}^{*}}^{\infty} r_{ij}(\varphi,\varepsilon)\,\mu_{ij}(\varphi,\varepsilon)\,f(\varepsilon)\,d\varphi d\varepsilon \tag{H.12}$$

$$= \frac{\eta\theta\kappa}{\eta\theta-(\sigma-1)}b_iJ_iw_jL_j\frac{\left(\frac{\sigma-1}{\sigma}\left(\frac{\eta}{\eta-1}T_{ij}w_i\right)^{\frac{\eta-1}{\eta}}(\eta w_i\tau_{ij})^{\frac{1}{\eta}}\right)^{-\theta\eta}\left(\frac{\sigma w_if_{ij}}{w_jL_j}\right)^{\frac{\sigma-1-\theta\eta}{\sigma-1}}}{P_{j\sigma}^{-\theta\eta}} \tag{H.13}$$

Firm's profits equals to the total fixed cost paid, which yields the free entry condition:

$$w_i f = \frac{1}{\sigma}\frac{1}{J_i}\sum_j X_{ij} = \frac{1}{\sigma}\frac{w_i L_i}{J_i} \tag{H.14}$$

where the last equality stems from that total income equals to total expenditure. Hence, the potential firm mass is

$$J_i = \frac{L_i}{\sigma f}$$

Now, the percentage change of U_j satisfies:

$$d\ln U_j = d\ln w_j - d\ln P_{j\sigma} \tag{H.15}$$

Now, λ_{jj} satisfies:

$$\lambda_{jj} = \frac{X_{jj}}{\sum_i X_{ij}} = \frac{b_j L_j \left(\left(T_{jj}^{\eta-1}\tau_{jj}\right)^{\frac{1}{\eta}} w_j\right)^{-\theta\eta} \left(w_j f_{jj}\right)^{\frac{\sigma-1-\theta\eta}{\sigma-1}}}{\sum_i b_i L_i \left(\left(T_{ij}^{\eta-1}\tau_{ij}\right)^{\frac{1}{\eta}} w_i\right)^{-\theta\eta} \left(w_i f_{ij}\right)^{\frac{\sigma-1-\theta\eta}{\sigma-1}}} \tag{H.16}$$

Consider the foreign shocks: τ_{ij}, T_{ij}, f_{ij} are changed to τ'_{ij}, T'_{ij}, f'_{ij} for $i \neq j$, respectively, such that $\tau_{jj} = \tau'_{jj}$ $T_{jj} = T'_{jj}$ and $f_{jj} = f'_{jj}$. Totally differentiating the previous equation implies:

$$d\ln \lambda_{jj} = \sum_i \lambda_{ij}\left[\left(\frac{\sigma}{\sigma-1}\theta\eta - 1\right)(d\ln w_i - d\ln w_j) - d\ln \xi_{ij}\right] \tag{H.17}$$

where $d\ln \xi_{ij}$ reflects any foreign shock, which satisfies:

$$d\ln \xi_{ij} = -\theta\eta\left(\frac{1}{\eta} d\ln \tau_{ij} + \frac{\eta-1}{\eta} d\ln T_{ij} + \left(\frac{1}{\sigma-1} - \frac{1}{\theta\eta}\right) d\ln f_{ij}\right) \tag{H.18}$$

The expression of $P_{j\sigma}$ implies that:

$$d\ln P_{j\sigma} = \sum_i \lambda_{ij}\left[d\ln w_i + \left(\frac{1}{\sigma-1} - \frac{1}{\theta\eta}\right)(d\ln w_i - d\ln w_j) - \frac{1}{\theta\eta} d\ln \xi_{ij}\right] \tag{H.19}$$

Hence, the percentage change in welfare satisfies:

$$\begin{aligned} d\ln U_j &= -\sum_i \lambda_{ij}\left[\left(\frac{\sigma}{\sigma-1} - \frac{1}{\theta\eta}\right)(d\ln w_i - d\ln w_j) - \frac{1}{\theta\eta} d\ln \xi_{ij}\right] \\ &= -\frac{1}{\theta\eta} d\ln \lambda_{jj} \end{aligned}$$

Integrating the previous expression between the initial equilibrium (before the shock) and the new equilibrium (after the shock), we finally get

$$\widehat{U}_j = \left(\widehat{\lambda}_{jj}\right)^{-\frac{1}{\theta\eta}} \tag{H.20}$$

It shows that the changes in welfare at country j can be inferred from changes in the share of domestic expenditure, λ_{jj}, using the parameter, $-\frac{1}{\theta\eta}$.

I Derivation for Welfare Comparison

I.1 Quality Case with T_{ij}

The representative consumer has preferences of:

$$U_j = \left[\sum_i \int_{\omega\in\Omega_{ij}} \left(q_{ij}(\omega)x^c_{ij}(\omega) + \overline{x}\right)^{\frac{\sigma-1}{\sigma}} d\omega\right]^{\frac{\sigma}{\sigma-1}} = \frac{y_j + \overline{x}P_j}{P_{j\sigma}} = \frac{\beta_\sigma}{\beta_\sigma - \beta}\frac{w_j}{P_{j\sigma}} \tag{I.1}$$

where $P_{j\sigma} = \left\{\sum_i J_i \int_0^\infty \int_{\varphi^*_{ij}(\varepsilon)}^\infty \tilde{p}_{ij}(\varphi,\varepsilon)^{1-\sigma} g_i(\varphi) f(\varepsilon) d\varphi d\varepsilon\right\}^{\frac{1}{1-\sigma}}$. Totally differentiating the previous equation, we have:

$$\begin{aligned} d\ln U_j &= d\ln w_j - d\ln P_{j\sigma} \\ &= d\ln w_j - \sum_i \lambda_{ij}\left(\frac{1}{\sigma-1} d\ln\left[J_i \int_0^\infty \int_{\varphi^*_{ij}(\varepsilon)}^\infty \tilde{p}_{ij}(\varphi,\varepsilon)^{1-\sigma} g_i(\varphi) f(\varepsilon) d\varphi d\varepsilon\right]\right) \end{aligned}$$

where

$$\begin{aligned} &\frac{1}{\sigma-1} d\ln\left[J_i \int_0^\infty \int_{\varphi^*_{ij}(\varepsilon)}^\infty \tilde{p}_{ij}(\varphi,\varepsilon)^{1-\sigma} g_i(\varphi) f(\varepsilon) d\varphi d\varepsilon\right] \\ = &-\frac{\int_0^\infty \int_{\varphi^*_{ij}(\varepsilon)}^\infty \tilde{p}_{ij}(\varphi,\varepsilon)^{1-\sigma} d\ln\tilde{p}_{ij}(\varphi,\varepsilon) g_i(\varphi) f(\varepsilon) d\varphi d\varepsilon}{\int_0^\infty \int_{\varphi^*_{ij}(\varepsilon)}^\infty \tilde{p}_{ij}(\varphi,\varepsilon)^{1-\sigma} g_i(\varphi) f(\varepsilon) d\varphi d\varepsilon} \\ &+\frac{1}{\sigma-1} d\ln J_i \\ &+\frac{1}{\sigma-1}\frac{\int_0^\infty \left(\tilde{p}^*_j\right)^{1-\sigma} g_i\left(\varphi^*_{ij}(\varepsilon)\right) \varphi^*_{ij}(\varepsilon) f(\varepsilon) d\varepsilon}{\int_0^\infty \int_{\varphi^*_{ij}(\varepsilon)}^\infty \tilde{p}_{ij}(\varphi,\varepsilon)^{1-\sigma} g_i(\varphi) f(\varepsilon) d\varphi d\varepsilon} d\ln\varphi^*_{ij} \end{aligned}$$

where the first term is the effects of changes in the prices of existing varieties calculated in ACDR; the second term is the effects of a change in potential firm entrants; the third term is the impact on welfare associated with the change in cutoff. Same as ACDR, the effects of changes in potential firm entrants, $d\ln J_i = 0$. However, the third term, the impact from a change in cutoff, is not infinitesimal, which should be larger than the gap between GT_j^{Bench} and $GT_j^{con\ mkp}$. The welfare change in our benchmark model are given by $-\frac{\sigma}{\sigma-1}\frac{1}{1+\theta\eta}\widehat{\lambda}_{jj}$ and the welfare change under the model without markup is given by $-\frac{\widehat{\lambda}_{jj}}{\theta\eta}$. Hence, their gap equals to

$$
\begin{aligned}
& -\frac{\sigma}{\sigma-1}\frac{1}{1+\theta\eta}\widehat{\lambda}_{jj}-\left(-\frac{\widehat{\lambda}_{jj}}{\theta\eta}\right) \\
= & -\frac{\theta\eta-(\sigma-1)}{\theta\eta\left[\sigma-1\right]}\frac{1}{1+\eta\theta}d\ln\lambda_{jj}
\end{aligned}
$$

In the following, we will prove that the third term is larger than this gap, $-\frac{\theta\eta-(\sigma-1)}{\theta\eta[\sigma-1]}\frac{1}{1+\eta\theta}d\ln\lambda_{jj}$. Hence, if we only focus on the first term by ignoring the extensive margin, the gain from trade in our benchmark model, GT_j^{bench}, is less than $GT_j^{con\ mkp}$. However, if including extensive, the gain from trade in our benchmark model, GT_j^{bench}, should be larger than $GT_j^{con\ mkp}$.

Proof: The third term could be rewritten as:

$$
\begin{aligned}
& \frac{1}{\sigma-1}\frac{\int_0^\infty\left(\tilde{p}_j^*\right)^{1-\sigma}g_i\left(\varphi_{ij}^*\left(\varepsilon\right)\right)\varphi_{ij}^*\left(\varepsilon\right)f\left(\varepsilon\right)d\varepsilon}{\int_0^\infty\int_{\varphi_{ij}^*(\varepsilon)}^\infty\tilde{p}_{ij}\left(\varphi,\varepsilon\right)^{1-\sigma}g_i\left(\varphi\right)f\left(\varepsilon\right)d\varphi d\varepsilon}d\ln\varphi_{ij}^* \\
= & \frac{1}{\sigma-1}\frac{\int_0^\infty g_i\left(\varphi_{ij}^*\left(\varepsilon\right)\right)\varphi_{ij}^*\left(\varepsilon\right)f\left(\varepsilon\right)d\varepsilon}{\beta\int_0^\infty\left[1-G_{ij}\left(\varphi_{ij}^*\left(\varepsilon\right)\right)\right]f\left(\varepsilon\right)d\varepsilon}d\ln\varphi_{ij}^* \\
= & \frac{1}{\sigma-1}\frac{\theta}{\beta}d\ln\varphi_{ij}^*
\end{aligned}
$$

where $\beta=\int_{\varphi_{ij}^*(\varepsilon)}^\infty\left[\frac{\tilde{p}_{ij}(\varphi,\varepsilon)}{\tilde{p}_j^*}\right]^{1-\sigma}\frac{g_i(\varphi)}{1-G_{ij}\left(\varphi_{ij}^*(\varepsilon)\right)}d\varphi$ is constant. Consider that $\frac{\tilde{p}_{ij}(\varphi,\varepsilon)}{\tilde{p}_j^*}>\frac{\tilde{c}_{ij}(\varphi,\varepsilon)}{\tilde{p}_j^*}=\left(\frac{\varphi}{\varphi_{ij}^*(\varepsilon)}\right)^{-\frac{1}{\eta}}$, we know that β could satisfy

$$
\begin{aligned}
\beta & < \int_{\varphi_{ij}^*(\varepsilon)}^\infty\left[\left(\frac{\varphi}{\varphi_{ij}^*\left(\varepsilon\right)}\right)^{-\frac{1}{\eta}}\right]^{1-\sigma}\theta\left(\varphi_{ij}^*\left(\varepsilon\right)\right)^{\theta+1}\varphi^{-\theta-1}d\frac{\varphi}{\varphi_{ij}^*\left(\varepsilon\right)} \\
& = \int_{\varphi_{ij}^*(\varepsilon)}^\infty\theta\left(\frac{\varphi}{\varphi_{ij}^*\left(\varepsilon\right)}\right)^{-\frac{\theta\eta-(\sigma-1)}{\eta}-1}d\left(\frac{\varphi}{\varphi_{ij}^*\left(\varepsilon\right)}\right)=\frac{\theta\eta}{\theta\eta-(\sigma-1)}
\end{aligned}
$$

The expression of $N_{ij}=J_i\int_0^\infty\left[1-G_{ij}\left(\varphi_{ij}^*\left(\varepsilon\right)\right)\right]f\left(\varepsilon\right)d\varepsilon$ implies that:

$$
d\ln\varphi_{ij}^*=-\frac{1}{\theta}d\ln N_{ij}
$$

which implies that the impact of cutoff on welfare satisfies:

$$
\begin{aligned}
& -\frac{1}{\sigma-1}\sum_i\lambda_{ij}\frac{\int_0^\infty\left(\tilde{p}_j^*\right)^{1-\sigma}g_i\left(\varphi_{ij}^*\left(\varepsilon\right)\right)\varphi_{ij}^*\left(\varepsilon\right)f\left(\varepsilon\right)d\varepsilon}{\int_0^\infty\int_{\varphi_{ij}^*(\varepsilon)}^\infty\tilde{p}_{ij}\left(\varphi,\varepsilon\right)^{1-\sigma}g_i\left(\varphi\right)f\left(\varepsilon\right)d\varphi d\varepsilon}d\ln\varphi_{ij}^* \\
= & -\frac{1}{\sigma-1}\frac{\theta}{\beta}\sum_i\lambda_{ij}d\ln\varphi_{ij}^*>-\frac{\theta\eta-(\sigma-1)}{\eta\left[\sigma-1\right]}\sum_i\lambda_{ij}d\ln\varphi_{ij}^* \\
= & \frac{\theta\eta-(\sigma-1)}{\theta\eta\left[\sigma-1\right]}\sum_i\lambda_{ij}d\ln N_{ij}=-\frac{\theta\eta-(\sigma-1)}{\theta\eta\left[\sigma-1\right]}\frac{1}{1+\eta\theta}d\ln\lambda_{jj}
\end{aligned}
$$

This implies that the impact on welfare associated with the change in cutoff should be larger than $-\frac{\theta\eta-(\sigma-1)}{\theta\eta[\sigma-1]}\frac{1}{1+\eta\theta}d\ln\lambda_{jj}$.

I.2 Fixed Quality Case without T_{ij}

The representative consumer has preferences of:

$$U_j = \left[\sum_i \int_{\omega\in\Omega_{ij}} \left(x_{ij}^c(\omega)+\overline{x}\right)^{\frac{\sigma-1}{\sigma}} d\omega\right]^{\frac{\sigma}{\sigma-1}} = \frac{\beta_\sigma}{\beta_\sigma-\beta}\frac{w_j}{P_{j\sigma}} \tag{I.2}$$

where $P_{j\sigma} = \left\{\sum_i J_i \int_{\varphi_{ij}^*}^{\infty} p_{ij}(\varphi)^{1-\sigma} g_i(\varphi)\, d\varphi\right\}^{\frac{1}{1-\sigma}}$. Totally differentiating the previous equation, we have:

$$\begin{aligned} d\ln U_j &= d\ln w_j - d\ln P_{j\sigma} \\ &= d\ln w_j - \sum_i \lambda_{ij}\left(\frac{1}{\sigma-1}d\ln\left[J_i\int_{\varphi_{ij}^*}^{\infty} p_{ij}(\varphi)^{1-\sigma} g_i(\varphi)\, d\varphi\right]\right) \end{aligned}$$

where

$$\begin{aligned} &\frac{1}{\sigma-1}d\ln\left[J_i\int_{\varphi_{ij}^*}^{\infty} p_{ij}(\varphi)^{1-\sigma} g_i(\varphi)\, d\varphi\right] \\ = \; & -\frac{\int_{\varphi_{ij}^*}^{\infty} p_{ij}(\varphi)^{1-\sigma} d\ln p_{ij}(\varphi)\, g_i(\varphi)\, d\varphi}{\int_{\varphi_{ij}^*}^{\infty} p_{ij}(\varphi)^{1-\sigma} g_i(\varphi)\, d\varphi} \\ & +\frac{1}{\sigma-1}d\ln J_i \\ & +\frac{1}{\sigma-1}\frac{\left(\tilde{p}_j^*\right)^{1-\sigma} g_i\left(\varphi_{ij}^*\right)\varphi_{ij}^*}{\int_{\varphi_{ij}^*}^{\infty} p_{ij}(\varphi)^{1-\sigma} g_i(\varphi)\, d\varphi}d\ln\varphi_{ij}^* \end{aligned}$$

where the first term is the effects of changes in the prices of existing varieties calculated in ACDR; the second term is the effects of a change in potential firm entrants; the third term is the impact on welfare associated with the change in cutoff. Same as ACDR, the effects of changes in potential firm entrants, $d\ln J_i = 0$. However, the third term, the impact from a change in cutoff, is not infinitesimal, which should be larger than the gap between $GT_j^{no\ q}$ and $GT_j^{no\ q,\ con\ mkp}$. The welfare changes under variable markups but no Washington Apples mechanism are given by $GT_j^{no\ q} = -\frac{\sigma}{\sigma-1}\frac{1}{1+\theta}\widehat{\lambda}_{jj}$ and the welfare change under the model without both endogenous quality and variable markup is given by $GT_j^{no\ q,\ con\ mkp} = -\frac{\widehat{\lambda}_{jj}}{\theta}$. Hence, their gap equals to

$$
\begin{aligned}
& -\frac{\sigma}{\sigma-1}\frac{1}{1+\theta}\widehat{\lambda}_{jj}-\left(-\frac{\widehat{\lambda}_{jj}}{\theta}\right) \\
= & -\frac{\theta-(\sigma-1)}{\theta\left[\sigma-1\right]}\frac{1}{1+\theta}d\ln\lambda_{jj}
\end{aligned}
$$

In the following, we will prove that the third term is larger than this gap, $-\frac{\theta-(\sigma-1)}{\theta[\sigma-1]}\frac{1}{1+\theta}d\ln\lambda_{jj}$. Hence, if we only focus on the first term by ignoring the extensive margin, the gain from trade under variable markups but no Washington Apples mechanism, $GT_j^{no\ q}$, is less than $GT_j^{no\ q,\ con\ mkp}$. However, if including extensive margin, the gain from trade under variable markups but no Washington Apples mechanism, $GT_j^{no\ q}$, should be larger than $GT_j^{no\ q,\ con\ mkp}$.

Proof: The third term could be rewritten as:

$$
\begin{aligned}
& \frac{1}{\sigma-1}\frac{\left(\tilde{p}_j^*\right)^{1-\sigma}g_i\left(\varphi_{ij}^*\right)\varphi_{ij}^*}{\int_{\varphi_{ij}^*}^{\infty}p_{ij}\left(\varphi\right)^{1-\sigma}g_i\left(\varphi\right)d\varphi}d\ln\varphi_{ij}^* \\
= & \frac{1}{\sigma-1}\frac{\theta}{\beta}d\ln\varphi_{ij}^*
\end{aligned}
$$

where $\beta=\int_{\varphi_{ij}^*}^{\infty}\left[\frac{p_{ij}(\varphi)}{p_j^*}\right]^{1-\sigma}\frac{g_i(\varphi)}{1-G_{ij}\left(\varphi_{ij}^*\right)}d\varphi$ is constant. Consider that $\frac{p_{ij}(\varphi)}{p_j^*}>\frac{c_{ij}(\varphi)}{p_j^*}=\left(\frac{\varphi}{\varphi_{ij}^*}\right)^{-1}$, we know that β could satisfy

$$
\beta<\int_{\varphi_{ij}^*}^{\infty}\theta\left(\frac{\varphi}{\varphi_{ij}^*}\right)^{-(\theta-(\sigma-1))-1}d\left(\frac{\varphi}{\varphi_{ij}^*}\right)=\frac{\theta}{\theta-(\sigma-1)}
$$

The expression of $N_{ij}=J_i\left[1-G_{ij}\left(\varphi_{ij}^*\right)\right]$ implies that:

$$
d\ln\varphi_{ij}^*=-\frac{1}{\theta}d\ln N_{ij}
$$

which implies that the impact of cutoff on welfare satisfies:

$$
\begin{aligned}
& -\frac{1}{\sigma-1}\sum_i\lambda_{ij}\frac{\left(\tilde{p}_j^*\right)^{1-\sigma}g_i\left(\varphi_{ij}^*\right)\varphi_{ij}^*}{\int_{\varphi_{ij}^*}^{\infty}p_{ij}\left(\varphi\right)^{1-\sigma}g_i\left(\varphi\right)d\varphi}d\ln\varphi_{ij}^* \\
= & -\frac{1}{\sigma-1}\frac{\theta}{\beta}\sum_i\lambda_{ij}d\ln\varphi_{ij}^*>-\frac{\theta-(\sigma-1)}{\sigma-1}\sum_i\lambda_{ij}d\ln\varphi_{ij}^* \\
= & -\frac{\theta-(\sigma-1)}{\theta\left[\sigma-1\right]}\frac{1}{1+\theta}d\ln\lambda_{jj}
\end{aligned}
$$

This implies that the impact on welfare associated with the change in cutoff should be larger than $-\frac{\theta-(\sigma-1)}{\theta[\sigma-1]}\frac{1}{1+\theta}d\ln\lambda_{jj}$.

J Supplementary Table: Welfare Comparison for All Countries

country	Bench	no q	con mkp	no q, con mkp
AUS	4.131	26.684	1.747	6.077
AUT	6.391	38.485	2.721	9.347
BEL	10.731	56.618	4.630	15.521
BRA	1.114	7.910	0.467	1.651
CAN	5.925	36.196	2.519	8.676
CHE	7.154	42.082	3.053	10.444
CHN	1.636	11.425	0.686	2.421
DEU	3.934	25.566	1.662	5.789
DNK	5.955	36.348	2.532	8.720
ESP	3.703	24.242	1.564	5.453
FIN	3.805	24.827	1.607	5.601
FRA	3.478	22.929	1.468	5.124
GBR	4.706	29.857	1.993	6.912
GRC	4.294	27.595	1.816	6.313
HKG	10.800	56.864	4.661	15.618
IDN	2.565	17.403	1.080	3.788
IND	1.037	7.384	0.435	1.537
IRL	7.951	45.638	3.401	11.583
ITA	2.273	15.565	0.956	3.359
JPN	1.292	9.125	0.542	1.914
KOR	2.314	15.820	0.973	3.418
MEX	4.513	28.805	1.910	6.632
MYS	6.530	39.154	2.781	9.547
NLD	5.977	36.453	2.541	8.750
NOR	5.187	32.420	2.200	7.609
POL	3.453	22.779	1.457	5.087
PRT	4.643	29.514	1.966	6.820
RUS	2.445	16.650	1.029	3.612
SAU	4.688	29.763	1.986	6.887
SGP	13.372	65.218	5.819	19.208
SWE	4.714	29.899	1.996	6.923
THA	4.962	31.231	2.103	7.283
TUR	2.436	16.595	1.025	3.599
TWN	5.045	31.672	2.139	7.404
USA	2.130	14.647	0.895	3.148
ZAF	2.112	14.533	0.888	3.122

K Supplementary Figure

Figure 9: Sales and Markup Distribution

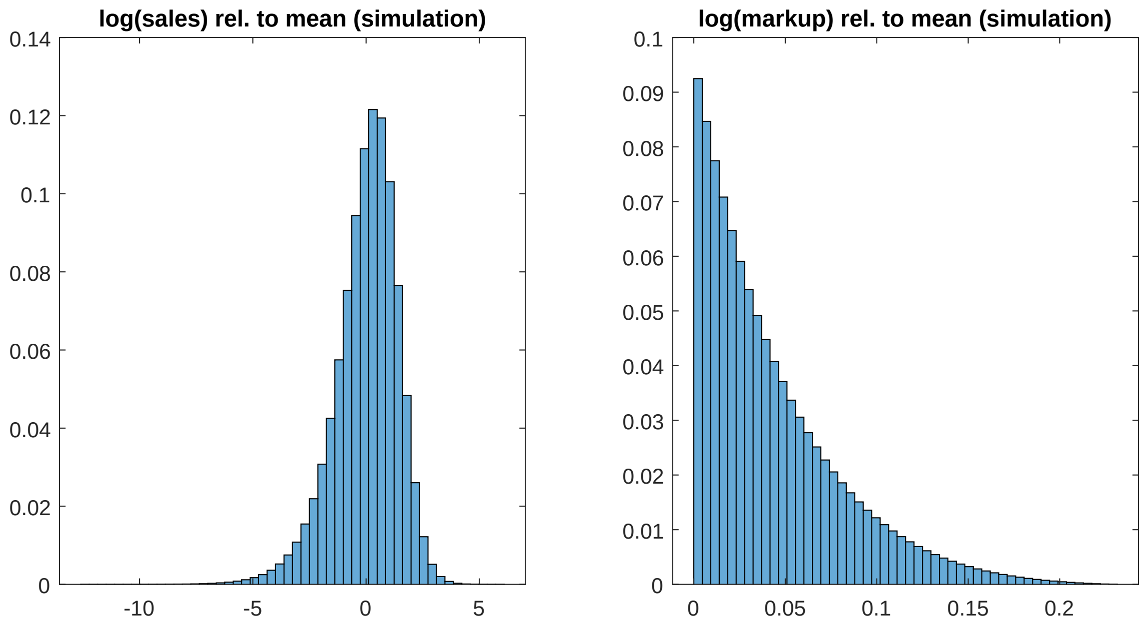


Figure 10: The relationship between market size and firm-level variables (prices, sales, and quality)

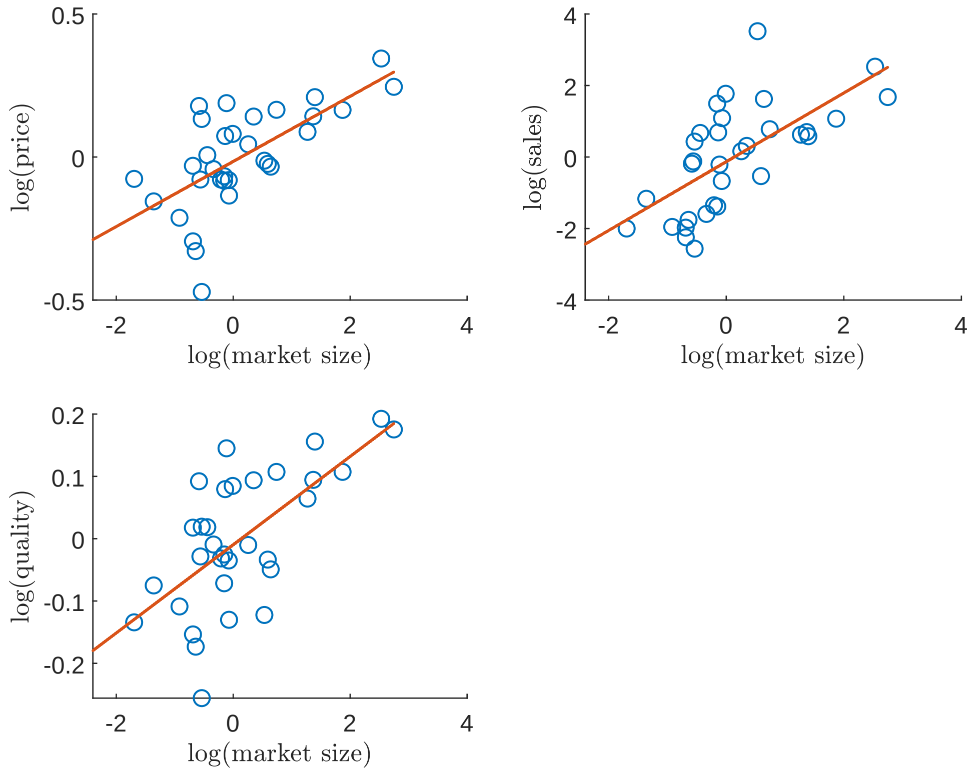
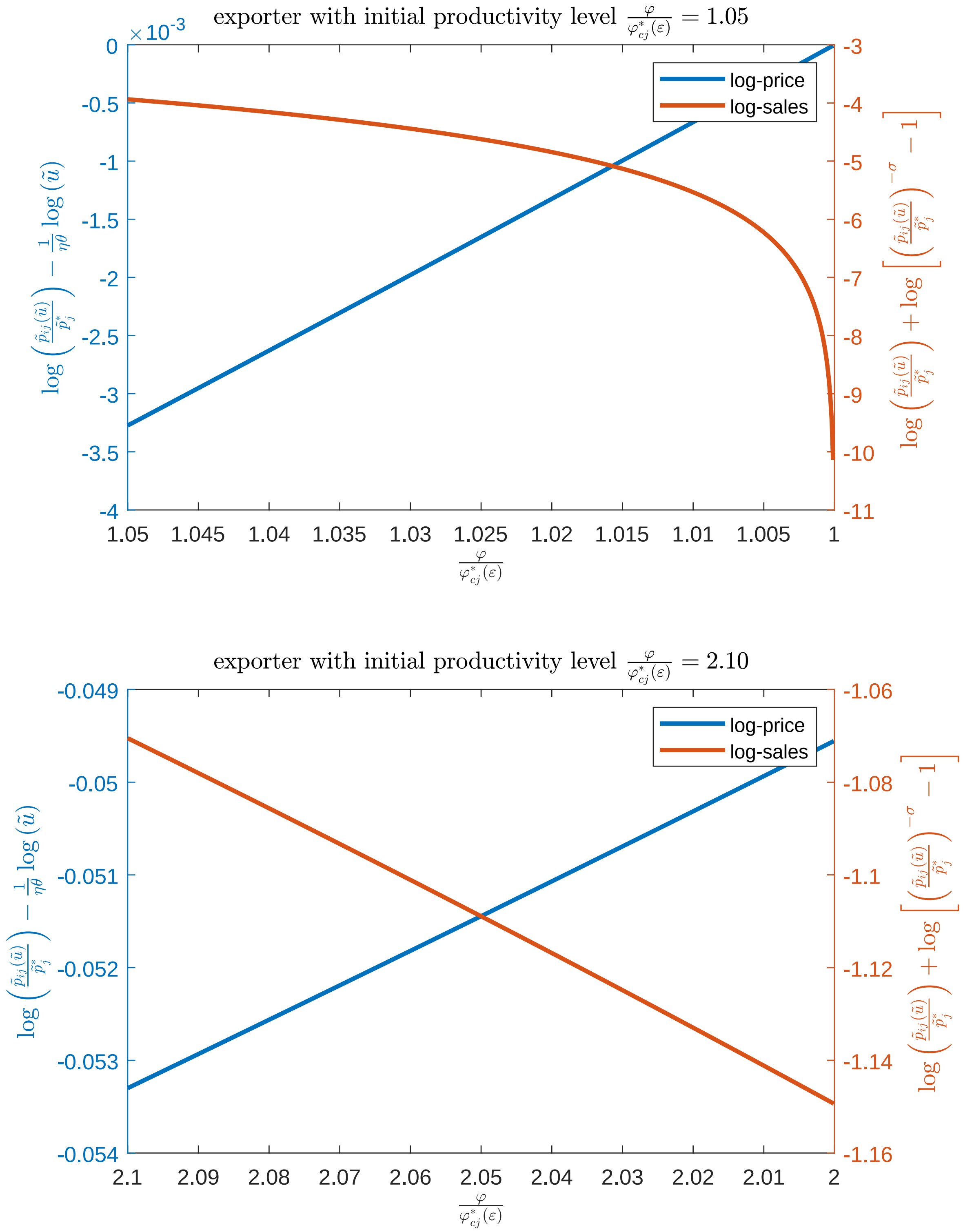


Figure 11: Illustration: the Changes in Prices and Sales by Low- vs. High-productivity Firms after Trade Cost Shock



Explanatory notes on Figure 11:

The upper panel plots a low-productivity firm whose productivity is only 5% above the cutoff productivity before the trade shock, i.e., $\frac{\varphi}{\varphi^*_{cj}(\varepsilon)} = 1.05$. When trade cost increases by

5% (either from τ or T), $\frac{\varphi}{\varphi^*_{cj}(\varepsilon)}$ goes to 1. Then, this producer starts to become a marginal exporter. The left y-axis plots the change of log(price), and the right y-axis plots the change of log(sales). Clearly, the variation in price changes is very small whereas the change in sales is large. Next, we turn to a initially high-productivity firm with $\frac{\varphi}{\varphi^*_{cj}(\varepsilon)} = 2.10$ shown in the lower panel. When it is hit by 5% increase in trade cost, the changes in log(price) is similar comparing with the low-productivity exporter in the upper panel, but the change in log(sales) is much smaller for this high-productivity firm.