Journal of Global Economic Analysis, Volume 1 (2016), No. 1, pp. 111-180.

GTAP-HET: Introducing Firm Heterogeneity into the GTAP Model

By Zeynep Akgul[a], Nelson B. Villoria[b] and Thomas W. Hertel[c]

Computable General Equilibrium (CGE) models incorporating firm heterogeneity can overcome the shortcomings of traditional Armington-based models in explaining changes in productivity and variety in the wake of reduced trade costs. In this paper, we present a new modeling framework where the firm heterogeneity theory of Melitz is introduced into the Global Trade Analysis Project (GTAP) model and calibrated to the GTAP 8 Data Base. The new mechanisms in the model are demonstrated in a stylized scenario with 3 regions (USA, Japan and the Rest of the World) and 2 sectors (manufacturing and non-manufacturing) where the elimination of tariffs levied by Japan on the import of US manufacturing goods is examined. Results are compared with those under monopolistic competition motivated by Krugman and under perfect competition motivated by Armington. The firm heterogeneity model incorporates endogenous variety, scale, productivity, and fixed cost effects into welfare change in addition to the traditional allocative efficiency and terms of trade effects. We observe that these effects are significant sources of welfare change. GTAP-HET presents the first GTAP implementation of firm heterogeneity. It is a powerful tool for policy analysis with improved abilities in tracing out productivity changes and entry/exit of firms following trade liberalization scenarios.

JEL codes: C68, D40, D58, F12

Keywords: Computable General Equilibrium, Firm Heterogeneity, Monopolistic Competition, Perfect Competition, Melitz, Krugman, Armington

1. Introduction

Traditional Computable General Equilibrium (CGE) models rely on the Armington (1969) assumption of national product differentiation to distinguish prefer-

[a] Center for Global Trade Analysis, Purdue University, 403 West State Street, West Lafayette, IN 47907 (zakgul@purdue.edu), Visiting Scholar at the US International Trade Commission.
[b] Department of Agricultural Economics, Kansas State University, 331 I Waters Hall, Manhattan, KS 66506 (nvilloria@ksu.edu).
[c] Center for Global Trade Analysis, Purdue University, 403 West State Street, West Lafayette, IN 47907 (hertel@purdue.edu).

ences between domestic and imported products. Changes in trade flows in these models are conditioned by pre-existing trade shares; therefore, they can only capture the trade adjustments that occur due to changes in current export volumes, i.e. intensive margin. This is at odds with the recent empirical trade literature that highlights the contribution of new varieties in export markets, i.e. extensive margin, to explain the expansion of trade following trade liberalization episodes (Hummels and Klenow, 2005; Chaney, 2008). The firm heterogeneity trade model proposed in the pioneering work of Melitz (2003) combines trade volume changes with expanding varieties as a result of trade liberalization by capturing the self-selection of firms into export markets based on their respective productivity levels. The resulting framework is solidly supported by empirical evidence (Eaton, Kortum, and Kramarz, 2004; Bernard, Jensen, and Schott, 2006). Given the importance of firm heterogeneity in both theory and empirical analysis, quantitative models that incorporate firm heterogeneity are needed in mainstream international trade policy analysis. Incorporating firm-heterogeneity into CGE models can improve the ability of these models to trace out trade and welfare implications of trade policies, which are previously unexplored in traditional models.

There have recently been some important efforts to incorporate Melitz (2003) into global CGE modeling (Zhai, 2008; Balistreri, Hillberry, and Rutherford, 2011; Balistreri and Rutherford, 2013; Oyamada, 2014; Dixon, Jerie, and Rimmer, 2015). However, a readily accessible Global Trade Analysis Project (GTAP) implementation with firm heterogeneity has not yet become available. Our paper addresses this gap by incorporating firm heterogeneity into the standard GTAP model, calibrating it to the GTAP 8 Data Base (Narayanan, Aguiar, and McDougall, 2012) and illustrating this framework with a stylized scenario. A comparison with the standard GTAP model with Armington assumption, as well as a monopolistically competitive GTAP model, allows us to shed light on the new elements which the Melitz model brings to bear on trade liberalization impacts.

One of the stylized facts shown by micro-level data is that there is significant variation across firms of the same industry. In particular, firms vary by their productivity, size, profitability, the number of markets served and responses to trade shocks (Bernard et al., 2003; Eaton, Kortum, and Kramarz, 2004; Bernard et al., 2007; Balistreri, Hillberry, and Rutherford, 2011; Melitz and Trefler, 2012). Moreover, only a small proportion of firms export and they tend to be larger and more productive than non-exporters (Balistreri, Hillberry, and Rutherford, 2011; Bernard et al., 2003; Bernard and Jensen, 1999). These stylized facts are captured by Melitz (2003) who examines the intra-industry reallocation effects of international trade in the context of a model with monopolistic competition and heterogeneous firms. In this framework, increasing the exposure to trade generates a redistribution of production across firms within the industry based on the productivity differences of firms. While firms with high productivity levels are induced to enter export markets, firms with low productivity levels continue to produce for the domestic market. The firms with the lowest productivity levels, on the other hand, are forced to exit the industry. These inter-firm reallocations generate a growth in the aggregate industry productivity which increases the welfare gains of trade. This channel is a unique feature of the firm heterogeneity model (Zhai, 2008). The main insight of the Melitz model is that trade induces changes in aggregate productivity even though the production technology of the country is the same. As opposed to the allocative efficiency gains in traditional trade models with homogeneous firms and Armington assumption, aggregate productivity changes in the firm heterogeneity model are brought about by the composition of firms within an industry.

Our main objective in the GTAP firm-heterogeneity model is to capture the effect of increased exposure to external markets on average industry productivity. This is accomplished when trade-induced changes in productivity thresholds are allowed to stimulate industry productivity, i.e. when productivity is endogenous. Thus, our main strategy is to endogenize the industry productivity, which is exogenous in the standard GTAP model, by linking it with endogenous productivity thresholds.

Melitz (2003) builds on Krugman's (1980) monopolistic competition framework to model trade; while he draws from Hopenhayn (1992) to model the endogenous self-selection of heterogeneous firms. Likewise, we build on the monopolistically competitive GTAP model (Swaminathan and Hertel, 1996), where variety effects (changes in the number of firms and hence distinct varieties offered) and scale effects (changes in output per firm) are captured. We draw from Zhai (2008) in modeling certain features of firm heterogeneity such as the specification of productivity thresholds for market entry and the calibration of fixed trading costs. This allows us to endogenize aggregate industry productivity in the monopolistically competitive sectors of the model, thereby capturing the intra-industry reallocation of resources in the wake of trade liberalization.

A contribution of our firm heterogeneity model is the decomposition of the welfare implications of trade policy. This is an extension of the existing GTAP welfare decomposition (Huff and Hertel, 2000), which now includes scale, variety, productivity, and fixed cost effects derived from the firm heterogeneity model, in addition to the traditional allocative efficiency and terms of trade effects.

In addition to the firm heterogeneity model, we also explore other model structures to highlight how trade policy impacts differ across various frameworks. These include monopolistically competitive GTAP model motivated by Krugman (1980) and perfectly competitive GTAP model motivated by the standard GTAP model with Armington (1969) assumption. Occasionally, we refer to these as Armington (1969) and Krugman (1980) models. However, the reader should keep in mind that even though these GTAP modules are motivated by Armington (1969) and Krugman (1980), they do not exactly follow the same structure as these seminal works. We mainly seek to bring the main features of these theories into a common framework.

The rest of the paper is organized as follows: Section 2 provides a brief introduction to the theory of firm heterogeneity. Section 3 details the implementation of firm heterogeneity theory into the standard GTAP model. Section 4 describes the data requirement for the firm heterogeneity model. Alternative closure rules for model switches are discussed in Section 5. Section 6 illustrates this framework with a stylized trade liberalization scenario. Section 7 concludes the paper.

2. Overview of firm heterogeneity

In this framework, there are two types of industries: (i) monopolistically competitive industries with heterogeneous firms that produce differentiated varieties and (ii) perfectly competitive industries with identical firms that produce homogeneous products which are assumed to be differentiated only at national scale. The characteristics of the standard GTAP model industries are retained in the perfectly competitive industries where a representative firm produces at constant returns to scale technology. The characteristics of firms in the monopolistically competitive industry, on the other hand, warrants a detailed discussion concerning the treatment of production, cost, and productivity.

The monopolistically competitive industry is characterized by a continuum of firms, each producing a single unique variety that is an imperfect substitute in demand to other varieties. In what follows we use firms and varieties interchangeably. While firms are free to enter or exit the market, entrance requires covering fixed set-up costs that are associated with expenses made during initial development of the differentiated variety. The existence of fixed set-up costs is a large impediment for start-up firms. However, it also creates potential scale economies in the monopolistically competitive industry. In this framework, until the firm makes a commitment to enter the industry by paying fixed set-up costs, there is no information on its productivity level. Since firms do not know their productivity with certainty until they begin production, they are assumed to be identical before entering the industry. Once they enter, their productivity levels are revealed and we observe that productivity is heterogeneous across firms within the industry. Consistent with the static nature of GTAP, we assume a static version of the Melitz (2003) model similar to Arkolakis, Costinot, and Rodriguez-Clare (2012) and Melitz and Redding (2013) where we abstract from the probability of a bad shock happening every period that might cause firm death.

In this context, productivity is defined as how much a firm can produce per composite input. It is inversely related to the marginal cost of production; therefore, a high-productivity firm is the one producing a similar variety at a lower marginal cost which follows from the simplification of Melitz (2003). Firm productivity is assumed to be identically and independently distributed with productivity following a Pareto distribution. Each firm draws its productivity out of this distribution and only then finds out where they stand on the productivity spectrum.

Once firms know their productive capabilities, they can choose whether or not to operate in the market. The decision to produce depends on the potential for making nonnegative profits given the productivity of the firm and fixed costs of market entry. Firms are assumed to face symmetric fixed costs, while they differ with respect to their productivity levels. Thus, production is carried out only by firms that are productive enough to cover the fixed costs of entering the market. High-productivity firms have a better chance of survival since they can produce at a lower cost compared to low-productivity firms. Competition in the market, therefore, forces low-productivity firms to exit (or not produce) and high-productivity firms to expand their shares in the domestic market.

Where does trade fit into this framework? Once a firm secures its niche in the domestic market, it has the choice to supply foreign markets as well as to satisfy home demand. The decision to export or not has its own challenges. Just as firms incur fixed set-up costs to start producing, they also incur region-specific fixed trading costs to start exporting, i.e. market access costs. They may arise due to expenses associated with distinguishing a firm's product to make it compatible with regional standards in the destination market. In addition, they may be associated with the expenses of finding local dealerships or with conforming the rules and regulations of export markets. For example, automobile companies incur the costs of redesigning certain features of their models in order to meet the needs of consumers in the destination market. The battery pack and the number of rows of seating in Prius 2010 differ between the European and Japanese markets, as does the placement of the steering wheel. Moreover international standards are different for car parts such as headlights, seat belts, and wiper blades. Another example can be the keyboard requirements of personal computers in different regions. A Dell sold in the Japanese market has a different keyboard design than the same Dell sold in the US market due to language differences.

Regardless of their nature, the very existence of fixed trading costs is the reason why only a subset of firms are able to export and why firms self-select into export markets based on their respective productivity levels. This mechanism works through the endogenous determination of the productivity threshold to export. Only the firms with productivity levels equal to or higher than this threshold find it profitable to supply that specific market. Hence the distribution of firms is such that while the most productive firms serve in the export markets, firms with lower productivity levels supply only the domestic market, and the lowest-productivity firms do not produce.

Self-selection of firms, first into the domestic market, then into export markets is a unique mechanism in the firm heterogeneity model and offers additional gains from trade due to improvements in industry productivity through inter-firm reallocation of resources. This is a channel that was previously unexplored in trade models. In conventional theory, trade leads to inter-sectoral reallocation of resources with scarce resources shifting towards the more profitable industry. However, in the presence of firm heterogeneity, competition for resources also occurs within the industry where high-productivity firms expand their market share and absorb the factors released by low-productivity firms. The expansion of high-productivity firms together with the exit of low-productivity firms in the face of trade liberalization, increases the productivity of the industry on average, generating additional gains from trade.

The discussion so far is based on the fact that productivity levels of firms are assumed to be constant. One could argue that trade also leads to 'learning by exporting' so that firms become more productive as they export. However, in this study we restrict our attention to Melitz (2003) and we abstract from endogenous changes in firm productivity levels in our model.

3. Implementation of firm heterogeneity in GTAP

This section describes the implementation of firm heterogeneity into the standard GTAP model. We explicitly show how to bring the theory into the GEMPACK programming language (Harrison and Pearson, 1996; Horridge, Pearson, and Rutherford, 2013) by providing code snippets where applicable. Definitions of variables used in the code are presented in Table A.1. In GEMPACK we work with endogenous variables in log-linear form. Therefore, the implementation of the model is accomplished by total differentiation of each equation introduced in subsequent sections.

The standard GTAP model is an industry-level framework that focuses on the behavior of a representative firm in the perfectly competitive industry where all firms are identical. However, in the Melitz model firms are heterogeneous with respect to their productivity levels. Since data on firm numbers and sales are limited, we aggregate firm-level variables into industry-level variables in the model. This aggregation eliminates the need for firm-level information to be included in the dataset.

We focus on the behavior of an average firm in order to incorporate productivity heterogeneity into the industry-level GTAP model without losing the firm-level insight. We define average firm as the firm whose productivity level equals the average productivity in the industry. This definition follows from Melitz (2003) and Dixon, Jerie, and Rimmer (2015). Average industry productivity is obtained by a weighted average of the productivity levels of active firms in the industry. As such, it fully summarizes the relevant firm-level information necessary to obtain aggregate outcomes in the industry (Melitz, 2003)[1]. Therefore, we can work with an average firm to develop a tractable model that provides aggregate outcomes without losing the firm-level distribution. Further details on the determination of average industry productivity are given in Section 3.2.4.

[1] In fact, as stated in Melitz (2003), the aggregate outcome of an industry with N firms that have identical productivity level $\widetilde{\Phi}$, is the same as the aggregate outcome of an industry with N firms of any distribution of productivity levels that yields the same aggregate productivity level $\widetilde{\Phi}$.

3.1 Demand

Demand structure in the GTAP firm heterogeneity model entails the implementation of *love-of-variety* in consumer utility and its implications for prices. These changes follow closely from the monopolistically competitive module of GTAP developed by Swaminathan and Hertel (1996).

We assume a multi-region world. Each region contains a perfectly competitive industry, where firms produce homogeneous products under constant returns to scale, and a monopolistically competitive industry, where heterogeneous firms produce differentiated products under increasing returns to scale.

Bilateral trade flows of homogeneous products are governed by the Armington (1969) assumption of national product differentiation, in which commodities sourced from different countries are imperfect substitutes in demand. Based on this assumption, a nested consumption structure is used wherein the geographical origin of the commodity matters only up to the border of the destination region. A composite imported commodity is formed at the border which no longer retains the respective geographical origin of the constituent commodity. This composite is, then, sourced to each agent in the economy and imperfectly substitutes for the domestically produced commodity. To do this stepwise aggregation, the determination of import sourcing is assumed to be independent of the price of domestic goods for trade in homogeneous products.

The Armington (1969) assumption does not apply to trade in differentiated commodities. Demand structure in the firm heterogeneity model is instead based on Krugman (1980)'s monopolistic competition. Under monopolistic competition, consumers are characterized by love-of-variety where they perceive each variety as a unique product and derive utility from this uniqueness. The sheer availability of different varieties benefits consumers. What matters for the consumer is the brand, irrespective of whether it is imported or produced domestically. Therefore, under this framework, as opposed to an import-domestic decision, consumers make a variety decision across differentiated products, where imports from different source regions directly compete with the domestic varieties. This implies that, unlike in the Armington case, the price of domestic varieties affect the sourcing of imported varieties for the monopolistically competitive products.

Tracking the source of varieties is still important in this structure as the geographical origin is associated with a particular set of exporters and hence varieties/brands/products. Therefore, in the firm heterogeneity model, we source imports to each agent, private household, government, and firms, and allow for direct competition of domestic and imported varieties in consumer demand. An important implication of this implementation is that the structure of the database needs to allow for sourcing varieties to agents in the model. To transform the data base accordingly, we follow Swaminathan and Hertel (1996) and define the market share of the source region s in the total imports of a product by importing region r. This share is then used to source out the imports consumed by the agent. In order to account for intra-regional imports, we include them in domestic sales. As a result, transformed trade flows represent intra-regional imports as well as the domestically produced goods when $s = r$, while they represent inter-regional imports when $s \neq r$. Therefore, the word "domestic" has a broader meaning in this paper. It refers to domestic as well as intra-regional changes. The data transformation is discussed in further detail in Section 4 and Appendix B.

In addition to love-of-variety, the model also incorporates the preference bias of consumers. In the GTAP model, consumer preference bias is built in via household expenditure shares in the spirit of Venables (1987). This treatment departs from the symmetric preference assumption of the monopolistically competitive models that follow Krugman (1980). However, empirical observations show that consumers prefer domestic varieties over imported varieties, often referred to as the "home bias" (McCallum, 1995; Obstfeld et al., 2001). As discussed in Lanclos and Hertel (1995), it is important to incorporate this empirical regularity in the model as trade policies distribute varieties from different sources in an asymmetric manner in the market (Venables, 1987). In this model, we retain the preference bias in expenditure shares as we believe that it is better suited to trade policy analysis.

3.1.1 Derived demand and price index

Following Dixit and Stiglitz (1977), we assume that preferences are given by a Constant Elasticity of Substitution (CES) utility function over a range of differentiated varieties. Let Q_{ir} be the aggregate product of i demanded in region r, which is equivalent to the utility $Q_{ir} \equiv U_{ir}$, and let $\omega \in \Omega_{isr}$ index varieties in the set of product i sourced from region s to r. Then Q_{ir} is a CES aggregate of all available varieties:

$$Q_{ir} = \left[\sum_{s} \int_{\omega \in \Omega_{isr}} Q_{isr}(\omega)^{\frac{\sigma_i - 1}{\sigma_i}} \, d\omega \right]^{\frac{\sigma_i}{\sigma_i - 1}} \tag{1}$$

where $\sigma_i > 1$ is the constant elasticity of substitution amongst varieties, $Q_{isr}(\omega)$ is consumer demand in region r for variety ω of product i sourced from region s.

The dual to the utility function is the CES unit expenditure function. Let P_{ir} be the CES price index of product i in region r:

$$P_{ir} = \left[\sum_{s} \int_{\omega \in \Omega_{isr}} P_{isr}(\omega)^{1-\sigma_i} \, d\omega \right]^{\frac{1}{1-\sigma_i}} \tag{2}$$

where $P_{isr}(\omega)$ is the price in region r of variety ω of product i sourced by region s.

Equations (1) and (2) aggregate demand and price of individual varieties over a continuum. To facilitate numerical implementation, we translate these equations into their discrete counterparts by using Melitz' definition of an average firm[2]. Melitz (2003) defines an average firm as the one that produces at a productivity level that is equal to the industry average. By using the average firm, we can define average firm price, $\widetilde{P}_{isr}$, and demand for an average product, $\widetilde{Q}_{isr}$, where the accent tilde (˜) above a variable denotes that it is an average.

Using these definitions, Equations (1) and (2) are discretized as:

$$Q_{ir} = \left[\sum_s N_{isr} \widetilde{Q}_{isr}^{\frac{\sigma_i - 1}{\sigma_i}} \right]^{\frac{\sigma_i}{\sigma_i - 1}}, \; P_{ir} = \left[\sum_s N_{isr} \widetilde{P}_{isr}^{1-\sigma_i} \right]^{\frac{1}{1-\sigma_i}} \tag{3}$$

where N_{isr} is the number of varieties of product i in region r sourced from s. Since each firm produces a unique product, N_{isr} also represents the number of active firms in industry i that sells from s to r. Equation (3) show that utility is an increasing function of the number of varieties sourced from s (N_{isr}). This implies that consumers can obtain a higher utility level when there are more varieties available. The love-of-variety is also reflected in the CES price index which is a decreasing function of N_{isr}. This implies that given constant prices, consumers spend less to attain the same level of utility as the number of varieties available for consumption increases.

Consumer utility maximization yields demand in market r for average product i sourced from s as:

$$\widetilde{Q}_{isr} = Q_{ir} \left[\frac{P_{ir}}{\widetilde{P}_{isr}} \right]^{\sigma_i} \tag{4}$$

Aggregate derived demand in r for all varieties of i sourced from s depends on the demand for an average product and the number of available varieties in the market, $Q_{isr} = N_{isr} \widetilde{Q}_{isr}$. Using Equation (4) we obtain the aggregate derived demand of the $s - r$ link as:

$$Q_{isr} = N_{isr} Q_{ir} \left[\frac{P_{ir}}{\widetilde{P}_{isr}} \right]^{\sigma_i} \tag{5}$$

According to Equation (5), demand increases with the availability of more varieties for consumption (N_{isr}) and larger market size (Q_{ir}), and decreases with higher prices ($\widetilde{P}_{isr}$) relative to import price index (P_{ir}). Therefore, consumer utility increases with the availability of more varieties, even when there is no change in price or consumption of an average variety.

We differentiate Equation (5) to obtain a linear form representation. This linearization yields the percentage change in demand for the differentiated product i

[2] The conversion of equations from their continuous to discrete forms follows from Balistreri and Rutherford (2013).

produced in region s and sold in region r as follows:

$$q_{isr} = n_{isr} + q_{ir} - \sigma_i[\tilde{p}_{isr} - p_{ir}] \tag{6}$$

Note that we use lowercase letters to denote percentage changes in the corresponding variables that are denoted by uppercase letters through the text. Equation (6) determines the derived demand in region r for product i sourced from s, q_{isr}. This refers to products sourced within the region when $s = r$, including both the domestically produced goods and intra-regional imports[3]. On the other hand, Equation (6) determines the demand for inter-regional imports when $s \neq r$.

Based on Equation (6), the change in consumer demand is decomposed into three parts. The first is the variety effect, n_{irs}. Consumer demand increases with the availability of new unique varieties. Given the same level of consumption, the sub-utility derived from consuming a differentiated product rises as the number of varieties sourced from r for consumption in s increases. The second is the expansion effect which is similar to the standard model. Higher aggregate demand in the destination market raises the demand for each source region's product. The last component is the substitution effect. This is also similar to the standard model. The substitution effect is the product of the constant elasticity of substitution amongst varieties, and the percentage change in the ratio of average price of the product relative to consumer's unit expenditure. Demand for a particular source country product increases as the product becomes cheaper relative to the composite price in the destination market.

The demand system in the GTAP model is composed of three agents: private household, government, and firms. In this paper, we focus on private household and firm demands to relate the linearized equations with the code. The rest of the agent-specific equations are similar in the model.

The application of Equation (6) to private households is implemented in the code as:

```
Equation PHLDSRCDF
# private household demand in r for differentiated commodity i sourced from s #
(all,i,MCOMP_COMM)(all,s,REG)(all,r,REG)
      qpmc(i,s,r)
            = - ams(i,s,r) + qp(i,r) + vp(i,s,r)
            - SIGMA(i) * [pps(i,s,r) - ams(i,s,r) - pp(i,r)];
```

where `qpmc(i,s,r)` is the aggregate demand of the private household in market r for differentiated product i that is sourced from region s, `qp(i,r)` is the private household demand in r for product i, `vp(i,s,r)` is the variety index for the private household, `pps(i,s,r)` is the private household price in r for product i that is sourced from region s, `pp(i,r)` is the private consumption price for product i in

[3] Sourced value flows in the data base are transformed such that consumer purchases include domestic as well as intra-regional products when $s = r$. This procedure is explained in more detail in Appendix B.

region r, `ams(i,s,r)` is the import-augmenting technical change parameter, and `SIGMA(i)` is the elasticity of substitution amongst varieties.

Equation `PHLDSRCDF` is similar to the private household demand equation in the standard GTAP model except for the variety effect, which is represented in the code as:

```
Equation PHLDVARIN
# private household variety index #
(all,i,MCOMP_COMM)(all,s,REG)(all,r,REG)
      vp(i,s,r) = ns(i,s,r) + vpslack(i,s,r);
```

where `ns(i,s,r)` is the percentage change in the number of varieties of i shipped from region s to r. This equation shows that the variety index is proportional to the number of sourced varieties. Note that `vpslack(i,s,r)` is an exogenous slack variable. The presence of `vpslack(i,s,r)` allows us to alter the demand structure. This is further discussed in Section 5.

The linearized representation of the composite price index in Equation (3) is given by:

$$p_{ir} = \sum_s \theta_{isr} \widetilde{p}_{isr} - \frac{1}{\sigma_i - 1} \sum_s \theta_{isr} n_{isr} \tag{7}$$

where θ_{isr} is the expenditure share of differentiated product i originating from source s in total expenditure of all varieties from all sources in region r.

$$\theta_{isr} = \frac{N_{isr} \widetilde{Q}_{isr} \widetilde{P}_{isr}}{Q_{ir} P_{ir}}, \tag{8}$$

For private households, Equation (7) and Formula (8) are implemented in the code as:

```
Equation PCOMPRICEMC
# private hhld price index for differentiated commodity i #
(all,i,MCOMP_COMM)(all,r,REG)
      pp(i,r)
            = sum(s,REG, PTHETA(i,s,r) * [pps(i,s,r) - ams(i,s,r)])
            - [1/(SIGMA(i) - 1)] * sum(s,REG, PTHETA(i,s,r) * vp(i,s,r));

Coefficient (all,i,MCOMP_COMM)(all,s,REG)(all,r,REG)
      PTHETA(i,s,r) # shr of demand for i sourced from s in private hhld exp #;
Formula (all,i,MCOMP_COMM)(all,s,REG)(all,r,REG)
      PTHETA(i,s,r) = VPAS(i,s,r) / sum(k,REG, VPAS(i,k,r));
```

where `PTHETA(i,s,r)` is the expenditure share of private households in market r for differentiated product i sourced from s and *VPAS(i,s,r)* is the value of private household expenditure in region r at agent's price by source s.

The private household price index for differentiated product depends on two components. The first component is the average private household price for differentiated composite commodity, `pps(i,s,r)`, weighted by the budget share of differentiated products from source s, `PTHETA(i,s,r)`. This is similar to the standard private household price index except for the data sourcing by agent and the average price.

The second component is the impact of available varieties for consumption in region *r* sourced from *s*, `ns(i,s,r)`, on the price index. This is represented by the private household's variety index, `vp(i,s,r)`. The variety component of the price index rises with an increase in the number of varieties originating from *s* or with an increase in the expenditure share on products originating from *s*. This log-differential form clearly shows that as the number of varieties increases, the same level of utility can be attained at a lower expenditure level. The expenditure shares reflect consumer preference for products sourced from region *s*. The shares typically show a consumer preference for domestic varieties over imported varieties.

3.1.2 Firm demand

The same modifications in derived demand and price index equations apply to all agents in the economy, i.e. private household, government, and firms. In the case of firms, Equation (6) represents the intermediate input cost of producing one unit of output. An important distinction with firms is that there is one more dimension to consider in derived demand equations for intermediate inputs. This additional dimension increases the data arrays relating to firms' purchases from three to four dimensions.

A point to note here is that the specification of firms' derived demand equation depends on the market structure of the input. What matters for firm demand is the nature of the intermediate input, not the nature of the industry that demands the input. For instance, if a firm in the perfectly competitive industry demands differentiated intermediate inputs, then its derived demand equation incorporates *love-of-variety*. Conversely, if a firm in the monopolistically competitive industry demands homogeneous intermediate inputs, then the derived demand equation has the standard form with Armington assumption and composite import commodity formed at the border.

Intermediate input demand for differentiated products, `qfmc(i,j,s,r)`, is governed by equation `INDSRCDF` below, which is an application of Equation (6):

```
Equation INDSRCDF
# industry j's demand in r for differentiated commodity i sourced from s #
(all,i,MCOMP_COMM)(all,j,PROD_COMM)(all,s,REG)(all,r,REG)
      qfmc(i,j,s,r)
            = - ams(i,s,r) + qf(i,j,r) + vf(i,s,r)
            - SIGMA(i) * [pfs(i,j,s,r) - ams(i,s,r) - pf(i,j,r)];
```

where `qf(i,j,r)` is the demand for the differentiated product *i* for use by industry *j* in region *r*, `vf(i,s,r)` is the firm variety index of differentiated product *i* shipped from source *s* to destination *r*, `pfs(i,j,s,r)` is firms' price in industry *j* of region *r* for the differentiated product *i* sourced from *s*, `pf(i,j,r)` is firms' price for differentiated product *i* for use by industry *j* in region *r*.

Dual to this is firm's price index, `pf(i,j,r)`, which is governed by Equation `ICOMPRICEMC`, an application of Equation (7):

```
Equation ICOMPRICEMC
# industry j's price index for differentiated commodity i in region r #
(all,i,MCOMP_COMM)(all,j,PROD_COMM)(all,r,REG)
      pf(i,j,r)
            = sum(s,REG, FTHETA(i,j,s,r) * [pfs(i,j,s,r) - ams(i,s,r)])
            - [1 / (SIGMA(i) - 1)] * sum(s,REG, FTHETA(i,j,s,r) * vf(i,s,r));
```

where `FTHETA(i,j,s,r)` is the expenditure share for differentiated product *i* sourced from *s* of firms' total purchases in industry *j* of region *r*. The interpretation of these equations is analogous to that of the private household demand and price index. Firms, too, benefit from the sheer availability of differentiated varieties of intermediate inputs. The cost of producing one unit of output decreases in the face of increased number of varieties, at constant prices. This has important implications for trade in intermediate inputs which is discussed in Section 6.

3.2 Production

This section introduces the production technology in the firm heterogeneity module. Similar to the standard GTAP model, production in the monopolistically competitive industry is modeled based on a nested structure. This is laid out in Figure 1. Panel A in Figure 1 shows the modeling of fixed costs and Panel B shows the production tree. The solid lines specify CES nests, while the dashed lines distinguish between the respective types of the top level variables. The difference will be clarified as we explore the production structure further below. The key characteristic that distinguishes the production technology in this industry from a perfectly competitive one is the difference between the variable and fixed components of costs. In this model, we assume that there are fixed set-up costs associated with industry entry and fixed trading costs associated with entry into bilateral markets.

Following Swaminathan and Hertel (1996), we assume that fixed costs are attributed solely to non-traded primary factors. Thus, we assume that a part of the value-added inputs of heterogeneous firms is devoted to cover fixed costs. Intermediate inputs are not used in this process. As discussed earlier, firms invest in research and development as well as advertising and distribution in order to differentiate their varieties for domestic and export markets. They learn about the rules and regulations on shipping, packaging, and labeling specific to each market they plan to supply. They adapt their production lines to ensure that their products are in line with the market regulations. Each of these activities require the employment of labor and capital.

Fixed set-up costs are one time only investments made prior to entry into the industry to develop the product and set up its initial production line. Particularly, the equipment used in the research and development laboratory is considered as capital, while the firm hires labor to advertise their products in foreign markets to inform new consumers.

Due to the distinction between fixed and variable costs, total value-added composite, `qva(j,r)` in Figure 1, Panel A, has two components: variable value-added,

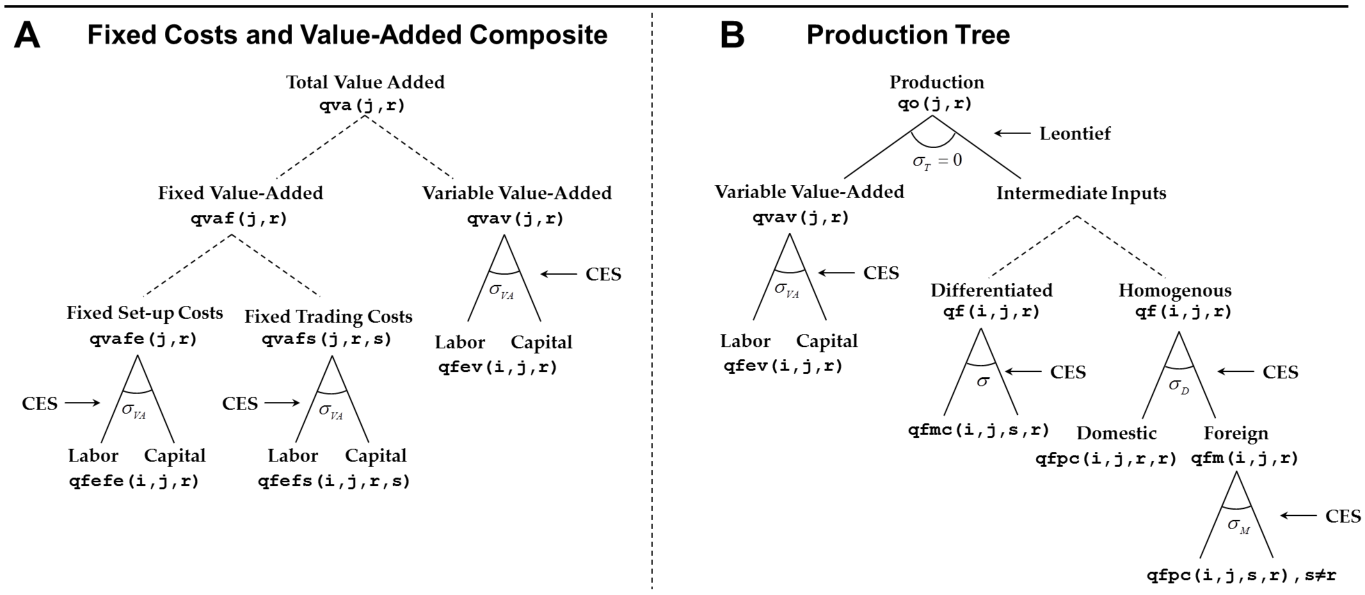

Figure 1. Production structure in the monopolistically competitive industry with firm heterogeneity.

Notes: `i` ∈ `ENDW_COMM` for endowment inputs, `i` ∈ `TRAD_COMM` for intermediate inputs, `j` ∈ `MCOMP_COMM` for monopolistically competitive industry, and `r,s` ∈ `REG` for regions.

Source: Author calculations.

`qvav(j,r)`, and fixed value-added, `qvaf(j,s)`. Variable value-added is used in the production of the differentiated product and therefore is proportional to output such that demand for variable value-added increases as firms expand production. On the other hand, fixed value-added is incurred only once and is invariant to how much the firm produces.

The fixed value-added is further split into set-up and trading components based on whether the primary factors are employed to cover fixed set-up costs, `qvafe (j,r)`, or region-specific fixed trading costs, `qvafs(j,r,s)`. This is shown at the bottom level of the tree in Figure 1 where both set-up and trading components of fixed value-added are produced by labor and capital according to a CES technology. Note that fixed trading costs are intra-regional when $s = r$, while they are inter-regional when $s \neq r$. Further details on how `qvafe(j,r)` and `qvafs(j,r,s)` are determined in the model can be found in Appendix C.4.

We adopt the assumption that the labor/capital intensity in fixed and variable value-added composites is the same (Swaminathan and Hertel, 1996). As a result, the substitution elasticity between primary factors, σ_{VA} (`ESUBVA(j)` in GTAP), is identical in each value-added nest. This simplifying assumption is based on the data availability pertaining to the composition of fixed costs as opposed to variable value-added.

Under certain conditions it can be more appropriate to consider research and development as more capital intensive and marketing as more labor intensive compared to production. In that case it becomes necessary to allow for varying labor/-capital intensity across different components of the value-added composite. While this can be achieved in the current model with only minor modifications, it also requires industry-specific information that is currently not available in our data base. Therefore, we restrict ourselves to the assumption of equal intensities in this study.

The set-up and trading components determine the fixed value-added composite based on their respective weights in total fixed costs. The total value-added bundle is then determined as a share-weighted aggregation of the fixed and variable components.

According to Panel B in Figure 1, output is produced by a combination of variable value-added and intermediate input composites at the top level of the production tree depending on a constant returns to scale technology. We should emphasize that the assumption of constant returns to scale technology in combining variable inputs does not mean that we abstract from potential scale economies. The existence of fixed costs generate internal increasing returns to scale in sales as firms expand production. Firms take advantage of falling average costs when they operate at a larger scale since each additional input brings about a more than proportional increase in output when fixed costs are present.

Intermediate input demand, `qf(i,j,r)`, is given in the lower nest of Panel B in Figure 1. Firms in industry j are supplied either with differentiated products or with homogeneous products. As discussed before, the specification of the derived demand equation depends on the nature of the intermediate input, not the nature of the industry demanding it. In particular, demand for differentiated intermediate inputs, `qfmc(i,j,s,r)`, incorporates the effect of *love-of-variety* regardless of the market structure in industry *j*. Demand for homogeneous intermediate inputs, `qfpc(i,j,s,r)`, follows from the Armington nests. Note that unlike homogeneous intermediate inputs, there is no domestic versus imports distinction for differentiated intermediate inputs. Imported varieties are assumed to compete directly with domestic varieties at the market based on the associated substitution elasticity, σ. That is why there are no additional nests for `qfmc(i,j,s,r)`.

3.2.1 Productivity distribution

There are N^p_{ir} firms in industry *i* of region *r* that pay the fixed set-up costs to participate in the productivity draw. Prior to the draw, each of these firms has the potential to be a producer in the industry. Although there is no particular productivity distribution in the original Melitz (2003) model, the convention in the subsequent literature is to assume Pareto distribution. This preference is based on the analytical tractability of the Pareto distribution (Chaney, 2008) and its empirical fit for the observed size distribution of firms (Axtell, 2001; Eaton, Kortum, and Kramarz, 2011).

We assume that firms draw their productivity, Φ, from a Pareto distribution with scale parameter Φ_{min}, shape parameter γ, and support $[\Phi_{min}, \infty)$. The associated probability density function, $g(\Phi)$, and the cumulative distribution function, $G(\Phi)$, are expressed as:

$$g(\Phi) = \frac{\gamma}{\Phi}\left(\frac{\Phi_{min}}{\Phi}\right)^{\gamma},\ G(\Phi) = 1 - \left(\frac{\Phi_{min}}{\Phi}\right)^{\gamma} \tag{9}$$

where $\Phi_{min} \in [1, \infty)$ is assumed. The shape parameter provides information on the dispersion of firm productivity within the industry and how this dispersion translates into price differences across firms. Note that γ is an inverse measure of heterogeneity. In particular, productivity is less dispersed in an industry that is characterized by a higher shape parameter which implies that the price charged by a new entrant will be similar to that of incumbents. In such an industry, inefficient firms account for a larger share of the industry output. Conversely, productivity is more dispersed in industries with lower shape parameters where more efficient firms represent a larger share of overall industry output. In this case new entrants are highly inefficient compared to the incumbents and charge relatively higher prices.

As discussed in Melitz (2003), the $(\sigma - 1)^{th}$ uncentered moment of $g(\Phi)$ must be finite in order to obtain a finite average productivity level in the industry. This condition imposes a parametric restriction on the size of the upper tail of the distribution in order for the model to be well-defined. When the productivity distribution takes on the Pareto density function, the parametric restriction reduces to $\gamma_i > \sigma_i - 1$. Note that under this parametric restriction, an industry with a high productivity heterogeneity (low γ_i) cannot be characterized by homogeneous preferences (high σ_i). Therefore, relative values of γ_i and σ_i become critical for quantitative outcomes such as export sales and welfare changes (Akgul, Villoria, and Hertel, 2015) as well as for calibration of fixed trading costs (see Section 4.2).

3.2.2 Markup pricing

Under perfect competition, identical firms operate at constant returns to scale and produce homogeneous products. Since firms are price takers in competitive industries, firm prices are equal to their marginal costs. However, heterogeneous firms in monopolistically competitive industries produce differentiated products which gives them a degree of monopoly power over their products. Since they are price setters for their unique products, they can afford to set a price above their marginal costs. The profit maximizing price for such firms is to charge a constant markup over their marginal costs.

Let Φ indicate the level of firm productivity which measures the number of units of output produced by one bundle of input and C_{ir} indicate the cost of the input bundle that is used for producing one unit of output in industry i of region r. By this definition, the marginal cost of a firm with productivity Φ in industry i of region r equals $\frac{C_{ir}}{\Phi}$. As reflected by this fraction, marginal cost increases with factor prices and decreases with productivity. Given bundle costs C_{ir}, the marginal cost of a high-productivity firm (higher Φ) is low, while that of a low-productivity firm (lower Φ) is high within the same industry.

The markup pricing equation translates differences in marginal costs into differences in prices. The price charged in market s by the firm with productivity Φ, in industry i of region r is given by $P_{irs}(\Phi)$ as follows:

$$P_{irs}(\Phi) = \frac{\sigma_i}{\sigma_i - 1}\frac{C_{ir}T_{irs}}{\Phi} \tag{10}$$

where $\frac{\sigma_i}{\sigma_i-1}$ is the constant markup ratio in industry i, T_{irs} is the cost associated with trade, transportation[4], and taxes, and $P_{irs}(\Phi) = P_{ir}(\Phi)T_{irs}$, is the price gross of trade and transportation costs. Equation (10) shows that the sales price is higher than the marginal cost by the amount of the markup in the industry[5]. Firms with higher productivity levels set lower price levels, receive higher markups, produce more and therefore earn higher profits compared to low-productivity firms.

Using Equation (10), the ratio of prices and outputs of any two firms can be

---

[4] Note that we follow the standard GTAP model for the treatment of international transportation margins. In particular, they are calculated as the wedge between the value of exports at fob prices by destination and the value of imports at cif prices by source. The effect of these transportation costs on prices enters the model via the price linkage.

[5] To clarify, markup is defined as the difference between firm price and marginal cost, while markup ratio is defined as the ratio of firm price to marginal cost.

written as the ratio of their respective productivity levels:

$$\frac{P_{irs}(\Phi_1)}{P_{irs}(\Phi_2)} = \frac{\Phi_2}{\Phi_1}, \ \frac{Q_{irs}(\Phi_1)}{Q_{irs}(\Phi_2)} = \left[\frac{\Phi_1}{\Phi_2}\right]^{\sigma_i} \tag{11}$$

To implement Equation (10) in the GTAP model, we use producer price charged by the average firm in the industry, $\widetilde{P}_{ir}$,

$$\widetilde{P}_{ir} = \frac{\sigma_i}{\sigma_i - 1}\frac{C_{ir}}{\widetilde{\Phi}_{ir}} \tag{12}$$

where $\widetilde{P}_{ir} = \frac{\widetilde{P}_{irs}}{T_{irs}}$, $\widetilde{\Phi}_{ir}$ is the average productivity in industry i of region r. Equation (12) is at the industry level and summarizes the firm level information in Equation (10). Total differentiation of Equation (12) yields:

$$p_{ir} = c_{ir} - \widetilde{\varphi}_{ir} \tag{13}$$

According to Equation (13), changes in the producer price is directly proportional to changes in the marginal cost at constant markup. Note that marginal cost is equal to the average variable cost in the monopolistically competitive industry. Therefore, average variable cost is determined by Equation (13), as well.

Equation (13) is implemented in the code as follows:

```
Equation MKUPRICE
# markup pricing (with constant markup) in the monop. comp. ind. j in r #
(all,j,MCOMP_COMM)(all,r,REG)
      ps(j,r) = avc(j,r) + mkupslack(j,r);
```

where `ps(j,r)` is the percentage change in the price received by the firm in the monopolistically competitive industry *j* in region *r*, `avc(j,r)` is the percentage change in the average cost of production in industry *j* in region *r*, and `mkupslack (j,r)` is a slack variable. The presence of mkupslack allows us to eliminate the mark-up pricing rule for any sector *j* in any region *r* if the market structure is not monopolistically competitive. The use of slack variables is discussed in Section 5.

The equation that governs `avc(j,r)` is implemented in the code as follows:

```
Equation AVERAGEVC
# average variable cost of production in the monop. comp. industry j in r #
(all,j,MCOMP_COMM)(all,r,REG)
      avc(j,r)
            = sum(i,TRAD_COMM, SVC(i,j,r) * [pf(i,j,r) - af(i,j,r)])
            + SVAV(j,r) * [pvav(j,r) - avav(j,r)] - ao(j,r);
```

where `SVC(i,j,r)` is the share of intermediate input *i* in variable costs of industry *j*, `SVAV(j,r)` is the share of variable value-added in variable costs of industry *j*, `pf(i,j,r)` is the price of intermediate input *i* employed in industry *j* of region *r*, `pvav(j,r)` is the price of variable value-added composite in industry *j* of region *r*, and `af(i,j,r)` and `avav(j,r)` are input augmenting technical change variables that capture the efficiency of inputs used in production[6]. The first two components on the right-hand side give the percentage change in the input bundle cost, corresponding to c_{ir} in Equation (13). `ao(j,r)` corresponds to the percentage change in the average industry productivity, $\widetilde{\varphi}_{ir}$ in Equation (13). The variable `ao(j,r)` is key in the GTAP firm-heterogeneity model and we will discuss it in further detail in Section 3.2.4.

Average variable cost rises with intermediate input prices, `pf(i,j,r)`, value-added prices, `pva(j,r)`, and with the share of inputs used in production, while it decreases with average industry productivity, `ao(j,r)`, and increased efficiency of inputs, `af(i,j,r)` and `avav(j,r)`.

3.2.3 Productivity threshold for market entry

Firm participation in a given bilateral market depends on the potential for making nonnegative profits. Each firm in industry i of region r with productivity draw Φ faces fixed and variable costs to operate in market s and makes profit $\Pi_{irs}(\Phi)$:

$$\Pi_{irs}(\Phi) = \frac{P_{irs}(\Phi)}{T_{irs}} Q_{irs}(\Phi) - \frac{C_{ir}}{\Phi} Q_{irs}(\Phi) - W_{irs} F_{irs} \tag{14}$$

for all regions r, s where W_{irs} is the price of fixed value-added and F_{irs} is the demand for value-added required for fixed costs of operating on the $r - s$ bilateral trade route. Firm profit increases with market size (Q_{irs}), productivity level (Φ), lower costs (C_{ir}), and lower barriers to trade (T_{irs}, F_{irs}). Based on Equation (14), firms with higher productivity levels charge lower prices with higher markups, make more sales, and therefore incur larger profits relative to low-productivity firms.

The existence of productivity heterogeneity and destination-specific fixed trading costs imply that there is a minimum level of productivity required for market entry as not all firms are able to cover their fixed costs. Let Φ^*_{irs} denote the minimum level of productivity required for a firm in industry i of region r to be active on market s. This trade route is profitable only for firms that are productive enough to make nonzero profits. This corresponds to the firms in the upper tail of the distribution with productivity levels on or above the threshold Φ^*_{irs}.

The threshold is determined by the marginal firm, whose productivity draw equals Φ^*_{irs} such that its variable profit is just enough to cover its fixed trading costs. Therefore, the marginal firm makes zero economic profit in market s, $\Pi_{irs}(\Phi^*_{irs}) = 0$. In order to find the productivity threshold for bilateral markets, we solve the zero profit condition of the marginal firm for Φ^*_{irs} by using the optimal demand and price of the marginal firm (see Appendix C.1 for the derivation). This yields

[6] For example, an increase in `avav(j,r)` implies that variable value-added becomes more efficient such that the same amount of output j can be produced by fewer variable value-added composite. The interpretation of changes in `af(i,j,r)` is analogous.

the productivity threshold for market entry as:

$$\Phi^{*}_{irs} = \frac{\sigma_i^{\frac{\sigma_i}{\sigma_i - 1}}}{\sigma_i - 1} \frac{C_{ir}}{P^{*}_{ir}} \left[\frac{P^{*}_{irs}}{T_{irs} W_{irs}} \frac{Q^{*}_{irs}}{F_{irs}} \right]^{\frac{1}{1-\sigma_i}} \tag{15}$$

where variables with superscript '*' are associated with the marginal firm. Based on Equation (15), the threshold depends on sales revenue ($P_{irs}Q_{irs}$), barriers to trade (T_{irs}, F_{irs}) and variable costs (C_{ir}). A firm with productivity draw above the threshold, $\Phi > \Phi^{*}_{irs}$, incurs positive profits and self-selects into the market. On the other hand, a firm with productivity draw below the threshold, $\Phi < \Phi^{*}_{irs}$, incurs negative profits and drops out of the market.

Market size reduces the threshold level of productivity required for firm entry by providing larger scale economies. With access to a larger market, the possibility of making positive profits increases since fixed costs are spread over larger sales. In addition, decreases in barriers to trade such as lower tariffs, transportation costs, or fixed costs also lead to lower productivity thresholds by increasing the potential for positive profits. As the productivity threshold for market entry decreases, firms in the lower portion of the productivity distribution will be able to compete with the existing firms. As new firms enter, the number of active firms in the market increases.

On the other hand, the productivity threshold for market entry is higher in industries with homogeneous preferences, i.e. high σ_i. In that case, markups are low and firms charge competitive prices. Since firm varieties are similar, price differences will determine which firm will likely to make a sale. Efficiency becomes advantageous in this case because consumers choose the lowest priced products when preferences are homogeneous. This implies that only the high-productivity firms will survive in the market.

It is important to note that the right-hand side of Equation (15) is defined in terms of the prices and sales of the marginal firm. Since information on the marginal firm is not readily available, we focus on the average firm instead. We will express the productivity threshold equation by using the prices and sales of the average firm rather than the marginal firm. This is achieved by relating marginal firm behavior to average firm behavior using the aggregation method in Melitz (2003).

Average productivity in a bilateral market is determined by productivity levels of active firms that make the cut, $\Phi > \Phi^{*}_{irs}$. The probability of a firm being active in a market is $1 - G(\Phi^{*}_{irs}) = \frac{N_{irs}}{N^{p}_{ir}}$. In order to find the average productivity in a bilateral market we aggregate productivity levels across all active firms in that market using a CES functional form. This CES aggregation is best explained by following the average firm definition in Dixon, Jerie, and Rimmer (2015). Similar to Dixon, Jerie, and Rimmer (2015) we define $\widetilde{\Phi}_{irs}$ as the average productivity of a firm in industry i that operates on the $r - s$ trade route and $\frac{\widetilde{Q}_{irs}}{\widetilde{\Phi}_{irs}}$ as the average amount of input composite required by the average firm to produce one unit of output. The average input composite is determined by:

$$\frac{\widetilde{Q}_{irs}}{\widetilde{\Phi}_{irs}} = \frac{1}{N_{irs}}\left[\int_{\Phi^*_{irs}}^{\infty} \frac{Q_{irs}(\Phi)}{\Phi} N^p_{ir} g(\Phi) d\Phi\right] \tag{16}$$

where $\frac{Q_{irs}(\Phi)}{\Phi}$ gives the amount of input composite for a firm with productivity Φ. We substitute the output ratio implied by Equation (11), $\frac{Q_{irs}(\Phi)}{\widetilde{Q}_{irs}} = \left[\frac{\Phi}{\widetilde{\Phi}_{irs}}\right]^{\sigma_i}$, into Equation (16). The substitution yields:

$$\widetilde{\Phi}^{-1}_{irs} = \widetilde{\Phi}^{-\sigma_i}_{irs} \frac{N^p_{ir}}{N_{irs}} \int_{\Phi^*_{irs}}^{\infty} \Phi^{\sigma_i - 1} g(\Phi) d\Phi \tag{17}$$

Rearranging Equation (17) to solve for $\widetilde{\Phi}_{irs}$ and substituting $\frac{N_{irs}}{N^p_{ir}} = 1 - G(\Phi^*_{irs})$ yields the CES weighted average productivity given in Melitz (2003) as follows:

$$\widetilde{\Phi}_{irs} = \left[\frac{1}{1 - G(\Phi^*_{irs})} \int_{\Phi^*_{irs}}^{\infty} \Phi^{\sigma_i - 1} g(\Phi) d\Phi\right]^{\frac{1}{\sigma_i - 1}} \tag{18}$$

where the weights reflect the relative output shares of firms with different productivity levels.

To obtain a closed form solution we use the probability density of Pareto distribution. Applying the Pareto distribution, Equation (18) reduces to:

$$\widetilde{\Phi}_{irs} = \Phi^*_{irs}\left[\frac{\gamma_i}{\gamma_i - \sigma_i + 1}\right]^{\frac{1}{\sigma_i - 1}} \tag{19}$$

where $\gamma_i > \sigma_i - 1$. Since the parameters are constant, Equation (19) shows that average productivity is proportional to marginal productivity. This proportionality implies a similar relationship between the prices and sales of the average firm and the marginal firm. Applying this relationship into Equation (15), we define the productivity threshold for market entry in terms of the average firm as follows:

$$\Phi^*_{irs} = \frac{\sigma_i^{\frac{\sigma_i}{\sigma_i - 1}}}{\sigma_i - 1} \frac{C_{ir}}{\widetilde{P}_{ir}} \left[\frac{\widetilde{P}_{irs}}{T_{irs} W_{irs}} \frac{\widetilde{Q}_{irs}}{F_{irs}}\right]^{\frac{1}{1 - \sigma_i}} \tag{20}$$

Now, we can incorporate endogenous bilateral thresholds into the model by total differentiation of Equation (20):

$$\varphi^*_{irs} = c_{ir} - \widetilde{p}_{ir} + \frac{1}{\sigma_i - 1}(f_{irs} - \widetilde{q}_{irs}) + \frac{1}{\sigma_i - 1}(w_{irs} - \widetilde{p}_{irs} + t_{irs}) \tag{21}$$

There are two factors at play in Equation (21). The first is the competition effect (first two terms on the right-hand side). It is the result of changes in costs (c_{ir}) relative to the supplier price ($\widetilde{p}_{ir}$). An increase in industry variable cost results in a loss in competitiveness against more efficient firms and makes it more costly for the firm to enter a new market. Therefore, productivity threshold increases with costs. This increase may be offset by the second component which is the market access effect (last two terms on the right-hand side). In trade liberalization scenarios, as markets integrate, firms gain access to larger markets (positive $\widetilde{q}_{irs}$). This reduces fixed trading costs per sale and increases potential for positive profits. As a result, the productivity threshold declines with larger market access.

Equation (21) is implemented in the code as:

```
Equation PRODTRESHOLD
# productivity threshold in industry j of region r to enter market s#
(all,j,MCOMP_COMM)(all,r,REG)(all,s,REG)
      aost(j,r,s)
            = sum(i,TRAD_COMM, SVC(i,j,r) * [pf(i,j,r) - af(i,j,r)])
            + SVAV(j,r) * [pvav(j,r) - avav(j,r)] - ps(j,r)
            + [MARKUP(j,r) - 1] * [qvafs(j,r,s) - qs(j,r,s)]
            + [MARKUP(j,r) - 1]
              * [pvafs(j,r,s) - pfob(j,r,s) - tx(j,r) - txs(j,r,s) - to(j,r)]
            + threshslack(j,r,s);
```

where `aost(j,r,s)` is the productivity threshold for market entry. The first three components on the right-hand side of Equation `PRODTRESHOLD` are as discussed in Section 3.2.2. Note that `qvafs(j,r,s)` is the demand for the value-added composite used in fixed trading costs, `qs(j,r,s)` is the amount of sales on the $r-s$ trade route, `pvafs(j,r,s)` is the price of value-added composite used in fixed trading costs, `pfob(j,r,s)` is the fob price, `tx(j,r)` is the destination generic tax/subsidy, `txs(j,r,s)` is the destination/source-specific tax/subsidy, `to(j,r)` is the output tax/subsidy, `MARKUP(j,r)` is the markup ratio $\frac{\sigma_i}{\sigma_i - 1}$, and finally `threshslack(j,r,s)` is a slack variable. The presence of `threshslack (j,r,s)` allows us to eliminate the productivity threshold for entry into any market if, for example, firms in the industry are identical or if there are no fixed trading costs to be active on that particular trade route.

Note that when $r = s$, `aost(j,r,s)` shows the productivity threshold for the domestic and intra-regional markets. If there are barriers to intra-regional trade, they will be captured by tax instruments and destination-specific fixed trading costs in Equation `PRODTRESHOLD`.

Given the productivity threshold `aost(j,r,s)`, the average productivity level of each bilateral trade route is implemented in the code using the proportionality between the threshold and average productivity, $\widetilde{\varphi}_{irs} = \varphi^*_{irs}$. This is implemented as:

```
Equation AVEPROD
# average productivity in industry j of region r to enter market s#
(all,j,MCOMP_COMM)(all,r,REG)(all,s,REG)
aos(j,r,s) = aost(j,r,s);
```

where `aos(j,r,s)` is the average productivity of firms that sell product j from the source region r to the destination market s. While equation `AVEPROD` is not necessary in terms of model computation, it is useful to demonstrate the link between changes in productivity thresholds and average productivities on the $r-s$

trade route in a transparent manner.

3.2.4 Average industry productivity

In the GTAP model, industry productivity is determined by the variable `AO(i,r)`, which refers to Hicks-neutral technical change (Hertel, 1997). `AO(i,r)` enters the top-level nest of the production function, where its increase uniformly reduces the input requirements to produce a given level of output. Under firm heterogeneity, we introduce a governing equation in the model for `AO(i,r)` to be endogenously determined via the endogenous changes in productivity thresholds. Therefore, we can trace the transparent mechanism of how trade leads to changes in average industry productivity even when there is no change in the industry's production technology. This has important implications in the model. Since `AO(i,r)` is exogenous in the standard model, its effect on the derived demand for intermediate inputs and value-added composite in production is zero. However, in the firm-heterogeneity model we can see the feedback of endogenous productivity changes on the economy such as the effect on input requirements in production, average variable costs, and average total cost.

In order to allow for endogenous productivity, we begin with the average productivity levels of domestic producers and exporters, $\widetilde{\Phi}_{irs}$, as defined in the weighted average function in Equation (18). The compositional changes in the domestic market are reflected in the domestic average $\widetilde{\Phi}_{irr}$, while the compositional changes in export markets are reflected in the export average, $\widetilde{\Phi}_{irs}$, where $r \neq s$. As discussed above, all producers in the industry sell in the domestic market, whereas only a chosen few among them sell in export markets. It is appealing to think that $\widetilde{\Phi}_{irr}$ summarizes the information on industry productivity as all producers supply the domestic market. However, $\widetilde{\Phi}_{irr}$ only reflects the domestic market share differences across firms. It should be noted that as high-productivity firms also sell in export markets, their export market shares have additional effects on the industry average and should be incorporated in the governing equation. The productivity differences between domestic producers and exporters have important implications for the industry average which are discussed below.

Let $\widetilde{\Phi}_{ir}$ be the weighted average productivity of all active firms in industry i of region r which is determined by:

$$\widetilde{\Phi}_{ir} = \Big[\sum_s \frac{N_{irs}}{\sum_t N_{irt}} \widetilde{\Phi}_{irs}^{\sigma_i - 1} \Big]^{\frac{1}{\sigma_i - 1}} \tag{22}$$

where $\widetilde{\Phi}_{irs}$ is the average productivity level on the $r - s$ bilateral trade route and the weights, $\frac{N_{irs}}{\sum_t N_{irt}}$, reflect the relative shares of active firms on $r - s$ route to all active firms in the industry. According to Equation (22), average productivity in each trade route contributes to the overall industry efficiency depending on their relative importance in industry output.

Linearization of Equation (22) yields:

$$\widetilde{\varphi}_{ir} = \sum_s \delta_{irs}\, \widetilde{\varphi}_{irs} \; + \; \frac{1}{\sigma_i - 1} \left[\sum_s \delta_{irs} \left(n_{irs} - n^t_{ir} \right) \right] \tag{23}$$

where δ_{irs} is the share of market s in region r's total sales of product i and n^t_{ir} is a firm number index calculated as a share-weighted average of the number of active firms on the $r - s$ trade route. These shares are given by:

$$\delta_{irs} = \frac{N_{irs}}{\sum_t N_{irt}} \left(\frac{\widetilde{\Phi}_{irs}}{\widetilde{\Phi}_{ir}} \right)^{\sigma_i - 1}, n^t_{ir} = \sum_s \frac{N_{irs}}{\sum_t N_{irt}} n_{irs} \tag{24}$$

Some firms in industry i may be actively operating in multiple destination markets. The firm number index takes this into account. As such, it represents the total mass of available varieties in industry i of region r. Due to lack of information on the number of firms, we calibrate $\frac{N_{irs}}{\sum_t N_{irt}}$ using the trade flow information in the data base.

There are two parts to Equation (23). The first part shows a 'within market' effect wherein average industry productivity rises with an increase in average productivity in domestic and export markets. Whereas, the second part shows a 'between markets' effect wherein higher productivity levels of exporters relative to domestic producers become more important in the industry average as exporters take more of the market. An important note here is that the second component equals zero for the domestic market. Since all active firms are assumed to supply the domestic market, n_{irr} is proportional to n^t_{ir}. Hence the difference, $n_{irr} - n^t_{ir}$, will be zero when $r = s$. However, it will be nonzero for exporters.

To illustrate the mechanism in Equation (23), let's assume that there is a reduction in tariffs. In such a scenario, average productivity in the domestic market increases as inefficient firms drop out of the industry due to higher import competition. This pulls up $\widetilde{\Phi}_{irr}$. On the other hand, as new exporters emerge, average productivity in export markets decreases since new exporters are less productive than the existing ones. This pulls down $\widetilde{\Phi}_{irs}$ for $r \neq s$. As such, domestic average raises the industry productivity, while export average reduces it. However, we should also factor in the second-order effects of changing market shares as exporters are becoming more important in the total weight. Thus, an expanding export market has an additional contribution to the industry average since exporters are more productive on average than domestic producers. The final change in the industry productivity depends on which of these effects dominate.

In order to implement Equation (23) into the code, we express market shares, δ_{irs}, in an alternative way. Equation (11) implies that the ratio of firm revenues equals the relative firm productivity levels raised to power $\sigma_i - 1$. When we substitute this into the expression in Equation (24), δ_{irs} becomes:

$$\delta_{irs} = \frac{N_{irs}}{\sum_t N_{irt}} \left(\frac{\widetilde{\Phi}_{irs}}{\widetilde{\Phi}_{ir}} \right)^{\sigma_i - 1} = \frac{N_{irs}}{\sum_t N_{irt}} \frac{\widetilde{Q}_{irs} \widetilde{P}_{irs}}{\widetilde{Q}_{ir} \widetilde{P}_{ir}} \tag{25}$$

This is implemented in the code as:

```
Coefficient (all,i,TRAD_COMM)(all,r,REG)(all,s,REG)
      SHRSMD(i,r,s) # share of sales of i to s in r #;
Formula (all,i,TRAD_COMM)(all,r,REG)(all,s,REG)
      SHRSMD(i,r,s) = VSMD(i,r,s) / VOM(i,r);
```

where δ_{irs} is denoted as `SHRSMD(i,r,s)`, the share of sales of product *i* to market *s* in region *r*, `VSMD(i,r,s)` is the value of sales of *i* from *r* to *s* at market prices, and `VOM(i,r)` is the value of commodity *i* output in region *r* at market prices.

Equation (23) is, then, implemented in the code as:

```
Equation AVEPRODIND
# average productivity in the monop. comp. industry i of region r #
(all,j,MCOMP_COMM)(all,r,REG)
      ao(j,r)
        = SHRSMD(j,r,r) * aos(j,r,r)
        + sum(s,REG, [1 - DELTA(r,s)] * SHRSMD(j,r,s) * aos(j,r,s))
        + [MARKUP(j,r) - 1] * sum(s,REG, SHRSMD(j,r,s) * [ns(j,r,s) - nt(j,r)])
        + aosec(j) + aoreg(r) + aoall(j,r) + prodslack(j,r);
```

where `ao(j,r)` is the percentage change in the average productivity of industry *j* in region *r* (associated with the upper-case variable AO(i,r)), `nt(j,r)` is the firm number index in industry *j* of region *r*, `aosec(j)` is output augmenting technical change in industry *j*, `aoreg(r)` is output augmenting technical change in region *r*, `aoall(j,r)` is output augmenting technical change in industry *j* of region *r*, and `DELTA(r,s)` is the Kronecker delta which equals one when $r = s$. We use Kronecker delta to distinguish between the average productivity in domestic (intra-regional) and export markets. This separation is purely for convenience in decomposing the relative contributions of domestic and export markets in productivity growth. Equation `AVEPRODIND` is useful in explaining the source of productivity gains, whether it is due to the exit of inefficient firms from the industry or the expansion of efficient ones into export markets.

It is important to note that `ao(j,r)` only captures productivity growth due to compositional changes. A positive `ao(j,r)` does not mean that firms are getting more productive. Rather it means that more productive firms constitute a larger part of the market than before. In other words, the expansion in the market share of high-productivity firms improves the efficiency of the industry on average. All other changes in productivity are assumed to be exogenous in this model. For example, there may be cases where total factor productivity in the industry changes due to technology shifts in addition to the trade-induced self-selection of firms. This can be achieved by introducing exogenous productivity shifter variables following a treatment similar to that of the perfectly competitive industry. The variables `aosec(j)`, `aoreg(r)`, and `aoall(j,r)` are included in the equation for this purpose.

3.2.5 Endogenous firm entry and exit

Firms face uncertainty about their productivity levels when making an entry decision into the industry. This uncertainty is removed once they pay the initial fixed set-up costs required for entry. With their productive capacity unknown to them *ex-ante*, they consider all possible outcomes regarding their net profits prior to entry. A low productivity draw would result in a negative net profit, while a high draw would result in a positive one. The expectation of positive profits is why potential entrants pay the fixed set-up costs despite the uncertainty they face in terms of productivity.

Free entry and exit in the monopolistically competitive industry implies that potential for positive profits will attract new firms into the industry resulting in an inflow of entrants. Conversely, negative profits will cause inefficient firms to drop out of the industry. Firm movement continues until there is no further incentive for entry/exit which is achieved when firm expected profits equal zero in equilibrium.

There is an important distinction between firm profits and industry profits when firms are heterogeneous. While industry-level profit is zero, firm-level profit will likely be non-zero depending on the productivity draw. As efficient firms produce more at lower cost, they have the potential to make positive profits in the industry. The higher their productivity level, the higher their profit. However, inefficient firms that cannot produce due to bad productivity draws lose all their initial investments. Note that inputs of these non-producers are not reallocated to other firms. Therefore, at the industry level, positive profits of efficient firms are completely exhausted on the losses incurred by inefficient firms until the zero profits condition of the industry is restored. The total number of potential firms in the industry is determined by this process.

Total industry profit is comprised of profits of active firms that operate in bilateral markets and fixed set-up costs paid by all entrants. Let each firm use H_{ir} units of the value-added for fixed set-up costs to enter industry i of region r, and let N^p_{ir} be the number of potential firms in the industry, i.e. total mass of entrants. These are the firms that pay the fixed set-up cost to participate in the productivity draw. Industry profit is, then, given by:

$$\Pi_{ir} = \sum_s N_{irs}\widetilde{\Pi}_{irs} - N^p_{ir} W_{irr} H_{ir} \tag{26}$$

where $\widetilde{\Pi}_{irs}$ is the profit of the average firm in industry i of region r made from supplying market s. Since there are N_{irs} active firms in that market, we simply multiply the profit of an average firm, $\widetilde{\Pi}_{irs}$, with N_{irs} to obtain the total profit made in industry i on that route. Note that while only N_{irs} firms are making profits, all of the potential N^p_{ir} firms incur the fixed set-up costs. In equilibrium, the total profit of these N_{irs} firms is fully exhausted on fixed set-up costs paid by all the N^p_{ir} firms that made the initial productivity draw. The zero profit condition is satisfied when total revenue equals total cost in the industry, $\Pi_{ir} = 0$. Equation (26) then becomes:

$$\sum_s \frac{N_{irs}Q_{irs}P_{irs}}{T_{irs}} = \sum_s \frac{N_{irs}Q_{irs}C_{ir}}{\widetilde{\Phi}_{irs}} + \sum_s N_{irs}W_{irs}F_{irs} + N^p_{ir}W_{irr}H_{ir} \tag{27}$$

which determines N^p_{ir}. In GTAP notation Equation (27) can be expressed as:

$$VOA(j,r) = VC(j,r) + \sum_{s \in REG} VAFS(j,r,s) + VAFE(j,r) \tag{28}$$

where `VOA(j,r)` is the value of output in industry j of region r, $VC(j,r)$ is the variable cost in industry j of region r, `VAFS(j,r,s)` is the fixed trading costs for shipping j from r to s, and `VAFE(j,r)` is the fixed set-up costs for entering industry j in region r.

To obtain the zero profit condition in log-linear form, we perform total differentiation of Equation (28). The detailed derivation is provided in Appendix C.2. Total differentiation yields the zero profits condition of the monopolistically competitive industry as:

$$\begin{aligned} VOA(j,r) * ps(j,r) = & \sum_{i \in TRAD\_COMM} VFA(i,j,r) * pf(i,j,r) - af(i,j,r) \\ & + VA(j,r) * pva(j,r) - VAV(j,r) * avav(j,r) \\ & - VC(j,r) * ao(j,r) \\ & + SHAPE(j) * VAF(j,r) * aost(j,r,r) \\ & - SHAPE(j) * \sum_{s \in REG} VAFS(j,r,s) * aost(j,r,s) \\ & - VAF(j,r) * qof(j,r) \end{aligned} \tag{29}$$

where `VA(j,r)` is the value of purchases of total value-added, `VAV(j,r)` is the value of purchases of variable value-added composite by industry j in region r, `VFA(i,j,r)` is the value of purchases of intermediate input i demanded in industry j of region r, `VAF(j,r)` is total fixed cost payments, `VAF(j,r)` $=$ `VAFE(j,r)` $+ \sum_{s \in REG}$ `VAFS(j,r,s)`, `pva(j,r)` is the price of value-added, `avav(j,r)` is value-added augmenting technical change, and `qof(j,r)` is the percentage change in per firm output level in industry j of region r. Finally, `SHAPE(j)` is the associated Pareto shape parameter.

Equation (29) shows that changes in average total cost can arise due to changes in input prices, input efficiency, industry productivity, and changes in per firm output (scale). In order to simplify the zero profits equation, we distinguish between the change in average total cost due to input prices and the change in average total cost due to per firm output. Let `scatc(j,r)` be the component of average total cost that is attributable to input prices, input efficiency and industry productivity, at constant firm scale:

```
Equation SCLCONATC
# average total cost at constant scale in ind. j in region r #
(all,j,MCOMP_COMM)(all,r,REG)
      VOA(j,r) * scatc(j,r)
        = sum(i,TRAD_COMM, VFA(i,j,r) * [pf(i,j,r) - af(i,j,r)])
        + VA(j,r) * pva(j,r) - VAV(j,r) * avav(j,r) - VC(j,r) * ao(j,r)
        + SHAPE(j) * VAF(j,r) * aost(j,r,r)
        - SHAPE(j) * sum(s,REG, VAFS(j,r,s) * aost(j,r,s));
```

We can now substitute `scatc(j,r)` into Equation (29) and calculate the zero profits condition in the monopolistically competitive industry as follows:

```
Equation ZEROPROFITSMC
# zero pure profits condition for firms in the monopolistically comp industry #
(all,j,MCOMP_COMM)(all,r,REG)
      VOA(j,r) * ps(j,r)
             = VOA(j,r) * scatc(j,r) - VAF(j,r) * qof(j,r)
             + VOA(j,r) * profitslackmc(j,r);
```

where `profitslackmc(j,r)` is an exogenous slack variable. The presence of `profitslackmc(j,r)` allows us to let the industry earn nonzero profits under certain conditions. For example, if there is no free entry in the industry, then the total number of firms is fixed. In such a case, the industry may make positive profits in the short-run.

Equation `ZEROPROFITSMC` shows that the percentage change in supplier price equals the percentage change in average total cost in equilibrium. One of the important differences between this zero profits condition and that in a perfectly competitive market is the effect of per firm output. In a perfectly competitive market, average total cost equals average variable cost in the absence of fixed costs. However, in a monopolistically competitive market the presence of fixed costs generates a wedge between average total cost and average variable cost. A key assumption in our model is that fixed costs are comprised of primary factors only. This implies that changes in intermediate input use and prices only affect variable costs, not fixed costs. Therefore, a tariff cut, for example, will cause divergence in variable and fixed costs by increasing imports of intermediate inputs. This divergence also creates a wedge between the scale constant average total cost (which includes fixed costs) and average variable costs (which equals producer price). If this wedge is large, then there is a big potential for gains from trade through expanding the scale of the firm and reducing fixed cost per output.

As discussed above, while all N_{ir}^{p} firms participate in the productivity draw, only a subset of them have a high enough productivity level to produce and operate in bilateral markets, $N_{irs} < N_{ir}^{p}$. Who makes the cut is determined by the productivity threshold. Among all the N_{ir}^{p} potential firms in the industry only those firms that pass the threshold level of productivity, $\Phi > \Phi_{irs}^{*}$, are able to operate in a market. Recall that $1 - G(\Phi_{irs}^{*})$ is the proportion of firms that are active on the $r - s$ trade route. Then, N_{irs} is given by $N_{irs} = N_{ir}^{p}[1 - G(\Phi_{irs}^{*})]$. Applying the Pareto distribution, this equation becomes:

$$N_{irs} = N_{ir}^{p}[\Phi_{irs}^{*}]^{-\gamma_i} \tag{30}$$

Equation (30) shows that successful market entry increases with larger mass of potential firms in the industry (N_{ir}^{p}), lower productivity threshold (Φ_{irs}^{*}), and less productivity heterogeneity in the industry (high γ_i). Total differentiation of Equation (30) yields:

$$n_{irs} = n_{ir}^{p} - \gamma_i \varphi_{irs}^{*} \tag{31}$$

The entry/exit mechanism is easily traced out in the linearized model. An increase in the productivity threshold (positive φ_{irs}^{*}) causes firms with lower productivity levels to be less competitive and lose sales. Since they incur negative profits, inefficient firms are forced to drop out of the market (negative n_{irs}). The final change in the number of active firms in a market also depends on the change in total mass of firms in the industry (n_{ir}^{p}).

Endogenous firm entry/exit mechanism in Equation (31) is implemented as follows:

```
Equation NFIRM
# number of active firms in industry i of region r that sell in market s #
(all,i,MCOMP_COMM)(all,r,REG)(all,s,REG)
      ns(i,r,s) = np(i,r) - SHAPE(i) * aost(i,r,s) + entryslack(i,r,s);
```

where `entryslack(i,r,s)` is a slack variable.

The Pareto-shape parameter, `SHAPE(i)`, plays an important role in determining the number of active firms. As discussed above, productivity is less dispersed in industries that are characterized by higher shape parameters. In such industries there is a larger mass of firms around the margin with similar productivity levels. Keeping everything else constant, the same decrease (increase) in the productivity threshold will result in the entry (exit) of a larger mass of firms in an industry with a higher shape parameter compared to the one with a lower shape parameter. In other words, shape parameter magnifies the importance of the threshold in firm activity in a market.

3.2.6 Firm scale

Each firm in the monopolistically competitive industry produces a different level of output with high-productivity firms producing more, while low-productivity firms producing less. We can incorporate this firm-level information in the distribution of output in the model by assuming an industry composed of N_{irr} active producers with productivity levels that equal the industry average. Total output in the monopolistically competitive industry, Q_{ir}, is then a product of the average firm output, $\widetilde{Q}_{ir}$, and the number of producers in the industry, N_{irr}, given by:

$$Q_{ir} = N_{irr}\widetilde{Q}_{ir} \tag{32}$$

Equation (32) shows that as inefficient firms exit (lower N_{irr}), the surviving producers expand their scales and produce more, given the industry output. Total differentiation of Equation (32) yields:

$$q_{ir} = n_{irr} + \widetilde{q}_{ir} \tag{33}$$

Equation (33) is implemented in the code as:

```
Equation INDOUTPUT
# total output in the monopolistically competitive industry #
(all,j,MCOMP_COMM)(all,r,REG)
      qo(j,r) = n(j,r,r) + qof(j,r);
```

where `qof(j,r)` is the percentage change in the output per firm in industry j of region r.

Note that changes in output per firm is dictated by changes in costs providing information on the available scale economies. The relationship between output per firm and costs can be expressed by using markup pricing rule (Equation `MKUPPRICE`) and zero profits condition (Equation `ZEROPROFITSMC`). As discussed in section (3.2.5), the zero profits condition in the monopolistically competitive industry is given as follows:

```
VOA(j,r) * ps(j,r) = VOA(j,r) * scatc(j,r) - VAF(j,r) * qof(j,r)
```

where `profitslack(j,r)` in the original equation is assumed to be zero as the industry makes zero profits. Note that the producer price and average variable cost in the monopolistically competitive industry are proportional, `ps(j,r) = avc(j,r)`, as discussed in section (3.2.2). Following Swaminathan and Hertel (1996), we solve the zero profits equation for `qof(j,r)` by substituting in the markup pricing rule. This yields the following relationship between per firm output and relative costs:

```
qof(j,r) = [VOA(j,r)/VAF(j,r)] * [scatc(j,r) - avc(j,r)]
```

Based on this expression the change in per firm output is determined by the change in scale constant average total cost relative to average variable cost. Note that `scatc(j,r)` and `avc(j,r)` diverge due to the presence of fixed costs. As discussed above, fixed costs are included in total costs and can be solely attributable to primary factor payments. Therefore, changes in input prices affect `scatc(j,r)` and `avc(j,r)` at differing rates depending on their relative factor intensities.

3.3 Welfare decomposition

Compared to the standard GTAP model, there are four new channels through which trade liberalization induces welfare changes in the firm heterogeneity model: (i) productivity effect, (ii) love-of-variety effect, (iii) scale effect, and (iv) fixed cost effect.

The productivity effect is a new source of gains from trade generated by the reallocation of factors of production from less productive firms into more productive ones, thereby generating an improvement in the overall efficiency of the industry. This is purely a compositional gain within the industry that arises from expanding market shares of efficient firms.

In firm heterogeneity, consumers are characterized by love-of-variety. As trade growth induces new firms to enter international markets, consumers benefit from the availability of new varieties in the market. This *love-of-variety* effect contributes positively to overall regional welfare. Kancs (2010) states that even though there are lost domestic varieties in the post-tariff cut economy, the empirical findings in the literature show that consumers usually benefit from the trade policy as there are more imported varieties available. However, if we account for the preference bias, we see that the loss of domestic varieties has a bigger impact on welfare as consumers prefer domestically produced varieties to imported varieties based on observed import shares.

The scale effect on regional welfare is associated with increasing returns to scale technology. The presence of fixed costs and imported intermediate inputs create a wedge between scale constant average total cost and average variable cost. Tariff cut leads to increases in imports of cheap intermediate inputs which reduces the average variable cost. As average variable cost decreases relative to the scale constant average total cost, per firm output increases. This trade-induced expansion in firm scale spreads fixed costs across more output, generating significant additional gains from trade.

Finally, the fixed cost effect is the change in welfare due to firms' fixed cost payments. As discussed above, while all the potential firms, N_{ir}^{p}, incur the initial fixed costs, not all of them are able to begin production. Fixed cost payments of inactive firms in the industry reduces regional welfare.

The existing welfare decomposition in the standard GTAP model is based on Huff and Hertel (2000). We extend this decomposition to include the new channels in firm heterogeneity and monopolistic competition. Similar to the derivation in Huff and Hertel (2000), we start with the expression for the change in household income. This is a function of primary factor payments net of depreciation, tax revenues net of subsidies, and the profits accruing to firms in the monopolistically competitive industry. We substitute the equilibrium conditions in the model into the household income equation to obtain the welfare decomposition. In this section, we provide the decomposition expression and discuss the new sources of welfare change in this expression. For the GTAP firm-heterogeneity model, the equivalent variation (EV) decomposition can be expressed as follows[7]:

```
Equation EV_DECOMPOSITION
# decomposition of Equivalent Variation #
(all,r,REG)
EV_ALT(r)
```

[7] Note that the firm-heterogeneity welfare decomposition in this model is based on the assumption of Leontief production function, where `ESUBT(j)` $= 0$.

```
= -[0.01 * UTILELASEV(r) * INCOMEEV(r)]
* [DPARPRIV(r) * loge(UTILPRIVEV(r) / UTILPRIV(r)) * dppriv(r)
+ DPARGOV(r) * loge(UTILGOVEV(r) / UTILGOV(r)) * dpgov(r)
+ DPARSAVE(r) * loge(UTILSAVEEV(r) / UTILSAVE(r)) * dpsave(r)]
+ [0.01 * EVSCALFACT(r)]
* [sum(i,NSAV_COMM, PTAX(i,r) * [qo(i,r) - pop(r)])
+ sum(i,ENDW_COMM, sum(j,PROD_COMM,
ETAX(i,j,r) * [qfe(i,j,r) - pop(r)]))
+ sum(i,PCOMP_COMM, sum(j,PROD_COMM, sum(s,REG,
SFTAX(i,j,s,r) * [qfpc(i,j,s,r) - pop(r)])))
+ sum(i,MCOMP_COMM, sum(j,PROD_COMM, sum(s,REG,
SFTAX(i,j,s,r) * [qfmc(i,j,s,r) - pop(r)])))
+ sum(i,PCOMP_COMM, sum(s,REG, SPTAX(i,s,r) * [qppc(i,s,r) - pop(r)]))
+ sum(i,MCOMP_COMM, sum(s,REG, SPTAX(i,s,r) * [qpmc(i,s,r) - pop(r)]))
+ sum(i,PCOMP_COMM, sum(s,REG, SGTAX(i,s,r) * [qgpc(i,s,r) - pop(r)]))
+ sum(i,MCOMP_COMM, sum(s,REG, SGTAX(i,s,r) * [qgmc(i,s,r) - pop(r)]))
+ sum(i,TRAD_COMM, sum(s,REG, STAX(i,r,s) * [qs(i,r,s) - pop(r)]))
+ sum(i,TRAD_COMM, sum(s,REG, DTAX(i,s,r) * [qs(i,s,r) - pop(r)]))
+ sum(i,ENDW_COMM, VOA(i,r) * [qo(i,r) - pop(r)])
- VDEP(r) * [kb(r) - pop(r)]
+ sum(j,PCGDS_COMM, VOA(j,r) * ao(j,r))
+ sum(j,MCOMP_COMM, VC(j,r) * ao(j,r))
+ sum(j,PCGDS_COMM, VA(j,r) * ava(j,r))
+ sum(j,MCOMP_COMM, VAV(j,r) * avav(j,r))
+ sum(j,MCOMP_COMM, VAFE(j,r) * avafe(j,r))
+ sum(j,MCOMP_COMM, sum(s,REG, VAFS(j,r,s) * avafs(j,r,s)))
+ sum(j,PROD_COMM, sum(i,ENDW_COMM, VFA(i,j,r) * afe(i,j,r)))
+ sum(j,PROD_COMM, sum(i,TRAD_COMM, VFA(i,j,r) * af(i,j,r)))
+ sum(m,MARG_COMM, sum(i,TRAD_COMM, sum(s,REG,
VTMFSD(m,i,s,r) * atmfsd(m,i,s,r))))
+ sum(i,TRAD_COMM, sum(s,REG, VPAS(i,s,r) * ams(i,s,r)))
+ sum(i,TRAD_COMM, sum(s,REG, VGAS(i,s,r) * ams(i,s,r)))
+ sum(i,TRAD_COMM, sum(j,PROD_COMM, sum(s,REG,
VFAS(i,j,s,r) * ams(i,s,r))))
+ sum(i,TRAD_COMM, sum(s,REG, VSWD(i,r,s) * pfob(i,r,s)))
+ sum(m,MARG_COMM, VST(m,r) * pm(m,r))
+ NETINV(r) * pcgds(r)
- sum(i,TRAD_COMM, sum(s,REG, VSWD(i,s,r) * pfob(i,s,r)))
- sum(m,MARG_COMM, VTMD(m,r) * pt(m))
- SAVE(r) * psave(r)
+ sum(j,MCOMP_COMM, VAF(j,r) * [qof(j,r) - pop(r)])
+ sum(i,MCOMP_COMM, sum(s,REG,
[1 / (SIGMA(i) - 1)] * VPAS(i,s,r) * vp(i,s,r)))
+ sum(i,MCOMP_COMM, sum(s,REG,
[1 / (SIGMA(i) - 1)] * VGAS(i,s,r) * vg(i,s,r)))
+ sum(i,MCOMP_COMM, sum(j,PROD_COMM, sum(s,REG,
[1 / (SIGMA(i) - 1)] * VFAS(i,j,s,r) * vf(i,s,r))))
- sum(i,MCOMP_COMM, sum(s,REG, VAFS(i,r,s) * [ns(i,r,s) - ns(i,r,r)]))
- sum(i,MCOMP_COMM, VAFE(i,r) * [np(i,r) - ns(i,r,r)])]
+ 0.01 * INCOMEEV(r) * pop(r);
```

The right hand side of this expression is the real income decomposition and is on a per capita basis - all of the quantity terms in the expression are deflated by population, `pop(r)`. The new set definitions used in the firm-heterogeneity EV decomposition include: `MCOMP_COMM`, `PCOMP_COMM`, `PCGDS_COMM`. These represent the sets of monopolistically competitive commodities, perfectly competitive commodities, and a combination of capital goods and perfectly competitive commodities.

There are ten components in this EV decomposition: allocative efficiency effect, endowment effect, technical change effect, population effect, terms of trade effect, investment-savings effect, preference effect, scale effect, variety effect, and fixed cost effect. The breakdown of the EV decomposition into its components is outlined in Table 1 and all variable definitions are listed in Table A.1.

Table 1. Breakdown of terms in welfare decomposition.

Effect	Expression
Preference	-[0.01 * UTILELASEV(r) * INCOMEEV(r)] * [DPARPRIV(r) * loge(UTILPRIVEV(r) / UTILPRIV(r)) * dppriv(r) + DPARGOV(r) * loge(UTILGOVEV(r) / UTILGOV(r)) * dpgov(r) + DPARSAVE(r) * loge(UTILSAVEEV(r) / UTILSAVE(r)) * dpsave(r)]
Allocative Efficiency	sum(i,NSAV_COMM, PTAX(i,r) * [qo(i,r) - pop(r)]) + sum(i,ENDW_COMM, sum(j,PROD_COMM, ETAX(i,j,r) * [qfe(i,j,r) - pop(r)])) + sum(i,PCOMP_COMM, sum(j,PROD_COMM, sum(s,REG, SFTAX(i,s,j,r) * [qfpc(i,s,j,r) - pop(r)]))) + sum(i,MCOMP_COMM, sum(j,PROD_COMM, sum(s,REG, SFTAX(i,s,j,r) * [qfmc(i,s,j,r) - pop(r)]))) + sum(i,PCOMP_COMM, sum(s,REG, SPTAX(i,s,r) * [qppc(i,s,r) - pop(r)])) + sum(i,MCOMP_COMM, sum(s,REG, SPTAX(i,s,r) * [qpmc(i,s,r) - pop(r)])) + sum(i,PCOMP_COMM, sum(s,REG, SGTAX(i,s,r) * [qgpc(i,s,r) - pop(r)])) + sum(i,MCOMP_COMM, sum(s,REG, SGTAX(i,s,r) * [qgmc(i,s,r) - pop(r)])) + sum(i,TRAD_COMM, sum(s,REG, STAX(i,r,s) * [qs(i,r,s) - pop(r)])) + sum(i,TRAD_COMM, sum(s,REG, DTAX(i,s,r) * [qs(i,s,r) - pop(r)]))
Endowment	+ sum(i,ENDW_COMM, VOA(i,r) * [qo(i,r) - pop(r)]) - VDEP(r) * [kb(r) - pop(r)]
Technical Change (Productivity)	+ sum(j,PCGDS_COMM, VOA(j,r) * ao(j,r)) + sum(j,MCOMP_COMM, VC(j,r) * ao(j,r)) + sum(j,PCGDS_COMM, VA(j,r) * ava(j,r)) + sum(j,MCOMP_COMM, VAV(j,r) * avav(j,r)) + sum(j,MCOMP_COMM, VAFE(j,r) * avafe(j,r)) + sum(j,MCOMP_COMM, sum(s,REG, VAFS(j,r,s) * avafs(j,r,s))) + sum(j,PROD_COMM, sum(i,ENDW_COMM, VFA(i,j,r) * afe(i,j,r))) + sum(j,PROD_COMM, sum(i,TRAD_COMM, VFA(i,j,r) * af(i,j,r))) + sum(m,MARG_COMM, sum(i,TRAD_COMM, sum(s,REG, VTMFSD(m,i,s,r) * atmfsd(m,i,s,r)))) + sum(i,TRAD_COMM, sum(s,REG, VPAS(i,s,r) * ams(i,s,r))) + sum(i,TRAD_COMM, sum(s,REG, VGAS(i,s,r) * ams(i,s,r))) + sum(i,TRAD_COMM, sum(j,PROD_COMM, sum(s,REG, VFAS(i,s,j,r) * ams(i,s,r))))
Terms of Trade	+ sum(i,TRAD_COMM, sum(s,REG, VSWD(i,r,s) * pfob(i,r,s))) + sum(m,MARG_COMM, VST(m,r) * pm(m,r)) - sum(i,TRAD_COMM, sum(s,REG, VSWD(i,s,r) * pfob(i,s,r))) - sum(m,MARG_COMM, VTMD(m,r) * pt(m))
Investment and Savings	+ NETINV(r) * pcgds(r) - SAVE(r) * psave(r)
Scale	+ sum(j,MCOMP_COMM, VAF(j,r) * [qof(j,r) - pop(r)])
Variety	+ sum(i,MCOMP_COMM, sum(s,REG, {1/[SIGMA(i) - 1]} * VPAS(i,s,r) * vp(i,s,r))) + sum(i,MCOMP_COMM, sum(s,REG, {1/[SIGMA(i) - 1]} * VGAS(i,s,r) * vg(i,s,r))) + sum(i,MCOMP_COMM, sum(j,PROD_COMM, sum(s,REG, {1/[SIGMA(i) - 1]} * VFAS(i,s,j,r) * vf(i,s,r))))
Fixed Cost	- sum(i,MCOMP_COMM, VAFE(i,r) * [np(i,r) - ns(i,r,r)])] - sum(i,MCOMP_COMM, sum(s,REG, VAFS(i,r,s) * [ns(i,r,s) - ns(i,r,r)]))
Population	+ 0.01 * INCOMEEV(r) * pop(r)

Source: Author calculations.

In this section, we only discuss the new effects in the firm-heterogeneity decomposition expression. We begin with the technical change (productivity) effect. The following expression, `CNTtechr(r)` captures the impact of technical change on regional welfare:

```
CNTtechr(r)
= [0.01 * EVSCALFACT(r)]
* [sum(j,PCGDS_COMM, VOA(j,r) * ao(j,r))
+ sum(j,MCOMP_COMM, VC(j,r) * ao(j,r))
```

```
+ sum(j,PCGDS_COMM, VA(j,r) * ava(j,r))
+ sum(j,MCOMP_COMM, VAV(j,r) * avav(j,r))
+ sum(j,MCOMP_COMM, VAFE(j,r) * avafe(j,r))
+ sum(j,MCOMP_COMM, sum(s,REG, VAFS(j,r,s) * avafs(j,r,s)))
+ sum(j,PROD_COMM, sum(i,ENDW_COMM, VFA(i,j,r) * afe(i,j,r)))
+ sum(j,PROD_COMM, sum(i,TRAD_COMM, VFA(i,j,r) * af(i,j,r)))
+ sum(m,MARG_COMM, sum(i,TRAD_COMM, sum(s,REG,
      VTMFSD(m,i,s,r) * atmfsd(m,i,s,r))))
+ sum(i,TRAD_COMM, sum(s,REG, VPAS(i,s,r) * ams(i,s,r)))
+ sum(i,TRAD_COMM, sum(s,REG, VGAS(i,s,r) * ams(i,s,r)))
+ sum(i,TRAD_COMM, sum(j,PROD_COMM, sum(s,REG,
      VFAS(i,j,s,r) * ams(i,s,r))))];
```

The right-hand side consists of value flows multiplied by their associated technical change variables, showing the contribution of technical change to regional welfare change. Since technical change variables are defined as exogenous, their contribution to welfare is zero unless there is an exogenous technology shock imposed. On the other hand, average industry productivity in the monopolistically competitive industry is endogenous. The expression `sum(j,MCOMP_COMM,VC(j,r) * ao(j,r))` incorporates endogenous productivity changes resulting from trade policies. For example, `ao(j,r)` of the exporter increases in the face of a tariff-cut which then leads to welfare gains based on the expression above.

Note that there are three new policy instruments introduced in the technical change effect: `avav(j,r)`, `avafe(j,r)`, and `avafs(j,r,s)`. These are exogenous variables that represent input augmenting technical change variables for value-added composites used in variable, fixed set-up, and fixed trading costs, respectively. An increase in these variables imply an improvement in the efficiency of inputs employed in covering associated costs. For example, a reduction in fixed trading costs will be implemented as an increase in `avafs(j,r,s)` which will have a direct contribution to the technical change effect. It will also have an indirect effect through the change in `ao(j,r)`.

The effect of scale economies on regional welfare change is captured by the expression `CNTqofr(r)` as follows:

```
CNTqofr(r)
= sum(i,MCOMP_COMM, [0.01 * EVSCALFACT(r)] * VAF(i,r) * [qof(i,r) - pop(r)]);
```

The right-hand side of `CNTqofr(r)` consists of percentage change in output per firm per capita multiplied by fixed value-added purchases of the monopolistically competitive industry. As explained in Section 3.2.6, `qof(i,r)` captures relative changes in costs as a result of trade policies. With a tariff cut, increased imports of cheap intermediate inputs reduces average variable cost relative to scale constant average total cost and increases firm scale. This constitutes an important source of gains from trade.

The effect of love-of-variety on regional welfare change is given by the expression `CNTvar_r(r)` as follows:

```
CNTvar_r(r)
= [0.01 * EVSCALFACT(r)]
```

```
* [sum(i,MCOMP_COMM, sum(s,REG, [1/(SIGMA(i) - 1)] * VPAS(i,s,r) * vp(i,s,r)))
+ sum(i,MCOMP_COMM, sum(s,REG, [1/(SIGMA(i) - 1)] * VGAS(i,s,r) * vg(i,s,r)))
+ sum(i,MCOMP_COMM, sum(j,PROD_COMM, sum(s,REG,
   [1/(SIGMA(i) - 1)] * VFAS(i,j,s,r) * vf(i,s,r))))];
```

On the right hand side, variety index of consumers are multiplied by their purchases of differentiated products. Each agent, i.e. private household, government, and firms, benefits from the availability of new varieties captured by the variety index. Therefore, regional welfare increases as the number of varieties sourced from a region increases. Since consumers devote a larger proportion of their budget to domestically produced varieties, domestic variety numbers have a larger effect on regional welfare compared to the number of imported varieties.

Finally, the effect of fixed costs on regional welfare is governed by the expression `CNTfixed_r(r)` as follows.

```
CNTfixed_r(r)
= - [0.01 * EVSCALFACT(r)]
* [sum(i,MCOMP_COMM, sum(s,REG, VAFS(i,r,s) * [ns(i,r,s) - ns(i,r,r)]))
+ sum(i,MCOMP_COMM, VAFE(i,r) * [np(i,r) - ns(i,r,r)])];
```

Unlike other expressions given above, this term enters negatively into the EV decomposition (pre-multiplied by -1). On the right-hand side we have relative changes in firm numbers multiplied by fixed costs. The first relative change, `np(i,r) - ns(i,r,r)`, is the percentage change in the number of potential firms relative to the percentage change in the number of producers in the industry. This difference tracks those firms that pay the fixed set-up costs but cannot produce. Their fixed set-up cost payments exhaust all the positive operating profit and reduces regional welfare. Therefore, if the gap between `np(i,r)` and `ns(i,r,r)` widens, the number of non-producers in the industry increases which reduces regional welfare. That is why the right-hand side is pre-multiplied by -1. The second relative change, `ns(i,r,s) - ns(i,r,r)`, is the difference between the number of exporters and producers. Similarly, as the number of exporters increases relative to producers, there is an increase in fixed trading costs, which is a source of welfare loss.

4. Data and calibration

We use GTAP 8 Data Base (Narayanan, Aguiar, and McDougall, 2012) for the illustrative experiments in this paper. There are several changes required for the standard GTAP data base to be compatible with the demand structure in the GTAP firm-heterogeneity model.

As discussed in Section 3.1.1, consumer demand incorporates *love-of-variety* and imported varieties compete directly with domestic varieties in the monopolistically competitive industries that produce differentiated products. This is unlike the import-domestic distinction in the standard GTAP Data Base where composite imports are imperfect substitutes for the domestic commodity under the Armington assumption. In order to allow for direct substitution between imported and domestically produced varieties, we relax the assumption of composite imports being formed at the border. This entails tracking the geographical origin of agent purchases of imported products which results in source-destination specific value flows.

In order to obtain source-destination specific value-flows we transform the standard GTAP Data Base by sourcing imports to agents. We define the market share of each source region in total imports of the destination market. This share is then used to source out the composite imports to various agents in the model. This transformation follows from Swaminathan and Hertel (1996) and we outline the details in Appendix B.

Sourcing imports to agents means that we distinguish between the import purchases of private households from that of firms and government. This is critical for empirical work. One of the important features of the standard GTAP Data Base is the fact that it allows for differing import intensities by agent. In particular, some economies have very import intensive capital goods sector, while their consumption is largely from domestic goods. As a result, the effect of a trade policy will vary across different agents depending on their relative import intensities. It is important to retain this empirical heterogeneity in the model.

The firm-heterogeneity model requires additional information that is not readily available in the standard GTAP Data Base, such as information on key firm-heterogeneity parameters, fixed set-up costs, and fixed trading costs. In this section, we discuss how to obtain this information.

4.1 Parameters

There are two key parameters in the firm-heterogeneity model. The first one is the shape parameter of Pareto distribution, γ_i, that shows the productivity heterogeneity in the industry. As discussed in Section 3.2.1, γ_i reflects differences in productivity levels and thereby differences in prices across firms within the same industry. The second important parameter is the elasticity of substitution amongst varieties, σ_i, which translates differences in prices into differences in market shares. Large values of σ_i implies that the market is competitive, where low-productivity entrants are at a disadvantage against high-productivity incumbents (low γ_i). As discussed in Hillberry and Hummels (2013), the largest extensive margin effects can be observed when price differences are small (high γ_i) and consumers are less sensitive to price differences (low σ_i). This shows that parametric choice is crucial for quantitative results.

The parameterization of the firm heterogeneity model depends on the mathematical constraint discussed in Section 3.2.1. Since $\gamma_i > \sigma_i - 1$ has to hold for the model to be well-defined, parameter values of γ_i and σ_i need to be chosen carefully so that they satisfy this condition.

The GTAP data base does not contain any information on γ_i. Therefore, we use the shape parameter estimates provided in Spearot (2016), where shape parameters are structurally estimated using sectoral trade flows and tariff data at the GTAP sectoral definition.

Even though the GTAP data base contains information on substitution elasticities, it may not be appropriate to adopt these existing elasticity values in the firm-heterogeneity model as they are estimated as Armington elasticities (Akgul, Villoria, and Hertel, 2015). While it is conventional to use Armington elasticities in traditional models that are based on the Armington assumption, using the same values in a Melitz (2003) model could lead to overestimation of trade flows and welfare gains from trade liberalization (Dixon, Jerie, and Rimmer, 2015). In fact, Dixon, Jerie, and Rimmer (2015) discuss that welfare implications are close in Armington and Melitz (2003) models if elasticities are chosen so that trade flows are the same. This implies lower values for substitution elasticities in Melitz (2003) compared to Armington.

Akgul, Villoria, and Hertel (2015) uses a two-stage estimation method in order to obtain elasticity values that are consistent with the firm-heterogeneity theory. They estimate a combination of γ_i and σ_i and use the γ_i estimates of Spearot (2016) to infer the theory-consistent substitution elasticities. They find that substitution elasticity values in the firm-heterogeneity model are lower than the GTAP Armington elasticities, in line with Dixon, Jerie, and Rimmer (2015). We follow Akgul, Villoria, and Hertel (2015) to infer the substitution elasticity values that are used in the illustrative simulations under firm-heterogeneity.

We use GTAP 8 Data Base (Narayanan, Aguiar, and McDougall, 2012) for the illustrative experiments in this paper, where we aggregate the data base to two industries, manufacturing and non-manufacturing. The manufacturing industry is assumed to be monopolistically competitive with heterogeneous firms. Table 2 presents the parameter values that are associated with the manufacturing industry. While the GTAP Armington elasticity for manufactures, `ESUBM(i)`, is 6.96 in the data base, the firm-heterogeneity elasticity for manufactures is found as 3.75. We will distinguish between these two elasticity values in the simulation experiments presented in Section 6. While a practical policy-oriented comparison be-

Table 2. Parameter values for the manufacturing industry.

Parameter	Value
GTAP Armington Elasticity, `ESUBM`	6.96
Elasticity of Substitution between Varieties, σ	3.75
Pareto Shape Parameter, γ	2.89
Parameter Constraint, $\frac{\gamma}{\sigma-1}$	1.05
Proportion of Fixed Costs in Sales, $\frac{\gamma_i-\sigma_i+1}{\sigma_i\gamma_i}$	0.01

Source: Author calculations.

tween firm-heterogeneity and perfect competition specifications will require both the shape parameter and substitution elasticity to be estimated, we do not have enough information to estimate a more general specification in this study. Our aim is to illustrate the new mechanisms introduced by firm heterogeneity rather than to offer a comprehensive parameterization. Clearly this is an important area for future work.

As a result of these changes, there are three new headers in the parameters file: (i) SIGM for the elasticity of substitution amongst varieties in the monopolistically competitive industries, (ii) SHPE for the shape parameter of Pareto distribution that governs the productivity heterogeneity in the industry, and (iii) IND which is for the new set distinction between the monopolistically and perfectly competitive industries.

4.2 Calibration of fixed costs

To implement the firm heterogeneity model into GTAP, we need to know more about fixed costs. Since the GTAP data base does not contain information on fixed costs, we rely on calibration to obtain their initial magnitude.

Industry-wide fixed value-added costs incurred in the monopolistically competitive industry can be calculated as the difference between industry-wide total costs, $N_{irr}\widetilde{Q}_{ir}\widetilde{P}_{ir}$, and industry-wide variable costs, $N_{irr}\widetilde{Q}_{ir}\frac{C_{ir}}{\widetilde{\Phi}_{ir}}$, as follows:

$$\sum_s N_{irs}W_{irs}F_{irs} + N_{ir}^{p}W_{irr}H_{ir} = N_{irr}\widetilde{Q}_{ir}\left[\widetilde{P}_{ir} - \frac{C_{ir}}{\widetilde{\Phi}_{ir}}\right] \tag{34}$$

where the left-hand side is the fixed component of the value-added in industry i of region r, which is the summation of fixed set-up costs and fixed trading costs as alluded to earlier in Section 3.2.5. We then substitute the mark-up price Equation (12) into Equation (34) which yields the initial value of fixed costs as:

$$\sum_s N_{irs}W_{irs}F_{irs} + N_{ir}^{p}W_{irr}H_{ir} = N_{irr}\widetilde{Q}_{ir}\widetilde{P}_{ir}\frac{1}{\sigma_i} \tag{35}$$

It follows from Equation (35) that $\frac{1}{\sigma_i}$ of total costs in industry i is devoted to industry-wide fixed costs, in the Dixit and Stiglitz (1977) tradition. Note that fixed costs decrease with the elasticity of substitution. If preferences are heterogeneous, i.e. low σ_i, there is room in the industry to differentiate products which means that fixed costs are high. In such industries firms are encouraged to invest in developing a new product and charge a higher mark-up for it. On the other hand, if preferences are homogeneous, i.e. high σ_i, fixed costs are low as there is not much incentive to differentiate the product. In the extreme case, i.e. perfect competition with σ_i approaching infinity, fixed costs reduce to zero since all value-added is allocated to the production of identical varieties.

The calibration in Equation (35) is implemented in the code as follows:

```
Formula (initial)(all,i,MCOMP_COMM)(all,r,REG)
 VAF(i,r) = VOA(i,r) * [1 / SIGMA(i,r)];
```

where `VOA(i,r)` corresponds to $N_{irr}\widetilde{Q}_{ir}\widetilde{P}_{ir}$ in Equation (35).

The rest of the value-added costs in the industry, `VA(i,r)`, is attributed to variable value-added, `VAV(i,r)`, which is used in the actual production of the output. We calculate `VAV(i,r)` in the code as the difference between total value-added and fixed value-added costs as follows:

```
Formula (initial)(all,i,MCOMP_COMM)(all,r,REG)
 VAV(i,r) = VA(i,r) - VAF(i,r);
```

As shown on the left-hand side of Equation (35), fixed costs are composed of two parts: fixed trading costs and fixed set-up costs. Since the data base does not contain information on either, we calibrate them as well. We first calibrate the initial value of fixed trading costs to the base year bilateral trade flows following Zhai (2008). Then, we use this new information to calculate fixed set-up costs.

Bilateral trade flows in industry i from region r to s equal $\frac{N_{irs}\widetilde{P}_{irs}\widetilde{Q}_{irs}}{T_{irs}}$. When we use the optimal demand equation (4) and the optimal price equation (10) in bilateral trade flows, we obtain the following gravity equation:

$$\frac{N_{irs}\widetilde{P}_{irs}\widetilde{Q}_{irs}}{T_{irs}} = N_{irs}\frac{Q_{is}{P_{is}}^{\sigma_i}}{T_{irs}}\left[\frac{\sigma_i}{\sigma_i - 1}\frac{C_{ir}T_{irs}}{\widetilde{\Phi}_{irs}}\right]^{1-\sigma_i} \tag{36}$$

which shows that trade patterns depend on market size (Q_{is}), active firms (N_{irs}), trade barriers (T_{irs}), competition (P_{is}), productivity ($\widetilde{\Phi}_{irs}$ and C_{ir}). To calculate the value of fixed trading costs, we use the productivity threshold equation (20) in Equation (36) which yields:

$$N_{irs}W_{irs}F_{irs} = \frac{N_{irs}\widetilde{P}_{irs}\widetilde{Q}_{irs}}{T_{irs}}\frac{\gamma_i - \sigma_i + 1}{\sigma_i\gamma_i} \tag{37}$$

where $\gamma_i > \sigma_i - 1$ (see Appendix C.3 for details of this derivation). It follows from Equation (37) that $\frac{\gamma_i - \sigma_i + 1}{\sigma_i\gamma_i}$ of sales revenue in industry i of operating on the $r - s$ trade route is devoted to the fixed trading cost associated with operating in that market. The calibration in Equation (37) is implemented in the code as follows:

```
Formula (initial)(all,i,MCOMP_COMM)(all,r,REG)(all,s,REG)
 VAFS(i,r,s) = VSMD(i,r,s) * [SHAPE(i) - SIGMA(i) + 1] / [SHAPE(i) * SIGMA(i)];
```

where the fixed trading cost of sales of i from r to s, `VAFS(i,r,s)`, corresponds to $N_{irs}W_{irs}F_{irs}$ and the value of sales of i from r to s, `VSMD(i,r,s)`, corresponds to $\frac{N_{irs}\widetilde{P}_{irs}\widetilde{Q}_{irs}}{T_{irs}}$ in Equation (37).

Once fixed trading costs are calibrated, the initial value for fixed set-up costs, `VAFE(i,r)`, is obtained as the difference between industry-wide fixed costs and fixed trading costs aggregated across all destination markets. This is implemented in the code as:

```
Formula (all,i,MCOMP_COMM)(all,r,REG)
 VAFE(i,r) = VAF(i,r) - sum(s,REG, VAFS(i,r,s));
```

There are two important checks to conduct in the calibrations described here. Since

`VAFE(i,r)` and `VAV(i,r)` are calculated as residuals, there is a possibility that the resulting values are extremely low suggesting that value-added cost is low relative to total cost. In extreme cases, the calibration may result in negative value flows for a given substitution elasticity and shape parameter. This may be the indicator of a misclassification of the market structure or inappropriate parameter values. In such a case, the modeler should decide whether a different market structure is more appropriate for that particular industry. An alternative solution is to use different parameter values. For example, if the substitution elasticity value that is required for nonnegative `VAFE(i,r)` and `VAV(i,r)` is too high, then it may be the case that the industry should be characterized by perfect competition rather than monopolistic competition with heterogeneous firms.

5. Varying closure to compare Armington, Krugman, and Melitz specifications

In an applied CGE work, it is important to choose a specification that is in line with the characteristics of a given industry. In particular, industries where firms produce homogeneous commodities, may be more appropriately modeled by perfect competition as opposed to firm heterogeneity or monopolistic competition. There is increasing evidence supporting the relative strengths of each mechanism depending on the industry, initial conditions, and the trade policy being explored. Especially, the recent work by Dixon, Jerie, and Rimmer (2015) highlights the connections between these three structures and allows for nesting between them. Motivated by this approach, we compare simulation results obtained by firm heterogeneity with those obtained by alternative trade specifications, i.e. monopolistic and perfect competition. We use closure swaps to allow for alternative trade specifications. We start with the GTAP firm-heterogeneity model and impose several restrictions to derive the monopolistically and perfectly competitive modules of GTAP.

The monopolistically competitive model differs from Melitz (2003) on two fronts: (i) there are no fixed export costs and (ii) firms are identical with respect to their productivity levels. These imply that all firms are active in all destination markets. In order to reduce the firm heterogeneity module to the monopolistic competition, we need to remove the feedback from endogenous productivity thresholds to average industry productivity. This is achieved by setting the bilateral productivity thresholds as exogenous in the closure, using swap statements. By exogenizing productivity thresholds we ensure that all new entrants become exporters via Equation (31). Hence the resulting change in average industry productivity in Equation (23) will be zero.

To modify the closure we use the slack variable `threshslack(i,r,s)` introduced in Equation `PRODTRESHOLD` and apply the following swap statement.

```
swap aost(MCOMP_COMM,REG,REG) = threshslack(MCOMP_COMM,REG,REG);
```

In closure swaps, we adopt the convention that the variables that are being exogenized are given on the left-hand side of the swap statement, while the variables that are being endogenized are given on the right-hand side. Thus, the above closure swap ensures that the productivity thresholds in bilateral markets, `aost`, are exogenous, while the associated slack variables, `threshslack`, are endogenous. As a result of exogenous productivity thresholds, the percentage change in the average productivity of the domestic or export markets is also zero. Needless to say, their contribution to changes in aggregate productivity is also zero based on Equation (23). Average productivity is automatically exogenous since the components that determine it are exogenous by the closure.

We impose further restrictions on the monopolistically competitive model to obtain the perfect competition model. The formulation based on the Armington assumption entails the standard GTAP model assumptions of perfect competition, and constant returns to scale, where identical firms produce identical products. Since there is no product differentiation, there are no fixed costs associated with production in this framework. Neither the firm, nor the industry makes positive profits. The key difference between Krugman and Armington specifications is twofold: (i) products are homogeneous therefore we do not observe the *love-of-variety* in demand, and (ii) there are no fixed costs associated with production in the perfectly competitive industry; therefore, there are no economies of scale. Thus, in order to reduce the model to the Armington specification, we switch off both the *love-of-variety* effect and the scale economies. This is achieved by imposing the following closure swaps:

```
swap vp(MCOMP_COMM,REG,REG) = vpslack(MCOMP_COMM,REG,REG);
swap vg(MCOMP_COMM,REG,REG) = vgslack(MCOMP_COMM,REG,REG);
swap vf(MCOMP_COMM,REG,REG) = vfslack(MCOMP_COMM,REG,REG);
swap qof(MCOMP_COMM,REG) = mkupslack(MCOMP_COMM,REG);
```

The first three swap statements remove the effect of *love-of-variety* on consumer demand by setting the variety index of agents as exogenous and the associated slack variables as endogenous. Note that this does not mean that there is no change in the number of varieties in the industry. Output variations in the industry are accommodated by variations in variety/firm numbers under perfect competition. However, these changes no longer generate a *love-of-variety* effect in consumer utility due to the closure swaps imposed.

The last swap statement removes the effect of scale economies on firm production. Firms cannot markup their prices under a perfectly competitive market structure. Instead they charge competitive prices. As a result, the markup equation becomes redundant when we switch to perfect competition. In order to relax the markup pricing rule, the associated slack variable, `mkupslack`, is defined to be endogenous in the closure. Moreover, in a perfectly competitive market, all output expansion in the industry occurs by adding more identical firms into the market at constant cost. Therefore, per firm output, `qof` does not change and is defined as exogenous in the closure.

6. Policy application

In this section, we investigate the behavioral characteristics of the GTAP firm-heterogeneity model and compare it with those of Armington and Krugman modules of GTAP in the context of a bilateral tariff cut scenario. The numerical implementation of these theoretical models is carried out by a stylized scenario in order to keep the analysis tractable and provide a relatively transparent interpretation of results.

We calibrate the model to GTAP 8 Data Base (Narayanan, Aguiar, and McDougall, 2012) for 2007. We aggregate the data base to 3 regions: USA, Japan and ROW; and 2 sectors: manufacturing and non-manufacturing. The manufacturing sector is treated as monopolistically competitive with heterogeneous firms, while the non-manufacturing sector retains the perfect competition structure with Armington assumption.

To compare simulation results across different trade specifications, we first need to define the equivalence of these models. We adopt the convention in Dixon, Jerie, and Rimmer (2015) and define equivalence of Armington and Melitz simulations as giving the same trade flow responses to tariff policy. To obtain this equivalence, we calibrate the tariff policy in Armington and Melitz simulations to give the same changes in trade flows. Based on this calibration, policy experiments in Armington and Krugman simulations involve complete elimination of tariffs levied by Japan on the import of US manufacturing goods, which is a 3.66% decrease in the power of the aggregate tariff imposed on US manufactures. On the other hand, the implemented shock in the Melitz simulation is a 3.31% decrease in the power of the aggregate tariff imposed on US manufactures. Note that the same level of manufactures sales from the US to Japan is attained with a lower tariff cut in the Melitz simulation compared to the Armington simulation.

In addition to the difference in the tariff cut, there is also a difference between the values of substitution elasticities used in these simulations. While we retain the GTAP Armington elasticities for Armington and Krugman simulations, we use lower values of substitution elasticities for Melitz simulations.

Table 3 presents simulation results for the three specifications. We first focus on analyzing the insights obtained from the tariff cut scenario in the firm heterogeneity model. These results are then compared with those obtained from the monopolistic and perfect competition models. Note that in all three specifications we assume the external account to be always on balance, dictated by changes in trade and the capital account. Trade balance is endogenously determined by changes in the real exchange rate, `pfactor(r)` in GTAP. We also assume that expected rates of returns are equalized across regions through the allocation of investment which governs the changes in the capital account.

Table 3. Simulation results for the manufacturing sector under Armington, Krugman, and Melitz specifications.

Variable	**Notation in Code**		**Model**								
			FH			**MC**			**PC**		
			USA	JPN	ROW	USA	JPN	ROW	USA	JPN	ROW
Sales (%)	`qs(j,r,s)`	USA	-0.07	26.90	-0.54	-0.13	26.90	-0.56	-0.13	26.90	-0.57
		JPN	1.65	-0.26	1.15	1.54	-0.41	1.07	1.52	-0.43	1.05
		ROW	0.46	-1.53	-0.01	0.45	-1.54	-0.01	0.45	-1.53	-0.01
Industry Output (%)	`qo(j,r)`		0.09	0.08	0.00	0.04	-0.05	0.00	0.03	-0.07	0.00
Supply Price (%)	`ps(j,r)`		-0.01	-0.30	-0.04	0.07	-0.19	-0.02	0.07	-0.18	-0.02
Average Variable Cost (%)	`avc(j,r)`		-0.01	-0.30	-0.04	0.07	-0.19	-0.02	0.06	-0.18	-0.02
Scale Constant Average Total Cost (%)	`scatc(j,r)`		0.07	-0.14	-0.04	0.07	-0.19	-0.02	0.07	-0.18	-0.02
Number of Potential Firms (%)	`np(j,r)`		-0.03	-0.16	0.00	0.00	-0.09	0.00	0.03	-0.07	0.00
Number of Active Firms (%)	`ns(j,r,s)`	USA	-0.20	12.79	-0.43	0.00	0.00	0.00	0.03	0.03	0.03
		JPN	0.46	-0.51	0.20	-0.09	-0.09	-0.09	-0.07	-0.07	-0.07
		ROW	0.23	-0.79	-0.01	0.00	0.00	0.00	0.00	0.00	0.00
Output per Firm (%)	`qof(j,r)`		0.29	0.60	0.01	0.03	0.03	0.00	0	0	0
Productivity Threshold for Market Entry (%)	`aost(j,r,s)`	USA	0.06	-4.09	0.14	0	0	0	0	0	0
		JPN	-0.21	0.12	-0.12	0	0	0	0	0	0
		ROW	-0.08	0.27	0.00	0	0	0	0	0	0
Average Productivity (%)	`ao(j,r)`		0.06	0.13	0.00	0	0	0	0	0	0
Terms of Trade (%)	`tot(r)`		0.08	-0.24	0.03	0.10	-0.18	-0.01	0.10	-0.16	-0.01
Real Exchange Rate (%)	`pfactor(r)`		0.11	-0.05	-0.04	0.10	-0.16	-0.02	0.09	-0.14	-0.02

Notes: FH: firm heterogeneity, i.e. Melitz, MC: monopolistic competition, i.e. Krugman, and PC: perfect competition, i.e. Armington. `j` ∈ `MCOMP_COMM` for monopolistically competitive industry, and `r,s` ∈ `REG` for regions.

Source: Author calculations.

6.1 Impacts on the US

The direct effect of tariff cut is a reduction in the price of US manufactures in the Japanese market by 3.32% which is accompanied by an increase in sales of US manufactures in Japan by 26.90%. The tariff cut causes a real appreciation in the relative prices of primary factors in the US by 0.11%, raising the cost of US varieties in the domestic and ROW markets. This diverts sales from the domestic and ROW markets (-0.07% and -0.54%, respectively) bringing the US economy back into external balance. These results constitute a familiar narrative of the immediate effects of a tariff cut in an exporting region. Firm heterogeneity adds another aspect to this picture as increased exposure to trade generates endogenous productivity changes under this framework.

Figure 2, Panel A indicates the percentage change in productivity thresholds in the US for each destination market, `aost(j,r,s)`, and percentage change in the number of active firms in these markets, `ns(j,r,s)`. We observe that the productivity threshold to produce in the US manufacturing industry increases by 0.06%. This rise forces the least efficient firms out of the industry since they can no longer compete in the post-tariff economy. On the other hand, more productive firms find it profitable to expand their sales into export markets which bids up factor prices in the US. Therefore, inefficient firms lose competitiveness against cheaper imports coming from Japan and the ROW. The number of active US firms in the domestic market decreases by 0.20%. This is an example of inter-firm reallocation of resources within the industry as more-productive firms absorb the factors released from the exiting firms while gaining a larger share of the home market.

Within industry reallocation of production extends to export markets through the shifts in bilateral productivity thresholds. In particular, the tariff cut in Japan lowers the productivity threshold for US manufacturing firms exporting into the Japanese market by 4.09% as depicted in Figure 2, Panel A. Unlike in the home market case, the marginal firm on the export threshold benefits from this tariff cut since its productivity level is now higher than the threshold such that it can make positive profits by exporting to Japan. The same applies to the mass of firms that are below the pre-tariff cut threshold, but above the post-tariff cut one.

There are two factors at play for US manufacturing firms exporting into Japan: (i) increased competitiveness, and (ii) larger market access. As noted above, US manufacturing firms are less competitive in domestic and ROW markets due to higher factor costs. On the other hand, the tariff cut allows US firms to be more competitive in the Japanese market and take advantage of greater market access. As a result, sales to Japan rise by 26.90% which lowers fixed export cost per sale. This drop in fixed cost per sale raises the potential for positive profits and induces a rise in the number of US firms exporting into the Japanese market by 12.79%.

It is appealing to think that higher competitiveness and market access should benefit all US firms equally by creating positive profits. However, the impact of

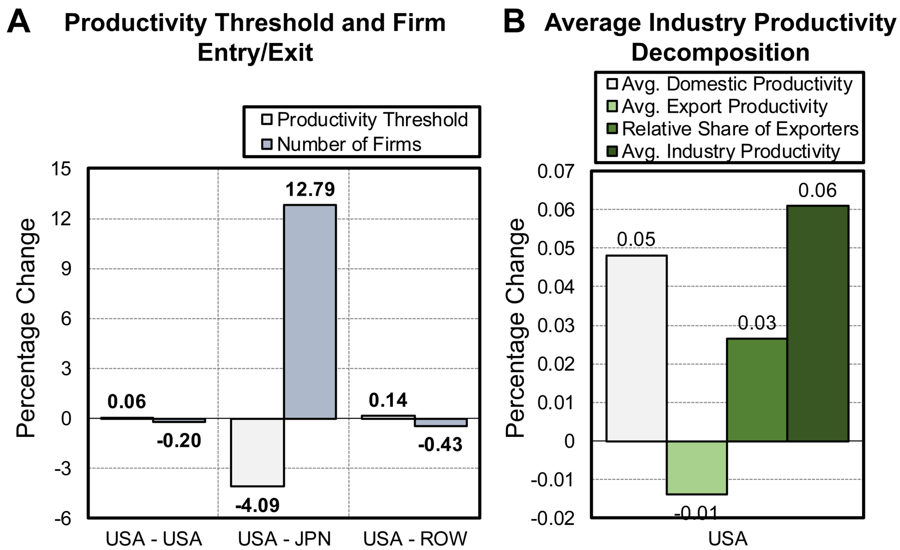

Figure 2. Productivity threshold, firm entry/exit and the decomposition of average industry productivity in the US.

Source: Author calculations.

tariff cut on each firm varies by efficiency. In the case of low-productivity firms, the impact of higher competition on firm profits dominates since their costs are too high to take advantage of bigger market size. Facing negative profits in the Japanese market, high-cost firms do not export to Japan, but continue to produce for the domestic market. On the other hand, firms with productivity levels above the new threshold are competitive enough to make use of the larger market. Therefore, they begin to sell in the Japanese market. Firm entry continues until potential profits from exporting are exhausted. As a result, even though there are fewer potential firms in the manufacturing industry (`np(j,r)`=-0.03%), more of them export to Japan. As depicted in Figure 2, panel A, there is an increase in the productivity threshold for exporting into the ROW market by 0.14%, which in turn generates a drop in the number of exporters by 0.43%.

Importantly, even though the new exporters have higher productivity levels compared to non-exporters, they are relatively less productive than the existing exporters. As a result, new exporters pull down the average productivity in export markets.

The overall effect on average productivity of the manufacturing industry is presented in Figure 2, panel B. The percentage change in the industry productivity is decomposed into share-weighted average productivity in domestic and export markets as well as the relative market shares of exporters. We observe that the rise in share-weighted domestic productivity by 0.05% dominates the reduction in share-weighted export productivity (-0.01%). This is due to the fact that home market has a much bigger share in sales compared to export markets. Moreover, we observe that the expansion of the export market contributes further to the industry average by 0.03% as average productivity of exporters is higher than that of domestic producers. Therefore, average productivity in the US manufacturing industry increases by 0.06%. This is purely a gain of inter-firm reallocation of resources within the manufacturing industry.

6.2 Impacts on Japan

Table 3 and Figure 3 present the effects of this tariff cut on the Japanese economy. We observe that cheaper US varieties increases competition and crowds out Japanese firms from the market. This results in a drop in domestic sales by 0.26%.

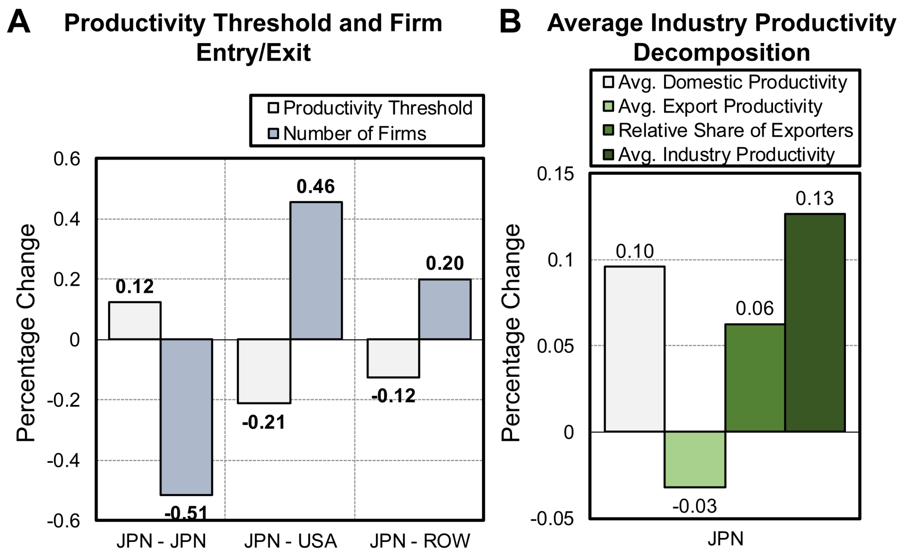

Figure 3. Productivity threshold, firm entry/exit and the decomposition of average industry productivity in Japan.

Source: Author calculations.

Although some firms are replaced by US competitors in the home market, surviving Japanese firms benefit from the cheap US manufactures. There is, in fact, a large increase in the demand for intermediate inputs sourced from the US, 26.80% increase in the manufacturing industry demand and 27.19% increase in the non-manufacturing industry demand for US manufactures. Lower prices for intermediate inputs reduce the average cost of production in Japan by 0.30%. This is good news for Japanese exporters as they are now more competitive in export markets. Relative prices of primary factors decrease in Japan implying a real depreciation of 0.05%. This restores the external balance in Japan by stimulating exports. In particular, Japanese exports to the US and ROW markets increase by 1.65% and 1.15%, respectively. As Japanese exporters gain access to larger markets, their fixed export costs per sale decline. This, together with the declining average variable costs, leads to reductions in productivity thresholds of exporting to US (-0.21%)

and ROW (-0.12%) markets as depicted in Figure 3, Panel A. Since export thresholds are now lower, the number of Japanese exporters in the US and ROW markets increases by 0.46% and 0.20%, respectively.

Despite the gain in export markets, the home market suffers from the loss of domestic varieties. As shown in Figure 3, panel A, the productivity threshold of supplying the domestic market increases by 0.12% which reduces the number of firms in the domestic market by 0.51%. Similar to the case of US, this tariff cut reallocates market share by shifting resources towards more productive firms, generating average productivity gains in the manufacturing industry. This is highlighted in the decomposition depicted in Figure 3, panel B. Average productivity in the domestic market rises by 0.10% overcompensating for the 0.03% drop in the average productivity of export markets. In addition, the export market becomes more important in the overall weight which pulls up the industry average further by 0.06%. Consequently, average industry productivity rises by 0.13%.

Overall, the tariff cut improves the industry efficiency not only in the US, but also in Japan. This is a good example of the importance of within industry reallocation of firms in facilitating trade through international supply chains.

6.3 Impacts on ROW

The impact of this tariff cut on the ROW region is less pronounced when compared to other regions. Figure 4, Panel A presents simulation results for percentage changes in productivity thresholds and firm entry/exit in the ROW.

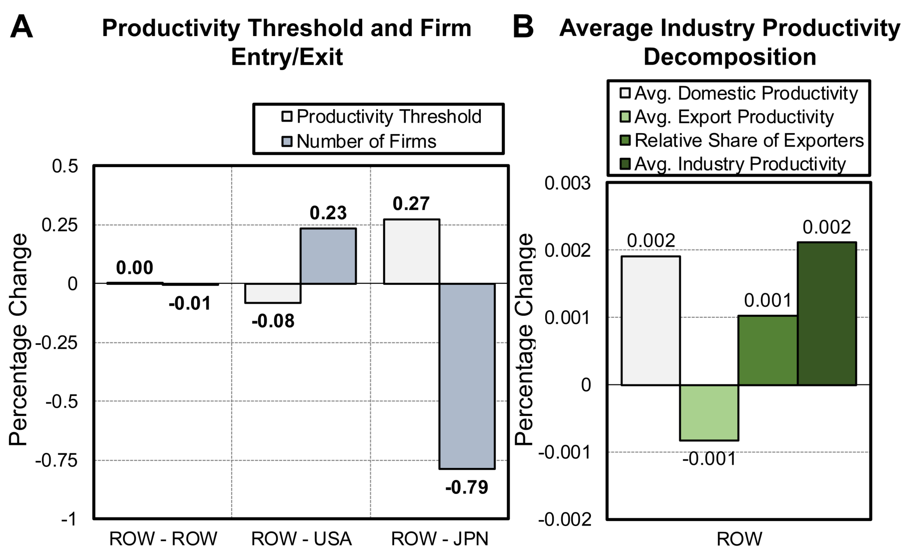

Figure 4. Productivity threshold, firm entry/exit and the decomposition of average industry productivity in the ROW.

Source: Author calculations.

The most striking change is observed in trade between ROW and Japan. The productivity threshold to export into Japan increases by 0.27% which is largely a result of the US competition. Demand for ROW manufactures in the Japanese market is displaced with US varieties leading to a drop in the number of ROW exporters by 0.79%. In addition, sales in the ROW home market decreases by 0.01%. We observe that the domestic productivity threshold increases by a very small amount which causes firm exit by 0.01%.

While there is trade diversion both in the home market and in the Japanese market, the external balance in the ROW is restored by trade creation in the US market (0.46%) as a result of real exchange rate depreciation (0.04%). There is a slight decrease in the productivity threshold by 0.08% which raises the number of ROW exporters into the US by 0.23%.

Consequently, aggregate productivity in the ROW manufactures sector increases by a very small amount as presented in Figure 4, Panel B. In practical terms, this is a negligible change and likely indistinguishable from zero. However, the productivity gain shows that firm reallocation in the ROW is similar to the experiences in Japan and in the US. In particular, the tariff cut leads to an efficiency gain in the industry where low-productivity firms contract and high-productivity firms expand their share into export markets.

6.4 Comparison across different model specifications

We start with firm heterogeneity and successively restrict the model to yield the simpler models: monopolistic and perfect competition. Then, we explore the tariff cut scenario between the US and Japan in the context of each model by calibrating the shock such that trade flows are equivalent under firm heterogeneity and perfect competition.

Table 3 reports the findings. A quick look at the results illustrates that the firm heterogeneity model captures the changes that occur in a conventional CGE model with the Armington assumption. Moreover, it includes the effect of changes in varieties as well as economies of scale delivered by the monopolistically competitive structure and furthermore incorporates the productivity channel that is linked with factor reallocation across firms within the same industry.

There are several differences in tariff-cut implications on prices, costs, and production under firm heterogeneity. A striking difference is observed in changes in average variable costs and thereby in supplier prices. Even though primary factor prices increase under all three specifications, the US average variable cost decreases under firm heterogeneity (-0.01%) as opposed to the increase in Armington and Krugman simulations (0.07% and 0.06%, respectively). This difference stems from the endogenous productivity changes captured by firm heterogeneity. As discussed before, average productivity increases in the US manufacturing industry following the tariff-cut. This productivity growth more than offsets the increase in factor prices; therefore, it reduces the average variable cost. The nature of this productivity growth should be kept in mind in interpreting this result. Lower average productivity does not mean that US firms are now more productive. It merely means that high-productivity firms have a larger share of the market and their lower costs pull down the average.

Another important difference under firm heterogeneity is the expansion of the manufacturing industry in Japan. In the Armington and Krugman simulations, the tariff cut leads to a contraction in the Japanese manufacturing industry (-0.072% and -0.054% respectively), while it leads to an expansion under firm heterogeneity (0.08%). We observe that, under firm heterogeneity, cheaper manufacturing imports from the US reduces average variable costs in Japan (-0.30%) more than it reduces average total costs (-0.14%). This leads to an increase in output per firm in the Japanese manufacturing industry (0.60%).

An important contribution of firm heterogeneity over monopolistic competition is the distinction between active firms and inactive firms in a market. The monopolistic competition model dictates that if a firm produces, it also exports into all destination markets. This is reflected in the results reported in Table 3. The percentage change in the number of exporters in each market, `ns(j,r,s)`, is equal to the percentage change in the number of potential varieties `np(j,r)`. However, it does not take the specific circumstances of each firm and each destination into account. Once we factor in the heterogeneity of productivity across firms and destination-specific fixed costs, we observe that not all firms are able to export into all destinations. In fact, the number of US firms that export to Japan increases (12.79%), while the number of US firms that export to the ROW declines (-0.44%) in contrast to the monopolistically competitive model which predicts an equal change in exporters to all destinations (0.00%).

6.5 Welfare effects

There is, currently, no consensus in the literature on the welfare implications of the Melitz model compared to those from traditional models with the Armington assumption. In order to do accurate policy analysis in a CGE setting, we need to understand how these models differ. Are there additional gains from trade that we are not accounting for when we choose one model over the other? If there are, do they matter in the overall welfare response? Do they contribute to aggregate welfare? These questions are getting more attention in the CGE literature.

In related work, welfare changes in the Melitz (2003) model are found to be larger than an Armington (1969) benchmark (Balistreri, Hillberry, and Rutherford, 2011; Kancs, 2010; Zhai, 2008). In fact, incorporating firm heterogeneity into standard CGE models raises the gains from trade liberalization by a multiple of two in Zhai (2008) and by a multiple of four in Balistreri, Hillberry, and Rutherford (2011). However, Arkolakis et al. (2008) and Arkolakis, Costinot, and Rodriguez-Clare (2012) argue that the impact of trade cost reductions is similar across models once their trade responses are equalized via the calibration of parameters. This argument suggests that the Melitz (2003) model does not offer additional gains from trade conditional on equal trade patterns. A similar finding is discussed by Dixon, Jerie, and Rimmer (2015). Having started from an undistorted initial equilibrium, Dixon, Jerie, and Rimmer (2015) observe that gains from productivity and preferences in firm heterogeneity offset each other which results in equal welfare change once the observed trade pattern is fitted with higher substitution elasticities in the Armington formulation. However, their model does not incorporate intermediate inputs in production and trade structure. The presence of imported intermediate inputs can be a major driver in welfare changes Lanclos and Hertel (1995). As discussed in Balistreri and Rutherford (2013), the incorporation of real-world complexities in quantitative CGE models can be important and generate divergence of results across model specifications.

Table 4 provides a summary of regional welfare change and decomposition in each model under the tariff cut scenario. We observe that the global welfare gain is much larger in magnitude under firm heterogeneity ($4002 million) compared to monopolistic competition ($393 million) and perfect competition ($423 million). Regional welfare change is also different under firm heterogeneity. In particular, welfare gain in the US ($4846 million) and welfare loss in the ROW (-$1314 million) are much greater. Another striking difference under firm heterogeneity is the sign reversal of welfare change in Japan. While the tariff cut causes welfare loss in Japan under monopolistic (-$1509 million) and perfect competition (-$1191 million), it leads to welfare gain ($469 million) under firm heterogeneity. These results show that variety, scale, fixed cost, and productivity effects captured under firm heterogeneity have significant implications for the magnitude of regional welfare change.

Under firm heterogeneity, we observe that the tariff cut leads to an increase in the average productivity of the US manufacturing industry. This is due to the expansion in market shares of high-productivity firms. Tariff-cut leads to a higher domestic productivity threshold due to higher import competition, and leads to a lower export threshold into the Japanese market due to lower barriers to trade. Thus, less efficient firms exit the industry, while more efficient firms expand into the Japanese market. This compositional change raises the overall efficiency in the industry contributing positively to the welfare in the US, $2665 million.

This is accompanied by the positive scale effects of $4738 million. The scale effect in the firm heterogeneity model is determined by changes in output per firm. We observe that the domestic market is supplied by fewer US firms due to higher domestic threshold. The surviving firms have to operate on a larger scale to allow for expanding output which enhances welfare in the US. In contrast, the variety effect is negative, -$2391 million, as consumers suffer from a loss in domestic varieties. Even though US enjoys a wider selection of Japanese and ROW varieties, the decreasing number of US varieties more than offsets this positive contribution. This confirms the home bias as the loss in domestic varieties is more dominant in the final variety effect. Finally, fixed cost payments reduce the welfare in US by

Journal of Global Economic Analysis, Volume 1 (2016), No. 1, pp. 111-180.

Table 4. Welfare decomposition of tariff removal under Armington, Krugman, and Melitz specifications: Equivalent variation in millions of US$.

Model	Region	Aggregate Welfare Effect	Allocative Efficiency Effects	TOT Effects	I-S Effects	Variety Effects	Scale Effects	Productivity Effects	Fixed Cost Effects
FH	USA	4846	539	1076	878	-2391	4738	2665	-2659
	JPN	469	476	-1858	11	-2671	4508	2632	-2629
	ROW	-1314	217	780	-889	-1779	356	400	-400
	Total	4002	1232	-1	-1	-6841	9603	5698	-5688
MC	USA	2758	287	1527	657	-16	303	0	0
	JPN	-1509	44	-1424	86	-341	125	0	0
	ROW	-856	150	-105	-744	-123	-35	0	0
	Total	393	481	-2	0	-479	393	0	0
PC	USA	2354	254	1490	611	0	0	0	0
	JPN	-1191	25	-1296	80	0	0	0	0
	ROW	-739	146	-195	-690	0	0	0	0
	Total	423	425	-2	0	0	0	0	0

Notes: FH: firm heterogeneity, i.e. Melitz, MC: monopolistic competition, i.e. Krugman, and PC: perfect competition, i.e. Armington.

Source: Author calculations.

$2659 million.

Contrary to the monopolistic and perfect competition models, Japan gains from this tariff removal scenario in the firm heterogeneity model. Similar to the US results, we see that the productivity, scale, variety, and fixed cost effects contribute significantly to the welfare change in Japan. Despite the negative terms of trade (-$1858 million), variety (-$2671 million) and fixed cost effects (-$2629 million), the positive productivity ($2632 million) and scale effects ($4508 million) increase the welfare in Japan. Even though Japan benefits from expanding US varieties, the loss in domestic and ROW varieties dominate the variety effect.

The welfare loss in the ROW is higher compared to the perfect and monopolistic competition cases. This is mostly due to the bigger negative impact of lost varieties. There is a relatively small increase in the aggregate productivity of the manufacturing industry in the ROW which brings about a modest improvement in the overall welfare ($400 million) along with the scale effect ($356). However, the variety effect in the ROW is larger and negative (-$1779 million). It is largely driven by the declining varieties sourced from the US Even though the number of Japanese varieties increases in the ROW, the drop in US varieties accompanied by the loss in domestic varieties dominate the variety effect. This is mostly dictated by the loss of intermediate inputs used by ROW firms.

This illustrative experiment demonstrates that the additional mechanisms of trade-induced welfare changes under firm heterogeneity lead to different welfare implications of the tariff cut scenario. Particularly, scale and productivity effects are observed to be significant sources of welfare gain.

6.6 Potential issues with large-scale models

In this study we used a stylized model with an aggregated data base and assumed that only the manufacturing industry is characterized by firm heterogeneity. Practical policy analysis will require working with a disaggregated data base with multiple heterogeneous sectors. There are several potential issues that will need to be addressed for numerical implementation of large-scale firm heterogeneity models.

Endogenous free entry and exit condition in a multi-sector, multi-region Melitz model that includes intermediate inputs may give rise to computational difficulties due to multiple corner equilibria (Zhai, 2008). Numerical implementations of such models are also associated with increased dimensionality and potential nonconvexities (Balistreri and Rutherford, 2013). The choice of computing environment, e.g. GAMS or GEMPACK, can also be a factor that could impose practical limitations on implementation depending on the use of non-linear vs. linear percentage change representations of governing equations (Dixon, Jerie, and Rimmer, 2015).

The issues of corner solutions have been remedied in Zhai (2008) by assuming no entry or exit of firms. We believe that the entry-exit feature is an important aspect of firm heterogeneity framework that needs to be included in the model. An alternative approach to avoid the corner solutions could be to restrict the mobility of some primary factors of production. The dimensionality and non-convexity issues have been addressed in Balistreri and Rutherford (2013) by developing a decomposition algorithm that iterates between a partial equilibrium Melitz model and a general equilibrium Armington model. We instead use the linear percentage change representation of governing equations in GEMPACK that is suggested to allow solution of large-scale Melitz models with relative ease and without resorting to decomposition (Dixon, Jerie, and Rimmer, 2015).

It should also be noted that computational resources required to solve the GTAP firm-heterogeneity model can be significantly more than those necessary for the standard GTAP model. This is due to the additional dimension introduced in data arrays by sourcing imports to agents. Since the model size in memory is increased as the data is further disaggregated, the availability of computational resources including memory and processing power may become a critical limiting factor for studies that employ high level of sectoral disaggregation.

7. Concluding remarks

While traditional CGE models based on the Armington assumption of national product differentiation have been successfully applied to various policy scenarios, they also have significant limitations in explaining the firm-level information prevalent in the recent international trade literature. The pioneering work of Melitz (2003) has provided a firm heterogeneity theory that can help address the shortcomings of Armington-based CGE models by introducing additional productivity mechanisms and extensive margin effects. Incorporation of firm heterogeneity in mainstream CGE models offers great potential to improve computational policy analysis.

This paper presents the implementation of firm heterogeneity theory in the GTAP model and illustrates the behavioral characteristics of this theory in a stylized tariff removal scenario whereby Japanese tariffs on US manufactures are eliminated. Results are compared across different model specifications such as monopolistic competition based on Krugman (1980) and perfect competition based on the Armington (1969) assumption.

We observe that productivity threshold for the US-Japan export market reduces mostly due to the reduction in fixed export costs per sale. This scale effect is the dominant factor in the reduction of the productivity threshold and the subsequent increase in the number of US manufacturing firms exporting to Japan. On the other hand, the reallocation in the domestic market is such that the lowest-productivity firms are forced to exit due to higher competition. As lowest-productivity firms exit the industry and high-productivity firms expand into export markets, the average productivity in the US manufacturing industry rises.

Welfare results under firm-heterogeneity shows significant gains from higher firm scale and increased industry productivity which lead to more pronounced welfare responses compared to the monopolistic and perfect competition models. We find that these positive effects offset the welfare losses that arise due to lost domestic varieties and fixed cost payments. It is important to note that the scale effect is found to be more dominant than the endogenous productivity effect in determining the welfare implications of the tariff removal scenario under firm heterogeneity.

The GTAP firm-heterogeneity model is a powerful tool for trade policy analysis with improved abilities in tracing out productivity changes and entry/exit of firms following trade liberalization episodes. By making the cutting edge trade theory more accessible for computational policy analysis, CGE models incorporating firm heterogeneity will enable exploration of the previously unobserved effects of trade agreements and will expand the scope of international trade policy analysis.

Acknowledgements

Earlier versions of this work were presented at the 17th and 18th Annual Conferences on Global Economic Analysis and 22nd and 23rd Annual Short Courses in Global Trade Analysis. We would like to thank the participants at these events for invaluable feedback and discussion. We thank Peter Dixon, Fan Zhai, Marinos Tsigas, Kazuhiko Oyamada, Jayson Beckman, and Shawn Arita for helpful suggestions and detailed comments. We thank Altug Ozcelikkale for his contributions to the graphical expositions in this manuscript.

References

Akgul, Z., N.B. Villoria, and T.W. Hertel. 2015. "Theoretically-consistent parameterization of a multi-sector global model with heterogeneous firms." 18th Annual Conference on Global Economic Analysis. https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=4731.

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare. 2012. "New trade models, same old gains?" *American Economic Review*, 102(1): 94–130. doi:10.3386/w15628.

Arkolakis, C., S. Demidova, P.J. Klenow, and A. Rodriguez-Clare. 2008. "Endogenous variety and the gains from trade." *American Economic Review*, 98(2): 444–450. doi:10.3386/w13933.

Armington, P.S. 1969. "Theory of demand for products distinguished by place of production." *International Monetary Fund Staff Papers*, 16(1): 159–178. doi:10.2307/3866403.

Axtell, R.L. 2001. "Zipf distribution of US firm sizes." *Science*, 293(5536): 1818–1820. doi:10.1126/science.1062081.

Balistreri, E.J., R.H. Hillberry, and T.F. Rutherford. 2011. "Structural estimation and solution of international trade models with heterogeneous firms." *Journal of International Economics*, 83(2): 95–108. doi:10.2139/ssrn.1153944.

Balistreri, E.J., and T.F. Rutherford. 2013. "Computing general equilibrium theories of monopolistic competition and heterogeneous firms." In *Handbook of Computable General Equilibrium Modeling*, edited by P. B. Dixon and D. W. Jorgenson.

Bernard, A.B., J. Eaton, J.B. Jensen, and S. Kortum. 2003. "Plants and productivity in international trade." *American Economic Review*, 93(4): 1268–1290. doi:10.3386/w7688.

Bernard, A.B., and J.B. Jensen. 1999. "Exceptional exporter performance: cause, effect, or both?" *Journal of International Economics*, 47(1): 1–25. doi:10.3386/w6272.

Bernard, A.B., J.B. Jensen, S.J. Redding, and P.K. Schott. 2007. "Firms in international trade." *Journal of Economic Perspectives*, 21(3): 105–130. doi:10.3386/w13054.

Bernard, A.B., J.B. Jensen, and P.K. Schott. 2006. "Trade costs, firms and productivity." *Journal of Monetary Economics*, 53(5): 917–937. doi:10.1016/j.jmoneco.2006.05.010.

Chaney, T. 2008. "Distorted gravity: the intensive and extensive margins of international trade." *American Economic Review*, 98(4): 1707–1721. doi:10.1257/aer.98.4.1707.

Dixit, A., and J. Stiglitz. 1977. "Monopolistic competition and optimum product diversity." *American Economic Review*, 67(3): 297–308. doi:10.1017/cbo9780511492273.004.

Dixon, P.B., M. Jerie, and M.T. Rimmer. 2015. "Modern trade theory for CGE modelling: the Armington, Krugman and Melitz models." *GTAP Technical Paper*, No. 36. doi:10.4337/9781845425647.00019.

Eaton, J., S. Kortum, and F. Kramarz. 2011. "An anatomy of international trade: evidence from French firms." *Econometrica*, 79(5): 1453–1498. doi:10.3386/w14610.

Eaton, J., S. Kortum, and F. Kramarz. 2004. "Dissecting trade: Firms, industries, and export destinations." *American Economic Review*, 94(2): 150–154. doi:10.3386/w10344.

Gohin, A., and T.W. Hertel. 2003. "A note on the CES functional form and its use in the GTAP model." *GTAP Research Memorandum*, No. 2. https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=1370.

Harrison, W.J., and K.R. Pearson. 1996. "Computing solutions for large general equilibrium models using GEMPACK." *Computational Economics*, 9(2): 83–127. doi:10.1007/bf00123638.

Hertel, T.W., ed. 1997. *Global trade analysis: Modeling and applications*. Cambridge University Press.

Hillberry, R., and D. Hummels. 2013. "Trade elasticity parameters for a computable general equilibrium model." In *Handbook of Computable General Equilibrium Modeling*, edited by B. D. Peter and W. J. Dale. Elsevier, vol. 1, book section 18, pp. 1213–1269.

Hopenhayn, H. 1992. "Entry, exit, and firm dynamics in long run equilibrium." *Econometrica*, 60(5): 1127–1150. doi:10.2307/2951541.

Horridge, M., K.R. Pearson, and T.F. Rutherford. 2013. *Handbook of computable general equilibrium modeling*, Elsevier, chap. Software Platforms: GAMS and GEMPACK. pp. 1331–1382.

Huff, K.M., and T.W. Hertel. 2000. "Decomposing welfare change in the GTAP model." *GTAP Technical Paper*, No. 5. https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=308.

Hummels, D., and P.J. Klenow. 2005. "The variety and quality of a nation's exports." *American Economic Review*, 95(3): 704–723. doi:10.1257/0002828054201396.

Kancs, D. 2010. "Structural estimation of variety gains from trade integration in Asia." *Australian Economic Review*, 43(3): 270–288. doi:10.1111/j.1467-8462.2010.00606.x.

Krugman, P. 1980. "Scale economies, product differentiation, and the pattern of trade." *American Economic Review*, 70(5): 950–959. doi:10.1016/0022-1996(81)90001-5.

Lanclos, D.K., and T.W. Hertel. 1995. "Endogenous product differentiation and trade policy: implications for the U.S. food industry." *American Journal of Agricultural Economics*, 77(3): 591–601. doi:10.2307/1243227.

McCallum, J. 1995. "National borders matter: canada - US regional trade patterns." *American Economic Review*, 85(3): 615–623. doi:10.1111/1467-9396.00158.

Melitz, M., and S.J. Redding. 2013. "Firm heterogeneity and aggregate welfare." *National Bureau of Economic Research*, Working Paper 18919. doi:10.3386/w18919.

Melitz, M.J. 2003. "The impact of trade on intra-industry reallocations and aggregate industry productivity." *Econometrica*, 71(6): 1695–1725. doi:10.3386/w8881.

Melitz, M.J., and D. Trefler. 2012. "Gains from trade when firms matter." *Journal of Economic Perspectives*, 26(2): 91–118. doi:10.1257/jep.26.2.91.

Narayanan, B., G., A. Aguiar, and R. McDougall, eds. 2012. *Global trade, assistance, and production: the GTAP 8 Data Base*. Center for Global Trade Analysis, Purdue University. https://www.gtap.agecon.purdue.edu/databases/v8/v8_doco.asp.

Obstfeld, M., K. Rogoff, O. Jeanne, and C. Engel. 2001. "The six major puzzles in international macroeconomics: Is there a common cause?" The MIT Press: NBER Macroeconomics Annual, vol. 15, pp. 339–390. http://papers.nber.org/books/bern01-1.

Oyamada, K. 2014. "Behavioral characteristics of applied general equilibrium models with an Armington, Krugman, and Melitz encompassing module." 17th Annual Conference of Global Economic Analysis, Dakar, Senegal. https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=4459.

Spearot, A. 2016. "Unpacking the long run effects of tariff shocks: New structural implications from firm heterogeneity models." *AEJ Microeconomics*, 8(2): 128–67. doi:10.1257/mic.20140015.

Swaminathan, P., and T.W. Hertel. 1996. "Introducing monopolistic competition into the GTAP model." GTAP Technical Paper No. 6. https://www.gtap.agecon.

purdue.edu/resources/res_display.asp?RecordID=309.

Venables, A.J. 1987. "Trade and trade policy with differentiated products: a Chamberlinian - Ricardian model." *The Economic Journal*, 97(387): 700–717. doi:10.2307/2232931.

Zhai, F. 2008. "Armington meets melitz: introducing firm heterogeneity in a global CGE model of trade." *Journal of Economic Integration*, 23(3): 575–604. doi:10.11130/jei.2008.23.3.575.

Appendix A. List of main variables and value flows

The variables and value flows used in this manuscript are listed in Table A.1 along with their definitions.

Table A.1. List of main variables and value flows.

Group	Variable	Definition
Price Variables	`ps(j,r)`	supply price of commodity j in region r
	`avc(j,r)`	average variable cost of production in industry j in region r
	`pf(i,j,r)`	firms' price for intermediate input i for use by j in r
	`pfe(i,j,r)`	firms' price for endowment commodity i in ind. j, region r
	`pva(j,r)`	price of value-added composite of industry j in region r
	`pcgds(r)`	price of investment goods
	`pm(j,r)`	market price of commodity j in region r
	`psave(r)`	price of savings in region r
	`pt(j)`	price of composite margins services, type
	`pfob(j,r,s)`	FOB world price of commodity j supplied from r to s
Quantity Variables	`qo(j,r)`	output of industry j in region r
	`qf(i,j,r)`	demand for intermediate input i for use by j in region r
	`qfe(i,j,r)`	demand for endowment i for use in ind. j in region r
	`qvaf(j,r)`	demand for fixed value-added in industry j in r
	`qvafe(j,r)`	demand for value-added in industry j in r to cover fixed set-up costs
	`qvafs(j,r,s)`	demand for value-added in industry j in r to cover fixed trading costs
	`qvav(j,r)`	demand for variable value-added in industry j in r
	`qof(j,r)`	quantity of output per firm in industry j of region r
	`qs(j,r,s)`	sales of commodity j from source r to destination s
	`qppc(j,s,r)`	private hhld demand in region r for homogeneous commodity j sourced from s
	`qpmc(j,s,r)`	private hhld demand in region r for differentiated commodity j sourced from s
	`qgpc(j,s,r)`	gov't demand in region r for homogeneous commodity j sourced from s
	`qgmc(j,s,r)`	gov't demand in region r for differentiated commodity j sourced from s
	`qfpc(j,i,s,r)`	industry i of region r's demand for homogeneous commodity j sourced from s
	`qfmc(j,i,s,r)`	industry i of region r's demand for differentiated commodity j sourced from s
	`vp(j,s,r)`	number of varieties of j sourced from s available to priv hhld in r
	`vg(j,s,r)`	number of varieties of j sourced from s available to government in r
	`vf(j,s,r)`	number of varieties of j sourced from s available to firms in r
	`np(j,r)`	number of potential firms in industry j of region r
	`ns(j,r,s)`	number of firms in industry j of region r that are active in market s
	`kb(r)`	beginning-of-period capital stock in r
Technology Parameters	`ao(j,r)`	output augmenting technical change in sector j of r
	`af(j,r)`	composite intermed. input i augmenting tech change by j of r
	`afe(i,j,r)`	primary factor i augmenting tech change by j of r
	`avafe(j,r)`	tech change in fixed set-up costs of j in r
	`avafs(j,r,s)`	tech change in fixed trading costs of j from r to s
	`avav(j,r)`	variable value-added augmenting tech change in sector j of r
	`ava(j,r)`	value-added augmenting tech change in sector j of r
	`atmfsd(m,j,r,s)`	tech change in m's shipping of j from region r to s
	`ams(j,r,s)`	import j from region r augmenting tech change in region s

(Continued)

Table A.1. List of main variables and value flows. (Continued)

Group	Variable	Definition
Preference Parameters	dppriv(r)	private consumption distribution parameter
	dpgov(r)	government consumption distribution parameter
	dpsave(r)	saving distribution parameter
	pop(r)	regional population
Value Flows	VOA(j,r)	value of commodity j output in region r at agent's prices
	VC(j,r)	variable cost in the production of the monop. comp. commodity j in region r
	VFA(i,j,r)	producer expenditure on input i by j in r valued at agent's prices
	VA(j,r)	value added in activity j in region r
	VAF(j,r)	fixed value added demanded by the monop. comp. industry j in region r
	VAFE(j,r)	fixed set-up costs of industry j in region r
	VAFS(j,r,s)	fixed trade costs of industry j in r to enter market s
	VAV(j,r)	variable value added demanded by the monop. comp. industry j in region r
	VPAS(j,s,r)	private hhld consumption expenditure in r for product j sourced from s
	VGAS(j,s,r)	government consumption expenditure in r for product j sourced from s
	VFAS(j,i,s,r)	purchases of intermediate input j sourced from s for use by industry i in region r
	VSWD(j,r,s)	value of sales of j from r to s, at world (fob) prices
	VTMD(j,s)	aggregate value of svces j in shipments to s
	VST(j,r)	exports of j from r for int'l trnsport valued at mkt price (tradeables only)
	VTMFSD(m,i,r,s)	international margin usage, by margin, freight, source, and destination
	VDEP(r)	value of capital depeciation in r (exogenous)
	NETINV(r)	regional NET investment in region r
	SAVE(r)	expenditure on NET savings in region r valued at agent's prices
	PTAX(j,r)	output tax on good j in region r
	SPTAX(j,s,r)	tax on private consumption in r of good i from source s
	SGTAX(j,s,r)	tax on private consumption in r of good i from source s
	SFTAX(j,i,s,r)	tax on use of intermediate input j from source s by industry i in r
	STAX(j,r,s)	tax on sales of good j from source r to destination s
	DTAX(j,s,r)	tax on demand for good j from source s to destination r
	ETAX(i,j,r)	tax on use of endowment good i by industry j in region r
Others	UTILPRIV(r)	utility from private hhld consumption
	UTILGOV(r)	utility from government consumption
	UTILSAVE(r)	utility from saving, for EV calcs
	UTILPRIVEV(r)	utility from private hhld consumption, for EV calc.
	UTILGOVEV(r)	utility from government consumption, for EV calcs
	UTILSAVEEV(r)	utility from saving, for EV calcs
	UTILELASEV(r)	elasticity of cost of utility wrt utility, for EV calc.
	INCOMEEV(r)	regional income, for EV calc.
	DPARPRIV(r)	private consumption distribution parameter
	DPARGOV(r)	government consumption distribution parameter
	DPARSAVE(r)	saving distribution parameter
	EVSCALFACT(r)	equivalent variation scaling factor

Source: Author calculations.

Appendix B. Data description and transformation

In the monopolistic competition model imports are sourced by agent as mentioned in the previous sections. The structure of the standard GTAP Data Base is not compatible with sourced imports. Therefore, we transform the standard GTAP Data Base following Swaminathan and Hertel (1996). This section outlines the steps for this data transformation. For more details, we refer the reader to Swaminathan and Hertel (1996).

There are three steps to generate the monopolistically competitive data base:

- Sourcing agent demand at market prices
- Sourcing agent demand at agents prices
- Trade data

B.1 Sourced imports at market prices

In the standard GTAP Data Base, consumption expenditure on domestic and imported products are given separately. For instance, the private household consumption expenditure is `VDPM(i,s)` for domestic goods and `VIPM(i,s)` for imported goods. The first step is to transform agents' domestic and import demands into sourced demands valued at market prices. Share of imports from a particular source country in all imports of the destination country is applied to value of agent purchases. Let $MSHRS(i,s,r)$ be the market share of source s in total imports of i by region r which is calculated as follows:

$$MSHRS(i,s,r) = \frac{VIMS(i,s,r)}{\sum_k VIMS(i,k,r)} \tag{B.1}$$

where $VIMS(i,s,r)$ is the value of imports of i by source s to destination r. Applying this share to agent purchases yields the consumption of imports of i from source s to destination r by agent. For instance, for the private household, we use VIPM(i,r) and the import share MSHRS(i,s,r) to generate VPMS(i,s,r). If the source region, s, is the same as the destination region, r, domestic sales are taken into account as well as the intra-regional imports. An example for private household is given as follows:

$$VPMS(i,s,r) = MSHRS(i,s,r) * VIPM(i,r), \; for\, s \neq r \tag{B.2}$$

$$VPMS(i,s,r) = MSHRS(i,s,r) * VIPM(i,r) + VDPM(i,s,r), \; for\, s = r \tag{B.3}$$

As a result, agents' domestic and import demands, i.e. `VDPM(i,r)` and `VIPM(i,r)`, are replaced by sourced demands, `VPMS(i,s,r)`. The change in GTAP notation is outlined in Figure B.1.

B.2 Sourced imports at agent's prices

The second step is to generate the sourced import demands valued at agents' prices. Sourced imports at market prices have already been obtained in step one.

Value flows at market prices will be used to generate sourced imports at agents' prices by using the power of average (ad volarem) tax on total demand by an agent ($TP(i,r)$, $TG(i,r)$, and $TF(i,j,r)$). The formula to calculate the power of the tax for private household is as follows:

$$TP(i,r) = \frac{VIPA(i,r) + VDPA(i,r)}{VIPM(i,r) + VDPM(i,r)} \tag{B.4}$$

The same method is used for private households, government and firm intermediate input demands. To obtain the sourced purchases at agents' prices, $TP(i,r)$ is applied to $VPMS(i,s,r)$ as follows:

$$VPAS(i,s,r) = TP(i,r) * VPMS(i,s,r) \tag{B.5}$$

As a result, agents' domestic and import demands, i.e. `VDPA(i,r)` and `VIPA(i,r)`, are replaced by sourced demands, `VPAS(i,s,r)`. The data transformation in this step is summarized in Figure B.1.

B.3 Trade data

The third step is to generate the trade data. Trade data does not go through sourcing since it is already sourced. There are just two changes: (a) notation (exports and imports are renamed as "sales" and "demands" respectively), and (b) inclusion of domestic sales to ensure market equilibrium (for r = s, aggregate domestic sales and intra-regional imports are both taken into account). The following formulas are used for sales:

$$VSMD(i,r,s) = VXMD(i,r,s),\ for\ r \neq s \tag{B.6}$$

$$VSMD(i,r,s) = VXMD(i,r,s) + VDM(i,s),\ for\ r = s \tag{B.7}$$

where $VDM(i,r)$ is the value of aggregate domestic sales of i in r at market prices:

$$VDM(i,r) = VDPM(i,r) + VDGM(i,r) + \sum_j VDFM(i,j,r) \tag{B.8}$$

The following formulas are used for imports:

$$VDMS(i,s,r) = VIMS(i,s,r),\ for\ s \neq r \tag{B.9}$$

$$VDMS(i,s,r) = VIMS(i,r,r) + VDM(i,s),\ for\ s = r \tag{B.10}$$

Swaminathan and Hertel (1996) note that there are hardly any consumption tax on domestic demand which allows the addition of domestic sales into value flows for exports and imports when $r = s$. However, they highlight the fact that if domestic sales are very large relative to intra-regional trade, then intra-regional trade may be distorted.

The same transformation is done for export and import flows at world prices. The data transformation in this step is summarized in Figure B.1.

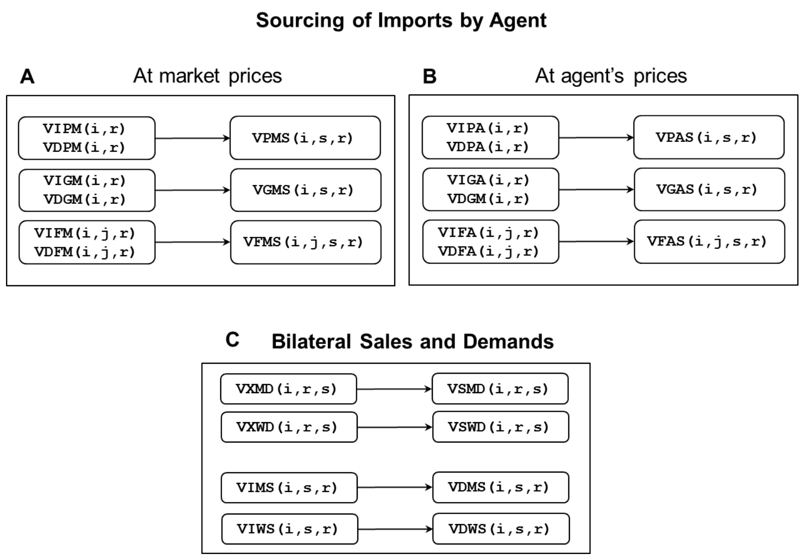

Figure B.1. Transformation of the data base: Sourcing imports by agent.

Notes: `i` ∈ `TRAD_COMM`, `j` ∈ `PROD_COMM`, and `r,s` ∈ `REG`.

Source: Author calculations.

Appendix C. Mathematical derivations

C.1 Productivity threshold

The profit of the marginal firm is given by Equation (14):

$$\Pi^*_{irs} = \frac{P^*_{irs}}{T_{irs}} Q^*_{irs} - \frac{C_{ir}}{\Phi^*_{irs}} Q^*_{irs} - W_{irs} F_{irs} \tag{C.1}$$

We solve $\Pi^*_{irs} = 0$ for Φ^*_{irs} to obtain the productivity threshold. Rearranging Equation (C.1) yields:

$$\frac{P^*_{irs}}{T_{irs}} Q^*_{irs} - \frac{C_{ir}}{\Phi^*_{irs}} Q^*_{irs} = W_{irs} F_{irs} \tag{C.2}$$

where we substitute the optimal demand for the marginal firm's product by using a version of Equation (4). Note that

$$\widetilde{Q}^*_{irs} = Q_{is} \left[\frac{P_{is}}{\widetilde{P}^*_{irs}} \right]^{\sigma_i} = Q_{irs} \left[\frac{P_{irs}}{\widetilde{P}^*_{irs}} \right]^{\sigma_i} \tag{C.3}$$

Substituting Equation (C.3) into Equation (C.2) and rearranging we obtain:

$$P^{*\;1-\sigma_i}_{irs} \frac{P_{irs}{}^{\sigma_i} Q_{irs}}{T_{irs}} - P^{*\;-\sigma_i}_{irs} P_{irs}{}^{\sigma_i} Q_{irs} \frac{C_{ir}}{\Phi^*_{irs}} = W_{irs} F_{irs} \tag{C.4}$$

Via the marginal firm version of the optimal price in Equation (10), (C.4) reduces to

$$\frac{1}{\sigma_i} \left[\frac{\sigma_i}{\sigma_i - 1} \frac{C_{ir} T_{irs}}{\Phi^*_{irs}} \right]^{1-\sigma_i} \frac{P_{irs}{}^{\sigma_i} Q_{irs}}{T_{irs}} = W_{irs} F_{irs} \tag{C.5}$$

Finally we solve Equation (C.5) for Φ^*_{irs} which yields Equation (15) as:

$$\Phi^*_{irs} = \frac{\sigma_i{}^{\frac{\sigma_i}{\sigma_i - 1}}}{\sigma_i - 1} \frac{C_{ir}}{P^*_{ir}} \left[\frac{P^*_{irs}}{T_{irs} W_{irs}} \frac{Q^*_{irs}}{F_{irs}} \right]^{\frac{1}{1-\sigma_i}} \tag{C.6}$$

where $P^*_{ir} = \frac{P^*_{irs}}{T_{irs}}$.

C.2 Zero profits condition

This section provides the derivation of the zero profits condition in the monopolistically competitive industry with heterogeneous firms. Total cost in the monopolistically competitive industry j of region r, $VOA(j, r)$, is composed of variable costs, $VC(j,r)$, and fixed costs, $VAF(j,r)$, as follows:

$$VOA(j,r) = VC(j,r) + VAF(j,r) \tag{C.7}$$

Value flows in Equation (C.7) can be written as products of associated prices and quantities as follows:

$$PS(j,r) * QO(j,r) = AVC(j,r) * QO(j,r) + PVAF(j,r) * QVAF(j,r) \tag{C.8}$$

Total differentiation of (C.8) yields:

$$\begin{aligned} VOA(j,r) * [ps(j,r) + qo(j,r)] = {} & VC\,(j,r) * [avc(j,r) + qo(j,r)] \\ & + VAF\,(j,r) * [pvaf(j,r) + qvaf(j,r)] \end{aligned} \tag{C.9}$$

where the lowercase letters denote percentage changes in the corresponding uppercase variables. Moving `qo(j,r)` to the right-hand side of (C.9) yields:

$$\begin{aligned} VOA(j,r) * ps(j,r) = {} & VC\,(j,r) * avc(j,r) + [VC(j,r) - VOA(j,r)] * qo(j,r) \\ & + VAF\,(j,r) * [pvaf(j,r) + qvaf(j,r)] \end{aligned} \tag{C.10}$$

Using (C.7), we can rewrite (C.10) as:

$$\begin{aligned} VOA(j,r) * ps(j,r) = {} & VC\,(j,r) * avc(j,r) - VAF(j,r) * qo(j,r) \\ & + VAF\,(j,r) * [pvaf(j,r) + qvaf(j,r)] \end{aligned} \tag{C.11}$$

Note that average variable cost is determined by Equation AVERAGEVC as follows:

$$\begin{aligned} VC\,(j,r) * avc\,(j,r) = {} & \sum_{i \in TRAD\_COMM} VFA\,(i,j,r) * [pf\,(i,j,r) - af\,(i,j,r)] \\ & + VAV\,(j,r) * [pvav\,(j,r) - avav\,(j,r)] - VC\,(j,r) * ao(j,r) \end{aligned} \tag{C.12}$$

Substituting (C.12) into (C.11) shows the effect of input prices and technical change on supply prices. This substitution yields:

$$\begin{aligned} VOA\,(j,r) * ps(j,r) = {} & \sum_{i \in TRAD\_COMM} VFA\,(i,j,r) * [pf\,(i,j,r) - af\,(i,j,r)] \\ & + VAV\,(j,r) * [pvav\,(j,r) - avav\,(j,r)] - VC\,(j,r) * ao(j,r) \\ & - VAF(j,r) * qo(j,r) \\ & + VAF\,(j,r) * [pvaf(j,r) + qvaf(j,r)] \end{aligned} \tag{C.13}$$

Rearranging (C.13) we have:

$$\begin{aligned} VOA(j,r) * ps(j,r) = & \sum_{i \in TRAD\_COMM} VFA(i,j,r) * [pf(i,j,r) - af(i,j,r)] \\ & + VAV(j,r) * pvav(j,r) + VAF(j,r) * pvaf(j,r) \\ & - VAV(j,r) * avav(j,r) - VC(j,r) * ao(j,r) \\ & - VAF(j,r) * qo(j,r) + VAF(j,r) * qvaf(j,r) \end{aligned} \tag{C.14}$$

Note that price of value-added composite, $pva(j,r)$, is a share-weighted summation of the price of variable value-added composite, $pvav(j,r)$, and the price of fixed value-added composite, $pvaf(j,r)$. This is given as follows:

$$VA(j,r) * pva(j,r) = VAV(j,r) * pvav(j,r) + VAF(j,r) * pvaf(j,r) \tag{C.15}$$

We substitute (C.15) into (C.14) to replace value-added prices:

$$\begin{aligned} VOA\,(j,r) * ps(j,r) = & \sum_{i \in TRAD\_COMM} VFA\,(i,j,r) * [pf\,(i,j,r) - af\,(i,j,r)] \\ & + VA\,(j,r) * pva\,(j,r) - VAV\,(j,r) * avav\,(j,r) \\ & - VAF(j,r) * qo(j,r) + VAF\,(j,r) * qvaf(j,r) \\ & - VC\,(j,r) * ao(j,r) \end{aligned} \tag{C.16}$$

Note that demand for fixed value-added composite, $qvaf(j,r)$, is determined by a share-weighted summation of $qvafe(j,r)$ and $qvafs(j,r,s)$ as explained in Appendix C.4. This is governed by Equation VADEMAND as follows:

$$VAF(j,r) * qvaf(j,r) = VAFE(j,r) * qvafe(j,r) + \sum_{s=REG} VAFS(j,r,s) * qvafs(j,r,s) \tag{C.17}$$

Substitution of (C.17) into (C.16) yields:

$$\begin{aligned} VOA\,(j,r) * ps(j,r) = & \sum_{i \in TRAD\_COMM} VFA\,(i,j,r) * [pf\,(i,j,r) - af\,(i,j,r)] \\ & + VA\,(j,r) * pva\,(j,r) - VAV\,(j,r) * avav\,(j,r) \\ & - VAF(j,r) * qo(j,r) \\ & + VAFE(j,r) * qvafe(j,r) + \sum_{s=REG} VAFS(j,r,s) * qvafs(j,r,s) \\ & - VC\,(j,r) * ao(j,r) \end{aligned} \tag{C.18}$$

As discussed in Appendix C.4, demand for value-added in fixed costs is proportional to firm numbers which is governed by the following two equations:

$$qvafe\,(j,r) = np\,(j,r) \tag{C.19}$$

$$qvafs\,(j,r,s) = ns\,(j,r,s) \tag{C.20}$$

Substitution of (C.19) and (C.20) into (C.18) yields:

$$\begin{aligned} VOA\,(j,r) * ps(j,r) = & \sum_{i \in TRAD\_COMM} VFA\,(i,j,r) * [pf\,(i,j,r) - af\,(i,j,r)] \\ & + VA\,(j,r) * pva\,(j,r) - VAV\,(j,r) * avav\,(j,r) \\ & - VAF(j,r) * qo(j,r) \\ & + VAFE(j,r) * np(j,r) + \sum_{s=REG} VAFS(j,r,s) * ns(j,r,s) \\ & - VC\,(j,r) * ao(j,r) \end{aligned} \tag{C.21}$$

Note that industry output is determined by per firm output and the number of producers in the industry. This is governed by:

$$qo\,(j,r) = qof\,(j,r) + ns\,(j,r,r) \tag{C.22}$$

Substituting (C.22) into (C.21) yields:

$$\begin{aligned} VOA\,(j,r) * ps(j,r) = & \sum_{i \in TRAD\_COMM} VFA\,(i,j,r) * [pf\,(i,j,r) - af\,(i,j,r)] \\ & + VA\,(j,r) * pva\,(j,r) - VAV\,(j,r) * avav\,(j,r) \\ & - VAF(j,r) * qof(j,r) - VAF(j,r) * ns(j,r,r) \\ & + VAFE(j,r) * np(j,r) + \sum_{s=REG} VAFS(j,r,s) * ns(j,r,s) \\ & - VC\,(j,r) * ao(j,r) \end{aligned} \tag{C.23}$$

As explained in Section 3.2.5, $ns(j,r,s)$ is determined by Equation NSFIRM. This is given by:

$$ns(j,r,s) = np(j,r) - SHAPE(j) * aost(j,r,s). \tag{C.24}$$

Substituting Equation (C.24) into Equation (C.23) yields:

$$\begin{aligned} VOA\,(j,r) * ps(j,r) = & \sum_{i \in TRAD\_COMM} VFA\,(i,j,r) * [pf\,(i,j,r) - af\,(i,j,r)] \\ & + VA\,(j,r) * pva\,(j,r) - VAV\,(j,r) * avav\,(j,r) \\ & - VAF(j,r) * qof(j,r) - VAF(j,r) * ns(j,r,r) \\ & + VAFE(j,r) * np(j,r) \\ & + \sum_{s \in REG} VAFS(j,r,s) * [np(j,r) - SHAPE(j) * aost(j,r,s)] \\ & - VC\,(j,r) * ao(j,r) \end{aligned} \tag{C.25}$$

Note that $[VAFE(j,r) + \sum_{s=REG} VAFS(j,r,s)] * np(j,r) = VAF(j,r)np(j,r)$. Substituting this into (C.25) and rearranging we obtain:

$$\begin{aligned} VOA\,(j,r) * ps(j,r) = & \sum_{i \in TRAD\_COMM} VFA\,(i,j,r) * [pf\,(i,j,r) - af\,(i,j,r)] \\ & + VA\,(j,r) * pva\,(j,r) - VAV\,(j,r) * avav\,(j,r) \\ & - VAF(j,r) * qof(j,r) - VAF(j,r) * ns(j,r,r) \\ & + VAF(j,r) * np(j,r) \\ & - SHAPE(j) \sum_{s \in REG} VAFS(j,r,s) * aost(j,r,s) \\ & - VC\,(j,r) * ao(j,r) \end{aligned} \tag{C.26}$$

Finally, note that based on Equation (C.24) we can write

$$VAF(j,r) * [ns(j,r,r) - np(j,r)] = -SHAPE(j) * VAF(j,r) * aost(j,r,r)$$

Plugging this into Equation (C.26) yields:

$$\begin{aligned} VOA\,(j,r) * ps(j,r) = & \sum_{i \in TRAD\_COMM} VFA\,(i,j,r) * [pf\,(i,j,r) - af\,(i,j,r)] \\ & + VA\,(j,r) * pva\,(j,r) - VAV\,(j,r) * avav\,(j,r) \\ & - VAF(j,r) * qof(j,r) \\ & + SHAPE(j) * VAF(j,r) * aost(j,r,r) \\ & - SHAPE(j) \sum_{s \in REG} VAFS(j,r,s) * aost(j,r,s) \\ & - VC\,(j,r) * ao(j,r) \end{aligned} \tag{C.27}$$

Let the scale constant average total cost, $scatc(j,r)$, be defined as:

$$\begin{aligned} VOA\,(j,r) * scatc(j,r) = & \sum_{i \in TRAD\_COMM} VFA\,(i,j,r) * [pf\,(i,j,r) - af\,(i,j,r)] \\ & + VA\,(j,r) * pva\,(j,r) - VAV\,(j,r) * avav\,(j,r) \\ & + SHAPE(j) * VAF(j,r) * aost(j,r,r) \\ & - SHAPE(j) \sum_{s \in REG} VAFS(j,r,s) * aost(j,r,s) \\ & - VC\,(j,r) * ao(j,r) \end{aligned} \tag{C.28}$$

We substitute Equation (C.28) into Equation (C.27) to obtain the zero profits condition in the industry as given in Equation `ZEROPROFITSMC` in Section 3.2.5:

$$VOA\,(j,r) * ps(j,r) = VOA\,(j,r) * scatc(j,r) - VAF(j,r) * qof(j,r) \tag{C.29}$$

C.3 Calibration of fixed costs

We use bilateral trade flows information in order to derive Equation (36). As described in Section 4.2, we first substitute the optimal demand of the average firm,

$\widetilde{Q}_{irs}$ (Equation (4)), into bilateral trade flows, $\frac{N_{irs}\widetilde{P}_{irs}\widetilde{Q}_{irs}}{T_{irs}}$. This substitution yields:

$$\frac{N_{irs}\widetilde{P}_{irs}\widetilde{Q}_{irs}}{T_{irs}} = N_{irs}\frac{\widetilde{P}_{irs}}{T_{irs}}Q_{is}\left[\frac{P_{is}}{\widetilde{P}_{irs}}\right]^{\sigma_i} \tag{C.30}$$

We, then, substitute in the optimal price of the average firm, $\widetilde{P}_{irs}$ which is governed by Equation (10), as follows:

$$\frac{N_{irs}\widetilde{P}_{irs}\widetilde{Q}_{irs}}{T_{irs}} = N_{irs}\frac{Q_{is}{P_{is}}^{\sigma_i}}{T_{irs}}\left[\frac{\sigma_i}{\sigma_i - 1}\frac{C_{ir}T_{irs}}{\widetilde{\Phi}_{irs}}\right]^{1-\sigma_i} \tag{C.31}$$

This provides Equation (36) in Section 4.2. We, then, use Equation (20) and Equation (19) to substitute in the productivity threshold, $\widetilde{\Phi}_{irs}$. This substitution yields:

$$\frac{N_{irs}\widetilde{P}_{irs}\widetilde{Q}_{irs}}{T_{irs}} = N_{irs}\frac{Q_{is}{P_{is}}^{\sigma_i}}{T_{irs}}\left[\frac{\sigma_i}{\sigma_i - 1}\frac{C_{ir}T_{irs}}{\frac{\sigma_i}{\sigma_i-1}^{\frac{\sigma_i}{\sigma_i-1}}\frac{C_{ir}}{\widetilde{P}_{ir}}\left(\frac{\widetilde{P}_{irs}}{T_{irs}W_{ir}}\frac{\widetilde{Q}_{irs}}{F_{irs}}\right)^{\frac{1}{1-\sigma_i}}\left(\frac{\gamma_i}{\gamma_i-\sigma_i+1}\right)^{\frac{1}{\sigma_i-1}}}\right]^{1-\sigma_i} \tag{C.32}$$

Several terms cancel out in Equation (C.32). It is further simplified by substituting $\widetilde{P}_{ir} = \frac{\widetilde{P}_{irs}}{T_{irs}}$. This yields:

$$\frac{N_{irs}\widetilde{P}_{irs}\widetilde{Q}_{irs}}{T_{irs}} = N_{irs}W_{ir}F_{irs}\frac{Q_{is}{P_{is}}^{\sigma_i}}{\widetilde{Q}_{irs}\widetilde{P}_{irs}^{\sigma_i}}\frac{\sigma_i\gamma_i}{\gamma_i - \sigma_i + 1} \tag{C.33}$$

Finally, the optimal demand $\widetilde{Q}_{irs}$ is substituted once again to obtain Equation (37). This yields:

$$N_{irs}W_{ir}F_{irs} = \frac{N_{irs}\widetilde{P}_{irs}\widetilde{Q}_{irs}}{T_{irs}}\frac{\gamma_i - \sigma_i + 1}{\sigma_i\gamma_i} \tag{C.34}$$

C.4 Demand for value-added

Since differentiated industries devote resources to fixed costs to adapt their products for new markets, demand for fixed value added is directly proportional to the number of successful firms in the monopolistically competitive market. This applies to both fixed set-up costs and fixed trading costs. Fixed set-up costs are faced by all firms that enter the industry. As new firms enter into the industry (as $np(j,r)$ increases), the need for primary factors increases, and demand for fixed value-added to cover set-up costs, $qvafe(j,r)$, rises. This is implemented as:

```
Equation VAFEDEMAND
# value added demand by the monop. comp. industry for fixed set-up costs #
(all,j,MCOMP_COMM)(all,r,REG)
      qvafe(j,r) = np(j,r);
```

A subset of the producers in the industry self-select into export markets based on their productivity levels and the fixed trading costs they face. Similar to fixed set-up costs, demand for value-added used in fixed trading costs is directly proportional to the number of suppliers in the market. This is implemented as:

```
Equation VAFXDEMAND
# value added demand by the monop. comp. industry for fixed trading costs #
(all,j,MCOMP_COMM)(all,r,REG)(all,s,REG)
      qvafs(j,r,s) = ns(j,r,s);
```

Fixed set-up costs and fixed trade costs are the two components that make up total demand for fixed value-added. To obtain the total demand for fixed value-added we aggregate $qvafe(j,r)$, and $qvafs(j,r,s)$ based on their respective shares in total fixed costs.

```
Equation VAFDEMAND
# demand for fixed value added in the monop. comp. industry j of region r #
(all,j,MCOMP_COMM)(all,r,REG)
      qvaf(j,r)
            = SVAFE(j,r) * qvafe(j,r) + sum(k,REG, SVAFS(j,r,k) * qvafs(j,r,k))
            ;
```

where `SVAFE(j,r)` is the share of fixed set-up cost in total fixed cost and `SVAFS (j,r,k)` is the share of fixed trade cost in total fixed cost.

In addition to the fixed component, there is also a variable component of value-added demand. The derived demand equation for variable value-added in monopolistically competitive industry is similar to that of the perfectly competitive industry and follows from the cost minimization problem in Gohin and Hertel (2003). It is implemented as:

```
Equation VAVDEMAND
# demand for variable value added in the monop. comp. industry j of region r #
(all,j,MCOMP_COMM)(all,r,REG)
      qvav(j,r)
            = - avav(j,r) + qo(j,r) - ao(j,r)
            - ESUBT(j) * [pvav(j,r) - avav(j,r) - ps(j,r) - ao(j,r)];
```

The variable value-added demand in industry *j* of region *r*, $qvav(j,r)$, is proportional to the industry output given the industry productivity level. If firms in the industry become more productive, they use less variable value-added to produce a given level of output. Hence demand for variable value-added declines. The last component is the substitution effect which captures the relative effective price of the variable value-added composite to the unit cost of production in the industry.

Demand for total value added in industry *j* in region *r*, qva(j,r), is a weighted aggregation of variable and fixed value-added demand based on respective weights of variable and fixed value-added in total value-added. It is implemented in the code as:

```
Equation VADEMANDMC
# demand for total value added in the monop. comp. industry j of region r #
(all,j,MCOMP_COMM)(all,r,REG)
      qva(j,r) = SHRVAV(j,r) * qvav(j,r) + SHRVAF(j,r) * qvaf(j,r);
```

where `SHRVAV(j,r)` and `SHRVAF(j,r)` are the respective shares of variable value-added and fixed value-added in total value-added purchases.