

Melitz in GTAP Made Easy: the A2M Conversion Method and Result Interpretation

BY PETER B. DIXON^a, MICHAEL JERIE^b AND MAUREEN T. RIMMER^c

Since the 1970s, Armington has been the workhorse specification of trade in computable general equilibrium (CGE) models. Under Armington, agents substitute between products from different countries. Conceptually, Melitz provides a more attractive approach in which substitution is between products from different firms rather than countries. Other attractive features of Melitz are monopolistic competition and economies of scale from fixed establishment costs for firms and fixed set-up costs on trade links. In this paper, we show how, with little change to existing code, an Armington model can be converted to Melitz by adding a few equations and introducing closure swaps. We apply our Armington-to-Melitz method to the Armington-based Global Trade Analysis Project (GTAP) model to derive GTAP-Armington-to-Melitz (GTAP-A2M). We show how results from a CGE model with Melitz industries can be interpreted via back-of-the-envelope calculations. In simulations of the effects of a tariff imposed by North America on imports of apparel, we find greater welfare losses for exporting regions under Melitz than under Armington principally because contraction of apparel output in these regions under Melitz generates an increase in the cost to their households of domestic apparel. Finally, we review two other Melitz-based versions of GTAP: GTAP-HET and a recently published model by Bekkers and Francois.

JEL codes: C68; D58; F12.

Keywords: Heterogeneous Firms; Armington to Melitz; GTAP-A2M; GTAP-HET; Result Interpretation.

^a Centre of Policy Studies, Victoria University, PO Box 14428, Melbourne, Victoria, 8001, Australia (e-mail: Peter.Dixon@vu.edu.au).

^b Centre of Policy Studies, Victoria University, PO Box 14428, Melbourne, Victoria, 8001, Australia (e-mail: Michael.Jerie@vu.edu.au).

^c Centre of Policy Studies, Victoria University, PO Box 14428, Melbourne, Victoria, 8001, Australia (e-mail: Maureen.Bleazby@vu.edu.au).

1. Introduction

This paper has two principal aims. The first is to describe how an existing Armington-based model such as standard Global Trade Analysis Project (GTAP) can be converted easily to a model with Melitz industries by the addition of some elementary equations and the use of closure swaps. We refer to the conversion method as Armington-to-Melitz, or A2M. The second is to demonstrate how results from a mixed Armington-Melitz model can be explained by back-of-the-envelope calculations that quantify the roles of mechanisms in Melitz beyond those in Armington.

The paper grows out of our book on trade theory in computable general equilibrium models [Dixon, Jerie and Rimmer (hereafter DJR), 2018]. However, here we are more succinct. We are targeting applied modelers who would like to take on board, in a GTAP setting, what Melitz has to offer but who don't have the time or inclination to make a massive investment in learning the technical details of Melitz theory. We aim to deliver something understandable and of practical relevance to people who use GTAP. Consequently, we step back from high theory and set out Melitz intuitively, at a relatively low level of technicality.

The paper is organized as follows. Subsection 1.1 gives a brief introduction to the main ideas in Armington and Melitz. Section 2 sets out Melitz theory and explains the A2M system in the context of the GTAP model. We refer to this as the GTAP-A2M system. Section 3 compares two GTAP simulations with similar shocks, one computed under pure Armington assumptions and the other computed with mixed Armington-Melitz assumptions. The emphasis is on interpreting the results and understanding why they are different. Ours is not the only attempt to provide a Melitz option in GTAP. In section 4, we discuss two other Armington to Melitz conversion approaches: the GTAP-HET model of Akgul et al. (2016) and the parsimonious model of Bekkers and Francois (2018). Concluding remarks are in section 5.

1.1 From Armington to Melitz in CGE modeling

Early economy-wide models treated imported and domestic commodities in the same classification as perfect substitutes. These models include Johansen (1960), which is often recognized as the first computable general equilibrium (CGE) model, and models in the linear-programming tradition such as Evans (1972). With the perfect-substitution assumption it is difficult to explain how imported and domestic varieties can co-exist in the same market. Models incorporating this assumption are susceptible to the flip-flop problem. Without artificial constraints, they produce unrealistically large changes in import shares in response to small tariff changes and unrealistic specialization in the commodity composition of a country's export bundle. Armington (1969, 1970) created a 15-country trade model in which each country produced just one good but consumed

all 15 goods, treating the goods from different countries as imperfect substitutes. Deardorff et al. (1977) and Dixon et al. (1977, 1982) adopted Armington's imperfect-substitution specification in multi-industry CGE models, effectively overcoming the flip-flop problem. Armington quickly became the dominant trade specification in CGE modeling.

Armington is still dominant in CGE modeling. However, its theoretical basis is unattractive: it treats domestic/import substitution at the country level rather than the firm level. Under Armington, Japan produces a single variety of car which from the point of view of consumers in all countries is an imperfect substitute for the single variety produced in Germany. This implies that consumers view Toyota, Mazda and Nissan as perfect substitutes and similarly with VW, Mercedes and BMW. Starting with Krugman (1980), trade theorists have been developing heterogeneous-firm models in which varieties are distinguished by firms rather than countries. In Krugman's model, each firm produces a unique variety under increasing returns to scale arising from fixed setup costs. Industries are monopolistically competitive and consumers have preferences that exhibit love of variety. Melitz (2003) not only includes these features, but also allows for fixed setup costs on trade links and productivity differences across firms. These extra features give endogenous changes in industry productivity levels in response to trade policies. For this reason, Melitz is now attracting considerable attention from CGE modelers. Pioneers in Melitz-style CGE modeling include Zhai (2008), Balistreri et al. (2010, 2011) and Balistreri and Rutherford (2013).

CGE models are about industries rather than firms. So how does a heterogeneous-firm theory such as Melitz fit into a CGE framework? What makes Melitz practical for CGE modeling is that although Melitz theory is developed at the firm level, individual firms can be eliminated. While productivity differs across firms in a Melitz industry, the differences follow exogenously given probability distributions. This allows Melitz theory to be expressed in terms of industries that are completely represented by average or typical firms: productivity movements at the industry level can be known by modeling productivity movements of typical firms.

2. The A2M conversion method applied to GTAP

The GTAP model (Hertel, 1997)⁴ has been in constant use for more than 20 years. Its database is the outcome of an enormous worldwide cooperative effort organized by Tom Hertel and his colleagues at Purdue University. Because GTAP is used intensively for policy analyses it contains many unavoidable complications in the treatments of taxes, tariffs, technologies, trade flows, capital flows and capital accumulation. Thousands of economists are familiar with GTAP and many

⁴ The latest documentation of the GTAP model is Corong *et al.* (2017).

have completed courses on its data, applications and Armington-based theory. In short, there is a huge worldwide investment by economists in GTAP.

In these circumstances, achieving widespread adoption of a Melitz version of GTAP with new theory or data structures has the potential to impose large human-capital costs on the world's community of policy modelers. Thus, in introducing Melitz to GTAP, we looked for a method that would leave the theory and data structures of the standard GTAP model essentially unchanged. We wanted to give users the opportunity to adopt Melitz assumptions in a model consisting of familiar GTAP plus a small number of add-on equations. We designed these add-on equations so that through closure swaps they could be turned on or off to introduce Melitz assumptions for user-chosen industries.

2.1 Preliminary modifications to standard GTAP

It was not possible to avoid entirely modifications to the GTAP model.⁵ We made five, none of which alter the fundamental theory of the standard Armington-based model, and none of which was difficult to implement. The five modifications are as follows.

Modification 1. We eliminated the difference between import-import substitution elasticities (*ESUBM*) and import-domestic substitution elasticities (*ESUBD*), effectively eliminating GTAP's two-tier treatment of substitution between domestic and imported goods. This is necessary for Melitz sectors because in Melitz theory substitutability between any pair of varieties is the same as that between any other pair, irrespective of the varieties' regions of origin. While the two-tier system is effectively eliminated, no substantive alteration is required to the GTAP code. We simply introduced a formula that made the values of *ESUBM* equal to those of *ESUBD*, overruling the initially read values of *ESUBM*.⁶

Modification 2. Standard GTAP includes the variable $ams(c,s,d)$ that allows for input-saving technical change in region *d* in the use of commodity *c* imported from region *s*. There is no corresponding variable that allows for input-saving technical change in the use of domestically-produced commodity *c*. We added such a variable, $ads(c,s)$, taking care to include it where necessary in GTAP equations for:

⁵ We worked with GTAP model 6.2 implemented with database 7.0 (November 2008). We chose the database with 57 commodities and 10 regions.

⁶ Although we could have retained the two-tier difference in substitution elasticities for Armington sectors, we chose not to do so. Working with U.S. data, Feenstra *et al.* (2018) find no statistical difference between import-import and import-domestic substitution elasticities for about 70 per cent of goods. We preferred the lower *ESUBD* values to the *ESUBM* values. High values for Melitz variety elasticities often lead to unrealistically large trade responses to tariff changes. Some of the *ESUBD* values in the standard GTAP parameter file are below 2.5. We adjusted these up to 2.5 to avoid excessive terms-of-trade effects for Armington industries.

the demand by agents in region s for domestically produced and imported commodity c ; the cost to agents in region s of composite units of commodity c ; and definitions of welfare and aggregate technical change. As we will see shortly, we need $ads(c,s)$ as well as $ams(c,s,d)$ to allow for the Melitz theory of love of variety.

Modification 3. In standard GTAP, the cost of transporting a unit of commodity c between regions s and d depends on the quantity transported and the cost of a unit of transport services. Consequently, changes in production costs in region s can affect the ratio of fob to cif prices for c sent from s to d . This is inconsistent with the iceberg view of trade costs favored by Melitz and other theorists. To implement Melitz theory accurately in GTAP, we added code that allows trade costs for selected commodities to be *ad valorem*. For these commodities, fob and cif prices can move by the same percentages. With this additional code, the *ad valorem* treatment for selected commodities can be implemented via closure swaps.⁷

Modification 4. We added several new variables to standard GTAP: $aoMel(c,s)$, $txMel(c,s,d)$, $d_revtxMel(c,s,d)$, $d_revMeltot(c,s)$, $pmarket(c,s,d)$ and $pbundle(c,s)$. The first of these, $aoMel(c,s)$, allows for percentage changes in output per input bundle (to be defined shortly) used in industry c in region s . There is already a variable in standard GTAP [$aoall(c,s)$] that can play this role. Including $aoMel(c,s)$ is mainly a matter of convenience. It gives us a dedicated variable for implementing endogenous movements, consistent with Melitz theory, in total-factor productivity for Melitz industries while leaving $aoall(c,s)$ available for simulating non-Melitz-related productivity shocks. For Armington industries, $aoMel(c,s)$ is exogenous on zero.

The second new variable, $txMel(c,s,d)$, is the percentage change in the power of a tax applying to sales of commodity c flowing from s to d . Standard GTAP includes a variable that could play this role [the negative of $txs(c,s,d)$].⁸ Again, duplication is convenient. It gives us a dedicated variable, a Melitz tax, for implementing the Melitz theory of pricing to market. Along with $txMel(c,s,d)$, we added to standard GTAP associated new variables and equations for the change in the collection of revenue [$d_revtxMel(c,s,d)$] from the Melitz tax on each commodity flow and the change in the total collection of Melitz revenue [$d_revMeltot(c,s)$] on all sales of c from region s . As we will see, if (c,s) is a Melitz industry, then $d_revMeltot(c,s)$ is set exogenously on zero. If (c,s) is an Armington industry, $txMel(c,s,d)$ is exogenous on zero for all d .

⁷ *Ad valorem* is more general than iceberg. Under iceberg, trade costs use up the commodity that is being transported. This is not a sensible assumption. Fortunately, iceberg is not essential to the Melitz story, but *ad valorem* is. In our A2M system for GTAP, we retain GTAP assumptions about the production of transport services, but for Melitz commodities we allow the quantity of required transport to vary with the value rather than quantity of what is being transported.

⁸ $txs(c,s,d)$ is the power of a subsidy rather than a tax.

Standard GTAP includes the variable $pm(c,s)$ defined as the percentage change in the market price of commodity c produced in s . We don't change the mathematical specification of $pm(c,s)$. It continues to be the percentage change in the cost per unit of output of commodity c in region s . However we now refer to $pm(c,s)$ as the percentage change in the *factory price* of (c,s) . We reserve the expression "market price" for the cost just beyond the factory door of commodity (c,s) destined for a specific region d . These destination-specific prices are new variables denoted by $pmarket(c,s,d)$. They differ from factory prices by the variable $txMel(c,s,d)$, that is

$$pmarket(c,s,d) = pm(c,s) + txMel(c,s,d) \quad (1)$$

Having added equation (1) to standard GTAP, we checked every place that pm appeared in the original GTAP code to see if it should remain as a factory price (pm) or be changed to a market price ($pmarket$).

The final new variable, $pbundle(c,s)$, is the percentage change in the cost of the standard bundle of inputs required in industry (c,s) . In the original Melitz theory, each industry used only one input, labor. In this case $pbundle(c,s)$ is the percentage change in the wage rate in region s . In our GTAP-A2M system, $pbundle(c,s)$ is the percentage change in an index of input prices (modified by input-saving technical changes)⁹ with weights reflecting input shares in production in industry (c,s) . With $pm(c,s)$ being the percentage change in the cost per unit of output of commodity c in region s and with $aoMel(c,s)$ being the percentage change in the output per input bundle we arrive at:

$$pm(c,s) = pbundle(c,s) - aoMel(c,s) \quad (2)$$

Modification 5. The database for standard GTAP includes non-zero intra-regional international trade flows, $VXMD(c,s,s)$. These are the result of aggregation of countries to form regions. In creating our A2M system for GTAP, we treat regions as countries. Thus, non-zero values for $VXMD(c,s,s)$ are awkward to handle. We chose to zero them out after adding them to domestic demand, $VMD(c,s)$. This procedure leaves balance conditions intact.

2.2. Additional equations to implement Melitz theory

The aim of this subsection is to set out and explain the equations that were added to the end of the modified GTAP code to form the GTAP-A2M system.

⁹ While fixing ideas, it is easiest to think of $pbundle(c,s)$ in a situation in which there are no changes in input-saving technology: in GTAP terms, ams , ava , afe , af , $aosec$, $aoreg$, $aoall$ and our new variable ads are set on zero. However, movements in these technology variables are taken into account in the determination of $pbundle(c,s)$ with weights reflecting the share in (c,s) 's costs of the input on which the technical change operates.

However, before going to the equations we will make a few preliminary comments to help interpretation.

First, in looking at the add-on equations it is necessary to keep in mind that in each region s the firms in a Melitz c -producing industry have an exogenously given distribution of marginal productivities. By marginal productivity we mean the addition to output that is generated by the addition of one bundle of inputs to current production. Invariably CGE modelers have assumed that marginal productivities form a Pareto distribution:

$$g_{c,s}(\Phi) = \alpha(c,s) * \Phi^{-\alpha(c,s)-1} \quad , \quad \Phi \geq 1 \quad (3)$$

where $g_{c,s}(\Phi)$ is the proportion of c producing firms in s that have marginal productivity level Φ and $\alpha(c,s)$ is a positive parameter. This distribution gives a high density of low-productivity firms and a few high productivity firms which, in the Melitz model, correspond to a high density of small firms and a few large firms. For each firm, marginal productivity is constant. Increasing returns to scale are generated for each firm because of upfront fixed costs.

A second important Melitz concept reflected in the add-on equations is the *typical* c -producing firm that exports from region s to region d (includes s to s). In Melitz, this is a c -producing firm that has the average marginal productivity over all those that send c from s to d . Melitz defines this average in a rather abstract way as a CES function of the marginal productivities of these firms. DJR (2018, page 17) show that the Melitz definition is equivalent to defining a typical c -firm on the s,d -link as one that employs the average number of input bundles (think production workers) in production (not fixed set-up costs) over all of the c -firms servicing the s,d link. For example, if there are 15 c -firms on the s,d link and in aggregate they use 300 bundles in all their production (not just their production for the s,d link), then a typical c -firm on the s,d -link is one that uses 20 bundles in production.

Third, in working through the add-on equations, it is easiest to assume initially that they refer only to a Melitz industry. In our GTAP-A2M code, these equations appear for all commodities/industries. Users of our code can make a choice as to which industries they want to treat as Melitz and which they want to treat as Armington. As explained at the end of this subsection, if c is Armington then the add-on equations are effectively turned off via closure swaps.

Fourth, consistent with the GEMPACK code used in GTAP computations, we express the add-on equations for the A2M system in percentage changes of variables.

The first two add-on equations have already been described. These are (1) and (2). The third add-on equation defines movements in the volume of c sent from s to d in terms of movements in the number of c -firms operating on the link and the volume sent by the typical firm:

$$qs(c,s,d) = numl(c,s,d) + qtypical(c,s,d) \quad (4)$$

where $qs(c,s,d)$ is the percentage change in the quantity of commodity c sent from s to d . In GTAP notation $qs(c,s,s)$ corresponds to $qds(c,s)$ and $qs(c,s,d)$ for $d \neq s$ corresponds to $qxs(c,s,d)$. Even with c being Melitz, these variables continue to be determined in the standard GTAP way as functions of: activity variables in region d ; relative prices of c from different sources; and preference variables.

$numl(c,s,d)$ and $qtypical(c,s,d)$ are Melitz concepts. They are the percentage changes in the number of c -producing firms in s that trade on the s,d -link and the quantity sent on the link by a typical (c,s) firm among those that trade on the link. Both these variables will be determined in later add-on equations.

The fourth add-on equation deals with the Melitz idea of pricing to market. Unlike Armington, in Melitz the price received by producers per unit of c sent from s to d varies across d . This variation reflects differences across destinations in the marginal productivity of the typical firms servicing these destinations. If the group of c -firms in region s that service destination d are typically high productivity firms, then the average price received by these firms per unit of sale will be low. To introduce this idea to GTAP without disrupting its Armington structure we use endogenously determined destination-specific phantom taxes, $txMel(c,s,d)$. In Armington terms, we should think of these taxes as applying at the factory door of industry c in region s . The c,s,d tax will be negative if destination d is serviced by a group of c -firms in s for which the typical firm has high productivity. As already mentioned, revenue collection for c -firms in region s added across destinations d is set at zero. Thus, the introduction of the phantom taxes doesn't require alteration to GTAP's government accounts, national accounts or zero-pure-profit conditions. The sole role of the phantom taxes is to get the GTAP-A2M system to generate prices in destination markets that are consistent with Melitz theory. We achieve this by specifying the phantom taxes according to:

$$txMel(c,s,d) = aoMel(c,s) - ptivity(c,s,d) + f_txMel(c,s,d) \quad (5)$$

where

$ptivity(c,s,d)$ is the percentage change in the marginal productivity of the typical c -producing firm on the s,d -link. To avoid confusion it is worth emphasizing that in Melitz theory changes in the marginal productivity of the typical firm are not caused by productivity changes for any individual firm. Rather they reflect changes in the membership of the group of c -producing firms that are able to trade on the s,d -link.

$f_txMel(c,s,d)$ is a shift variable, which for Melitz industries is set exogenously on zero. In combination with (1) and (2), (5) then implies for Melitz industries that

$$pmarket(c,s,d) = pbundle(c,s) - ptivity(c,s,d) \quad (6)$$

Equation (6) gives us Melitz' pricing to market. Consistent with Melitz theory, (6) implies that the average price received by c-firms in s from their sales to destination d [$p_{market}(c,s,d)$] moves with the marginal cost of the typical c-firm on the s,d link [$p_{bundle}(c,s) - p_{tivity}(c,s,d)$]. This doesn't mean that prices equal marginal costs. Rather, it means that mark-ups over marginal costs are constant.

In Melitz, users in destination d of commodity c from region s exhibit love of variety. They benefit from an increase in the number of varieties embodied in the c,s,d trade flow. More varieties increase the likelihood that users are able to closely match their requirements. The number of varieties is simply the number of c-firms operating on the s,d-link. The higher the substitution elasticity between varieties, the smaller the benefit that users derive from any given percentage increase in variety. The Melitz specification of these ideas is introduced to GTAP-A2M by the add-on equation:

$$as(c,s,d) = \frac{1}{\sigma(c)-1} * numl(c,s,d) + f_as(c,s,d) \quad , \quad \sigma(c) > 1 \quad (7)$$

where $\sigma(c)$, which is always set at a value greater than 1, is the elasticity of substitution between varieties of commodity c. In GTAP notation this corresponds to $ESUBD(c)$, see *Modification 1*.

$as(c,s,d)$ is the percentage change in the (c,s)-augmenting technology or preference variable for all agents in region d. If $as(c,s,d)$ equals one, then there is a one per cent increase in the ability of each unit of c sent from s to d to satisfy the requirements of users in region d, or more formally, if $as(c,s,d) = 1$, then users in d can satisfy any given requirement for c with 1 per cent less c from s while holding constant their use of c from all other sources. In GTAP notation, $as(c,s,d)$ for $d \neq s$ corresponds to $ams(c,s,d)$ and $as(c,s,s)$ corresponds to our new variable $ads(c,s)$.

$f_as(c,s,d)$ is a shift variable, which for Melitz industries is set exogenously on zero.

The most detailed add-on equation and the one requiring most explanation is the definition of the percentage change, $qbundle(c,s)$, in the total number of input bundles used by industry (c,s):

$$\begin{aligned} qbundle(c,s) = & \\ & \frac{\sigma(c)-1}{\sigma(c)} * \sum_d SMV(c,s,d) * \{qs(c,s,d) - p_{tivity}(c,s,d)\} \\ & + \frac{\alpha(c,s) - (\sigma(c)-1)}{\alpha(c,s) * \sigma(c)} * \sum_d SMV(c,s,d) * \{numl(c,s,d) + ff(c,s,d)\} \quad (8) \\ & + \frac{\sigma(c)-1}{\alpha(c,s) * \sigma(c)} * \sum_d SMV(c,s,d) * \{num(c,s) + hf(c,s)\} \\ & + fqbundle(c,s) \end{aligned}$$

where $SMV(c,s,d)$ is the share of industry (c,s) 's total costs represented by the market value of commodity (c,s) sent to d . In GTAP notation, this is $VXMD(c,s,d)/VOM(c,s)$ for $d \neq s$ and $VDM(c,s)/VOM(c,s)$ for $d = s$. The $SMV(c,s,d)$ s sum over d to 1. For Armington industries, this reflects the GTAP assumption that the total market value of sales equals total costs, that is zero pure profits. But what about Melitz industries where we allow the Melitz tax, $txMel(c,s,d)$, to come between costs and market values? As mentioned earlier, for Melitz industries we fix the total collection of Melitz taxes on sales of industry (c,s) at zero. Thus, it remains true that the sum of (c,s) 's sales at market value equals (c,s) 's total costs. In this way, consistent with Melitz theory, we preserve zero-pure-profits in Melitz industries.

$num(c,s)$ is the Melitz variable giving the percentage change in the number of (c,s) firms.

$hf(c,s)$ and $ff(c,s,d)$ are the percentage changes in the Melitz variables for the number of input bundles required to set up a (c,s) firm and the number of input bundles for a (c,s) firm to set up trading on the s,d -link. These will normally be exogenous and shocked only in simulations concerned with the effects of changes in set-up technologies.

$\alpha(c,s)$ is the shape parameter for the Pareto distribution of marginal productivities that we assume for (c,s) firms, see equation (3). It is a positive parameter whose value is greater than $\sigma(c) - 1$.¹⁰

$fqbundle(c,s)$ is a shift variable, which for Melitz industries is set exogenously on zero.

A key assumption in the derivation of (8) is that the input composition of fixed costs in industry (c,s) , that is the inputs required to set up firms and open trade links, is the same as that in current production. This assumption is avoided in GTAP-HET, see section 4. However, it has considerable simplifying benefits. Most obviously, it allows us to use a uniquely defined input bundle for industry (c,s) .

While the derivation of equation (8) is too long for inclusion here¹¹, the ideas behind it are transparent.

¹⁰ If demand for each variety of c on the s,d -link is highly elastic [large $\sigma(c)$] and the distribution of marginal productivities for c firms in s is relatively flat [low $\alpha(c,s)$] implying that there is a significant proportion of high productivity firms, then supply of c on the s,d -link is likely to be dominated by high-productivity, low-price firms each of which makes large sales. Thus, the average size and productivity of (c,s) firms servicing the s,d -link will be large. It can become unbounded if $\sigma(c)$ is sufficiently large relative to $\alpha(c,s)$. The necessary and sufficient condition to avoid this outcome is: $\alpha(c,s) > \sigma(c) - 1$, see DJR [2018, especially the derivation of equation (2.24)].

¹¹ Readers who would like to work through the derivation are referred to chapter 4 in DJR (2018), particularly the derivation of equation (4.71).

The first idea is that under Melitz assumptions, variable costs are the fixed fraction, $(\sigma(c)-1)/\sigma(c)$, of the market value of industry (c,s) 's sales on each link d . For example, if $\sigma(c) = 3$, then 2/3rds of the market value of (c,s) 's sales on the s,d -link is the cost of input bundles required to produce (c,s) 's product for supply on the s,d -link. Thus, the share of industry (c,s) 's total purchase of input bundles (or total costs) accounted for by variable costs to supply the s,d -link is $(2/3)*SMV(c,s,d)$. This explains the first term on the right hand side of (8): if sales on the s,d -link increase by 10 per cent [$qs(c,s,d) = 10$], then on this account industry (c,s) increases its purchase of bundles by $10*(2/3)*SMV(c,s,d)$ per cent, or if production productivity on the s,d -link increases by 10 per cent [$ptivity(c,s,d) = 10$], then on this account, industry (c,s) reduces its purchase of bundles by $10*(2/3)*SMV(c,s,d)$ per cent.

The second term on the right hand side of (8) depends on the idea that in a Melitz model with a Pareto distribution for firm productivities, the share of the market value of industry (c,s) 's sales on the s,d -link accounted for by link set-up costs is $[\alpha(c,s) - (\sigma(c)-1)]/[\alpha(c,s)*\sigma(c)]$. For example, if $\sigma(c) = 3$ and $\alpha(c,s) = 4.5$, then 0.185 of the market value of (c,s) 's sales on the s,d -link is the cost of input bundles required to set up the link. If link set-up requirements increase by 10 per cent either because the number of firms on the link increases by 10 per cent or because link set-up requirements per firm increase by 10 per cent [$numl(c,s,d) = 10$ or $ff(c,s,d) = 10$], then as indicated by the second term on the right hand side of (8), the industry requires an increase in the number of input bundles of $10*0.185*SMV(c,s,d)$ per cent.

With $(\sigma(c)-1)/\sigma(c)$ and $[\alpha(c,s) - (\sigma(c)-1)]/[\alpha(c,s)*\sigma(c)]$ being the fractions of the market value of (c,s) 's sales on the s,d -link accounted for by variable costs and link set-up costs, we can conclude that the remaining fraction, $(\sigma(c)-1)/[\alpha(c,s)*\sigma(c)]$, is a contribution towards the cost of establishing firms. Continuing our previous example, this fraction is 0.148. With zero pure profits in industry (c,s) , the sum of these contribution over d must equal the fixed costs of establishing firms in (c,s) , implying that these fixed costs equal (c,s) 's total costs (or market value of its sales) multiplied by 0.148. Thus, as indicated by the third term on the right hand side of (8), if establishment requirements for firms in industry (c,s) increase by 10 per cent either because the number of firms increases by 10 per cent or because requirements per firm increase by 10 per cent [$num(c,s) = 10$ or $hf(c,s) = 10$], then the industry requires an increase in the number of input bundles of $10*0.148$ per cent.

Given the determination in (8) of the percentage change in the number of input bundles required in each industry, we can now derive the percentage change in productivity for Melitz industries as $aoMel(c,s)$ in the add-on equation:

$$qbundle(c,s) = qo(c,s) - aoMel(c,s) \quad (9)$$

where: $qo(c,s)$ is the percentage change in the output of industry (c,s) . As defined in GTAP, $qo(c,s)$ is a weighted average over d of the percentage changes in sales, $qs(c,s,d)$, with the weights being shares in factory values.

A question that readers may have is: how does equation (9) work if c is an Armington industry? In this case, $aoMel(c,s)$ is exogenous and equation (9) determines $qbundle(c,s)$. The Melitz determination of $qbundle(c,s)$ in equation (8) is turned off by endogenization of $f_qbundle(c,s)$.

The next group of add-on equations, (10) to (13), are directly from Melitz theory with a Pareto distribution of firm productivities:

$$numl(c,s,d) = num(c,s) - \alpha(c,s) * ptivityMin(c,s,d) \quad (10)$$

$$ptivityMin(c,s,d) = qmin(c,s,d) - ff(c,s,d) \quad (11)$$

$$ptivity(c,s,d) = ptivityMin(c,s,d) \quad (12)$$

$$qtypical(c,s,d) = qmin(c,s,d) \quad (13)$$

Variables newly introduced here are:

$ptivityMin(c,s,d)$ which is the percentage change in the marginal productivity of the c -producing firm operating on the s,d -link with the lowest productivity. This is not the productivity change for a particular firm. It is a comparison between the productivity of the firm that happens to be the lowest productivity c -firm on the s,d -link in one situation (after the shock under investigation) with that of the c -firm that happens to be the one with the lowest productivity on the s,d -link in another situation (before the shock).

$qmin(c,s,d)$ which is the percentage change in the volume of sales to d of the minimum-productivity (c,s) firm operating on the s,d -link. Again, this is not about a particular firm. It is a comparison between the sales on the s,d -link of the low-productivity c -firm on the link in one situation with that of the low-productivity c -firm on the link in another situation.

Each of these four equations can be understood in intuitive terms.¹² Equation (10) says that if there is an increase in the minimum productivity required for profitable operation by a (c,s) firm on the s,d -link, then a smaller proportion of these firms will operate on the link because a smaller proportion will meet the minimum-productivity requirement. Equation (11) can be interpreted as meaning that if, for whatever reason, profitable operation by (c,s) firms on the s,d -link requires higher productivity, then there is a corresponding increase in the required minimum volume of sales relative to the volume of link set-up inputs. Equations (12) and (13) are a convenient implication of the Pareto assumption for firm productivities. These two equations exploit the idea that if x is a random variable

¹² See DJR (2018, chapter 2) for rigorous explanations.

with a Pareto distribution, then the average value of x , conditional on x being greater than any given minimum value, x_{min} , is proportional to x_{min} .

The final pair of add-on equations define the change in the collection of revenue from the Melitz tax on each commodity flow and the change in the total collection of Melitz revenue on all sales of c from region s :

$$d_revtxMel(c,s,d) = 0.01 * MV(c,s,d) * [pmarket(c,s,d) + qs(c,s,d)] - 0.01 * FV(c,s,d) * [pm(c,s) + qs(c,s,d)] \quad (14)$$

$$d_revMeltot(c,s) = \sum_d d_revtxMel(c,s,d) \quad (15)$$

where $MV(c,s,d)$ and $FV(c,s,d)$ are the market and factory values of c sold from s to d (includes s to s).¹³ The first term on the right hand side of (14) is the change in the market value of the c,s,d flow and the second term is the change in the factory value. The gap is the change in Melitz tax revenue.

For Melitz industries, $d_revMeltot(c,s)$ is set exogenously on zero. With total revenue from the Melitz taxes on the sales of industry (c,s) starting on zero and staying on zero, we ensure that the factory value of (c,s) 's production equals the market value of (c,s) sales summed across destinations. In this way we impose zero pure profits, consistent with Melitz theory.

Exogenizing $d_revMeltot(c,s)$ can be thought of as determining the number of firms, $num(c,s)$, that can operate in Melitz industry (c,s) . For Armington industries, $num(c,s)$ is exogenous and $d_revMeltot(c,s)$ is endogenously determined at zero. For Armington industries, the Melitz tax rates start at zero and their movements are set exogenously at zero.

Table 1 shows the closure swaps required in GTAP-A2M to move between Armington and Melitz specifications for industry c .

3. GTAP-A2M: illustrative Armington-Melitz comparison

This section compares two GTAP-A2M simulations of the effects of a 10 per cent increase (from 1.11 to 1.22) in the power of the North American (NAmerica) tariff on imports of Wearing apparel (wap). In the first simulation, all industries are treated as Armington. In the second, the wap industry is Melitz while all others remain Armington¹⁴. The switch from Armington to Melitz for the wap industry

¹³ $MV(c,s,d)$ is the same as the GTAP coefficients $VXMD(c,s,d)$ for $d \neq s$ and $VDM(c,s)$ for $s = d$.

¹⁴ Balistreri *et al.* (2011) and Balistreri and Rutherford (2013) also considered a model with a limited number of Melitz industries and a larger number of Armington industries. As we

as we moved from the first simulation to the second was implemented by the closure swaps in Table 1.

Table 1. Closure swaps for moving between Armington and Melitz specifications for industry c

Exogenous for Armington (endogenous for Melitz)	Exogenous for Melitz (endogenous for Armington)
txMel(c,s,d) for all s,d	f_txMel(c,s,d) for all s,d
aoMel(c,s) for all s	fqbundle(c,s) for all s
as(c,s,d) for all s,d	f_as(c,s,d) for all s,d
num(c,s) for all s	d_revMeltot(c,s) for all s

Source: Author construction's.

We conducted the simulations in a 10-region, 57-commodity version of GTAP-A2M using GEMPACK software¹⁵. In both simulations, we assumed that an increase in the NAmerica wap tariff has no effect on the balance of trade, investment and public consumption in any region. These are convenient simplifying assumptions. They allow us to measure welfare effects by the percentage changes in private consumption. For the wap industry under Melitz, we set the Pareto shape parameter (α) at 2 for all regions and the substitution elasticity between varieties (σ) at 2.5. For the wap industry under Armington, the shape parameter plays no role. For the wap inter-regional substitution elasticity in Armington we used the value 7.0 for both domestic/import and import/import substitution. Why 7.0? As we have argued elsewhere (Dixon *et al.*, 2016), Armington-Melitz comparisons are facilitated by parameter choices that imply similar levels under the two specifications for the sensitivity of trade flows to tariff changes. We chose the Melitz and Armington σ values so that the effect on NAmerican wap imports of the 10 per cent tariff was the same in both simulations. With the values at 2.5 and 7.0, both simulations gave a 32 per cent reduction in the quantity of NAmerica's wap imports.¹⁶

Figure 1 shows regional percentage welfare effects of the NAmerican wap tariff under Armington and Melitz specifications. These results are subdued (smaller welfare gains and losses) relative to those in DJR (2018, ch. 7). Unlike the DJR wap

will see, this structure generates important welfare effects associated with movements of resources between constant-returns-to-scale and increasing-returns-to-scale industries.

¹⁵ See Harrison *et al.* (2014) and Horridge *et al.* (2013).

¹⁶ As explained later, in measuring changes in quantities we use quality-adjusted units.

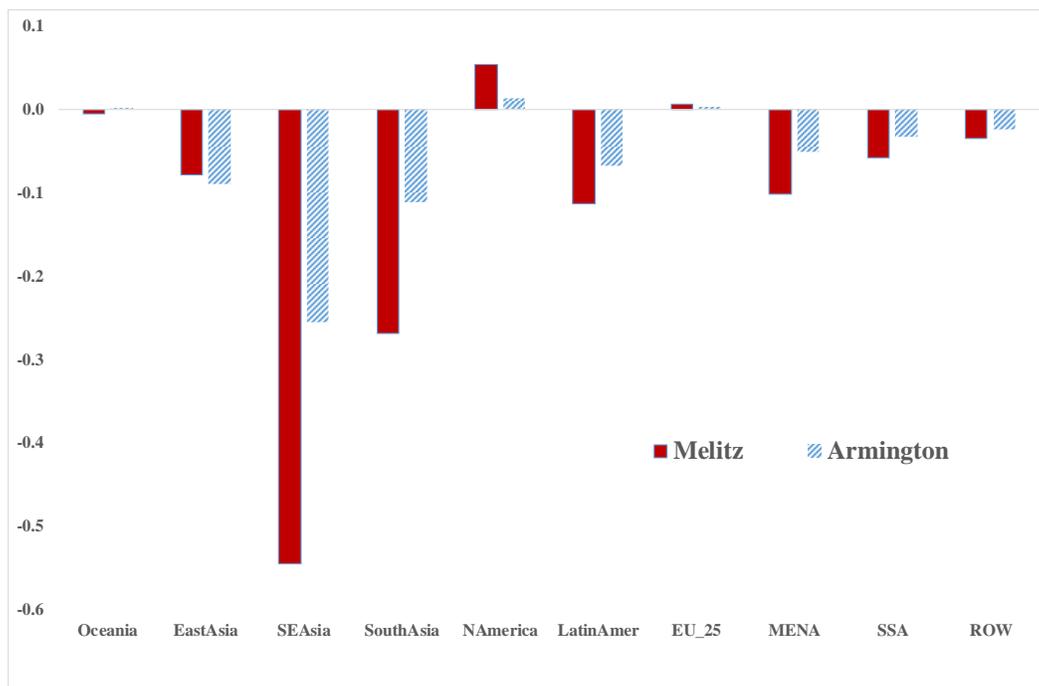


Figure 1. A 10 per cent increase in the power of the Wap tariff on North American imports: percentage welfare effects in Armington and Melitz

Source: Author calculations.

simulations, the current simulations were run with the GTAP database modified by the elimination of domestic wap inputs to the wap industry in each region. While this modification was not essential for obtaining stable solutions for simulations in which wap is treated as Melitz, stability was dramatically improved when the modification was adopted in simulations in which another industry, Electronic equipment, was Melitz. We found that large flows of domestically produced own product to a Melitz industry (as is the case for Electronic equipment) can cause stability problems by accentuating scale effects leading to a strongly downward sloping supply curve for quality adjusted output, with the possibility that the supply curve crosses the demand curve from above.

As can be seen from Figure 1, under either Melitz or Armington, the main losers are SE Asia and South Asia and the main winner is NAmerica. These qualitative results are unsurprising. SE Asia and South Asia have the biggest dependence of all the regions on exports of wap to NAmerica. For SE Asia the value of these exports in the GTAP database is 2.31 per cent of the total value of SE Asia's private consumption (welfare). The corresponding percentage for South Asia is 1.20. In a simulation in which NAmerica makes a modest tariff increase from a low rate (averaging 11 per cent across all sources) without retaliation, we would expect it to derive a welfare gain via the optimal-tariff argument.

While Figure 1 shows the same winners and losers for Armington and Melitz, the Melitz results are generally more extreme: under Melitz the main losers lose more and the main winner wins more. In the remainder of this section we focus on the differences between the Armington and Melitz results for SE Asia, South Asia and NAmerica.

3.1. Decomposing the Armington and Melitz welfare effects

The GEMPACK code for GTAP-A2M includes an exact decomposition of the percentage changes in real private consumption attributable to policy changes and other shocks under investigation. Given the assumptions in our Melitz and Armington wap-tariff experiments, decomposition of private consumption becomes a decomposition of welfare. For this paper we use a stripped-down version of the decomposition equation which includes only terms relevant for our wap experiments:¹⁷

$$\begin{aligned}
 \text{welfare}(s) = & \frac{1}{\text{AGGCON}(s)} * \left\{ \right. \\
 & \text{DUM_NA}(s) * \sum_{s \neq d} \text{MTAX}(\text{"wap"}, d, s) * [\text{qs}(\text{"wap"}, d, s) + \text{as}(\text{"wap"}, d, s)] \\
 & \qquad \qquad \qquad \text{[Tariff efficiency contribution for NAmerican wap tariff]} \\
 & + \sum_c \sum_{d \neq s} \left\{ \text{VXWD}(c, s, d) * [\text{pfob}(c, s, d) - \text{as}(c, s, d)] \right. \\
 & \qquad \qquad \qquad \left. - \text{VXWD}(c, d, s) * [\text{pfob}(c, d, s) - \text{as}(c, d, s)] \right\} \\
 & \qquad \qquad \qquad \text{[Terms of trade contribution]} \\
 & + \text{DUM_M} * \frac{1}{\sigma(\text{"wap"})} * \sum_d \text{MV}(\text{"wap"}, s, d) \\
 & \qquad \qquad \qquad * [\text{qs}(\text{"wap"}, s, d) + \text{as}(\text{"wap"}, s, d)] \qquad \text{[Melitz scale contribution]} \\
 & + \sum_{c, d, a, s} \text{SALESTAX}(c, d, a, s) * \text{qf}(c, d, a, s) \left. \right\} \\
 & \qquad \qquad \qquad \text{[Other tax efficiency contributions]}
 \end{aligned} \tag{16}$$

¹⁷ In the stripped-down version and in the decomposition calculations reported in Table 2, but not in the GEMPACK code and the welfare results in Table 2, we leave out the influence on welfare of transport margins. This presentational simplification introduces a small discrepancy between the “true” welfare results in Table 2 and the sum of the contributions. The stripped-down version also leaves out changes in region d’s factor endowments, investment, public expenditure, the trade balance and numerous technology terms that are held constant in the present simulations. The full decomposition equation is set out in DJR (2018, section 7.4 and Appendix 7.2).

where

$welfare(s)$ is the percentage change in welfare in region s (percentage change in real private consumption);

$DUM_NA(s)$ is a dummy coefficient with value one for NAmerica and zero otherwise;

$AGGCON(s)$ is the value of private consumption;

$MTAX("wap",d,s)$ is tariff collection by region s on imports of wap from region d ;

$qs("wap",d,s)$ is the percentage change in the quantity (count) of wap sent from d to s (includes d to d);

$as(c,d,s)$ is the percentage change in the (c,d) -augmenting technology or preference variable applying to flows of c from d to all agents in s [discussed earlier in connection with equation (7)];

$VXWD(c,s,d)$ is the fob value of c sent from s to d (i.e. market value plus genuine export taxes);

$pfob(c,s,d)$ is the percentage change in the fob price (per count) of c sent from s to d ;

DUM_M is a dummy coefficient with value one when wap is treated as Melitz and zero when it is Armington;

$\sigma("wap")$ is the wap inter-variety substitution elasticity (2.5 in the Melitz simulation);

$MV("wap",s,d)$ is the market value of wap produced in s and sent to d ($VXMD$ and VDM in GTAP notation);

$SALESTAX(c,d,a,s)$ is sales tax collection by region s on flows of commodity c from d to agent a (household, firms and government) in s ; and

$qf(c,d,a,s)$ is the percentage change in the quantity of flow (c,d,a,s) .

A key feature of equation (16) is the use of *quality-adjusted* units rather than *count* units. An example is the easiest way to explain what we mean by quality-adjusted units. If there is a 10 per cent increase in the number (count) of cars sent from d to s [$qs(cars,d,s) = 10$] and the suitability of these cars for satisfying the needs of users in s increases by 5 per cent [$as(cars,d,s) = 5$], then the quality-adjusted flow from d to s increases by 15.5 per cent [$= 100*((1 + 10/100)*(1 + 5/100) - 1)$]. If at the same time the price per car sent from d to s increases by 8 per cent, then the price per quality-adjusted car increases by 2.86 per cent [$= 100*((1 + 8/100)/(1 + 5/100) - 1)$].

The first contribution identified on the right hand side of (16) is the percentage change in NAmerican welfare generated by the tariff-induced percentage change in this region's imports of quality-adjusted units of wap. This is the familiar triangle/rectangle effect. Using GEMPACK, we impose the 10 per cent increase in the power of NAmerica's wap tariff (from 1.11 to 1.22) in a series of small steps. The effect on welfare is obtained by adding up the effects from each step. This

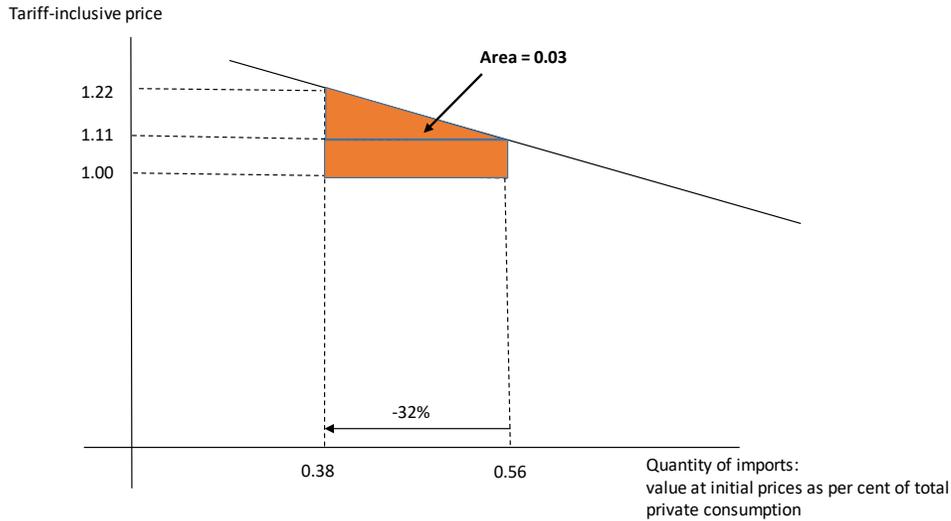


Figure 2. North American imports of Wap
 Source: Author calculations.

multi-step approach enables GEMPACK to track the change in $MTAX("wap", d, "NAmerica")$ as the power of the NAmerican tariff moves from its initial value to its final value. In this way, GEMPACK captures the welfare effect from the tariff increase as the negative of the shaded triangle/rectangle area in Figure 2.

The second contribution identified in (16) is the percentage change in the welfare of region s generated by the change in the region's terms of trade. The terms of trade for region s is the difference between weighted averages of the percentage changes in the quality-adjusted prices of its exports and imports. In our stripped-down version, equation (16), of the decomposition equation we use fob prices and weights for imports as well as exports. As mentioned in footnote 14, for the stripped-down version we simplify by omitting international transport costs.

The third contribution identified in (16) is the percentage change in the welfare of region s generated by changes in the scale of quality-adjusted output in the region's Melitz industries (only wap in the illustrative Melitz simulation presented here). In explaining this contribution, we start by defining the percentage change in quality-adjusted output in industry c in region s [$q_{quality}(c,s)$] by

$$q_{quality}(c,s) = \sum_d SMV(c,s,d) * (qs(c,s,d) + as(c,s,d)), \quad (17)$$

where the symbols on the right hand side have been defined earlier, see equations (4), (7) and (8). We show that Melitz industries exhibit increasing returns to scale, that is a Melitz industry c in region s can generate a 1 per cent increase in its quality-adjusted output with less than a 1 per cent increase in input bundles [$q_{quality}(c,s) - q_{bundle}(c,s) > 0$]. It can do this by exploiting economies of scale at either the firm level or industry level.

Consider firm-level economies of scale. If each existing c -firm in region s increases its sales on each link by 1 per cent and there is no change in the number of firms, the numbers on links, the marginal productivity of any firm, and firm and link setup requirements [$num(c,s) = numl(c,s,d) = ptivity(c,s,d) = hf(c,s) = ff(c,s,d) = 0$ for all d], then $qs(c,s,d) = 1$ for all d while $as(c,s,d) = 0$ for all d [see (7)¹⁸]. Thus, quality-adjusted output at the industry level increases by 1 per cent [see (17)]. But the industry increases its input bundles by only $(\sigma(c)-1)/\sigma(c)$ per cent, see (8). This is because existing firms do not need to incur additional costs either to start business or to set up on links. With quality-adjusted output increasing by 1 per cent and inputs increasing by $(\sigma(c)-1)/\sigma(c)$ per cent, total-factor productivity in the industry increases by $1/\sigma(c)$ per cent.

Alternatively, economies of scale could be achieved at the industry level, but not at the firm level. Assume that each existing firm stays unaltered, producing an unchanged amount for each link while at the same time, the number of firms and the number on each link increases by $(\sigma(c) - 1)/\sigma(c)$ per cent [$num(c,s) = numl(c,s,d) = (\sigma(c) - 1)/\sigma(c)$]. Assume no change in sales per typical firm on each link [$qtypical(c,s,d) = 0$] and continue to assume that there are no changes in the marginal productivity of typical firms on links or in firm and link setup requirements [$ptivity(c,s,d) = hf(c,s) = ff(c,s,d) = 0$ for all d]. Then, there is a 1 per cent increase in quality-adjusted output for the industry achieved by a $(\sigma(c) - 1)/\sigma(c)$ per cent increase in count sales on each link [$qs(c,s,d) = (\sigma(c) - 1)/\sigma(c)$, see (4)] and a $1/\sigma(c)$ per cent increase in quality, generated by increases in $as(c,s,d)$ of $1/\sigma(c)$ per cent for all d [see (7)]. From (8), the percentage increase in input bundles is $(\sigma(c) - 1)/\sigma(c)$. Again, with quality-adjusted output increasing by 1 per cent and inputs increasing by $(\sigma(c)-1)/\sigma(c)$ per cent, total-factor productivity in the industry increases by $1/\sigma(c)$ per cent.

It is the contribution to welfare in region s of the total factor productivity changes in Melitz industries that is reflected in the third term on the right hand side of (16). These total factor productivity changes are generated at the rate of $1/\sigma(c)$ per cent for each 1 per cent increase in quality-adjusted output for a Melitz industry.

In our experience with Melitz-based models, economies of scale at the industry level are generally more important than those at the firm level. As demonstrated in DJR (2018, Appendix 7.2), if input requirements for establishing a (c,s) firm and setting up on all trading links are constant [$hf(c,s) = ff(c,s,d) = 0$ for all d] then, at least in a simplified model, the average size of (c,s) firms is constant implying that on average there are no realized firm-level economies of scale. However, the scale contribution in (16) remains legitimate regardless of whether the change in

¹⁸ For a Melitz industry, $f_{as}(c,s,d) = 0$.

quality-adjusted industry output takes place via more firms or more output by existing firms.

The scale effect does not apply to Armington industries. For these industries, GTAP-A2M specifies constant returns to scale (no fixed costs) and no endogenous changes in the quality of output [$as(c,s,d) = 0$ for all d]. Consequently, in our first simulation in which all industries are Armington, inputs move in line with output, generating zero scale contributions to welfare. This is recognized in equation (16) through DUM_M which is zero when the wap industry is specified as Armington.

The fourth contribution identified in (16) is the percentage change in the welfare of region s generated by changes in other tax-carrying flows (other than the wap import flows to NAmerica). This contribution is fully specified in the GEMPACK code for the GTAP-A2M system and in DJR (2018, chapter 7). Since it is not a focus of this paper, we represent it in (16) only in stylized form. What this term recognizes is that increases in taxed flows boost welfare because the purchasing agents put a higher value on these flows (the tax-inclusive price) than the opportunity cost of supplying the flows (the tax-exclusive price).

Results from the welfare decomposition

The first column of results in Table 2 shows the welfare effects on SE Asia, South Asia and NAmerica of NAmerica's 10 per cent increase in tariffs on imports of wap under Armington and Melitz assumptions for the wap industry. The other four columns show the contributions from the factors identified in (16). As explained in footnote 14, the stripped-down decomposition in (16) is not comprehensive. However, the discrepancies in Table 2 between the results in column (1) and the decomposition contributions added over columns (2) to (5) are small.

As mentioned earlier, NAmerica's wap imports in both simulations contract by 32 per cent in quality-adjusted units. Thus, on this account, column (2) of Table 2 shows the same tariff efficiency effect on NAmerican welfare in the Armington and Melitz simulations (a reduction of 0.03 per cent, explained in Figure 2).

With wap specified as an Armington industry, Table 2 shows negative terms-of-trade contributions for SE Asia and South Asia and a positive contribution for NAmerica. The NAmerica tariff reduces SE Asia's and South Asia's wap exports to NAmerica. Under our assumption of no change in trade balances, SE Asia and South Asia experience rebalancing real devaluations. These real-exchange-rate movements reduce SE Asia's and South Asia's imports and stimulate their exports. The reduction in SE Asia's and South Asia's imports barely affects foreign-currency supply prices to these regions: SE Asia's and South Asia's imports are generally only moderate fractions of other regions' exports. On the other hand, under the Armington assumption that each region produces a distinctive version of each product, stimulation of export demand for SE Asia's and South Asia's products requires reduction in their foreign-currency prices. This generates reductions in the terms of trade for the two regions. For NAmerica, the opposite argument applies. NAmerica's wap imports fall. Rebalancing takes place via real

appreciation which reduces in NAmerica's exports with corresponding increases in their foreign-currency prices.

With wap specified as a Melitz industry, Table 2 gives positive rather than negative terms-of-trade contributions for SE Asia and South Asia and a negative rather than a positive contribution for NAmerica. The rebalancing argument from the last paragraph continues to apply. However, it is outweighed by Melitz effects on quality-adjusted wap prices. Contraction in wap output in SE Asia and South Asia causes loss of scale economies. This results in an increase in wap supply prices of quality-adjusted units from these regions which is sufficient to generate terms-of-trade improvements despite trade rebalancing via real devaluation. For NAmerica, the negative terms-of-trade contribution under Melitz arises partly from scale-related reductions in the quality-adjusted prices of its wap exports. But this is a minor effect: NAmerica's wap exports are small. The main reason that the terms-of-trade contribution to NAmerican welfare switches from positive to negative as we go from Armington to Melitz is that NAmerica pays higher prices for quality-adjusted units of wap imports. Under Melitz (but not Armington), NAmerica's tariffs reduce the ability of SE Asia and South Asia to supply cheap units of quality-adjusted wap.

The welfare contributions of the total factor productivity changes associated with scale economies in the wap industry are shown in column (4) of Table 2. These are zero under Armington assumptions: there are no scale effects. Under Melitz, they are positive for NAmerica, reflecting expansion of the wap industry, and negative for SE Asia and South Asia reflecting contraction of the wap industry.

Column (5) of Table 2 shows welfare contributions from changes in other tax-carrying flows. For SE Asia and South Asia, the negative contributions (-0.04 under Armington and -0.06 under Melitz for both regions), reflect reductions in consumption: in both regions sales to households carry substantial indirect taxes. Thus, the contribution in column (5) is not an explaining factor for the negative outcomes for SE Asian and South Asian welfare. Rather, it is a reinforcing factor, exacerbating negative outcomes caused by other factors.

3.2. Towards a fundamental explanation of the difference between Armington and Melitz welfare effects

Moving up and down the columns of Table 2 tells us a lot about how Armington and Melitz models operate. It tells us: that Melitz produces scale effects but Armington doesn't; that Melitz terms-of-trade effects can be in the opposite direction to those in Armington; and that contributions from tax-carrying flows are broadly similar between the two models. This last finding depends on our decision, explained earlier, to calibrate the Melitz and Armington substitution elasticities so that the two models imply similar sensitivity of NAmerica's wap imports to changes in NAmerica's wap tariffs.

Table 2. Decomposition of the welfare effects of a 10 per cent tariff imposed by NAmerica on imports of Wearing Apparel (wap)

	Welfare (%change) (1)	Contributions from:			
		Tariff efficiency (2)	Terms of trade (3)	Scale effect (4)	Other tax-carrying flows (5)
<i>Armington assumptions for all industries</i>					
SEAsia	-0.26	0	-0.21	0	-0.04
SouthAsia	-0.11	0	-0.07	0	-0.04
NAmerica	0.01	-0.03	0.04	0	0.00
<i>Melitz assumptions for wap, Armington for all other industries</i>					
SEAsia	-0.55	0	0.13	-0.61	-0.06
SouthAsia	-0.27	0	0.18	-0.38	-0.06
NAmerica	0.05	-0.03	-0.01	0.09	0.00

Source: Author calculations.

Table 3. Percentage welfare effects of a 10% tariff imposed by NAmerica on imports of Wearing Apparel (wap): decomposition of welfare differences between Melitz and Armington

	Armington	Melitz	Difference	Explained difference	Domestic supply cost	Rebalancing terms of trade	Other tax-carrying flows
	(1)	(2)	(3) = (2) - (1)	(4) = (5) + (6) + (7)	(5)	(6)	(7)
SEAsia	-0.26	-0.55	-0.29	-0.30	-0.19	-0.09	-0.02
SouthAsia	-0.11	-0.27	-0.16	-0.14	-0.06	-0.06	-0.02
NAmerica	0.01	0.05	0.04	0.06	0.07	-0.005	-0.01

Source: Author calculations.

By comparing columns or moving across rows in Table 2, we hoped to obtain a satisfying explanation of the net differences between Armington and Melitz welfare effects. What this comparison shows is that the welfare results for SE Asia and South Asia are more negative under Melitz than Armington because the negative Melitz scale contributions (-0.61 and -0.38) for these regions outweigh the positive differences between the Melitz and Armington terms-of-trade contributions (0.34 and 0.25). For NAmerica, Table 2 tells us that the welfare effect is more favorable with Melitz than Armington because the positive Melitz scale contribution (0.09) outweighs the negative difference between the Melitz and Armington terms-of-trade contributions (-0.05). However, Table 2 doesn't tell us why, as we go from Armington to Melitz, scale losses for SE Asia and South Asia outweigh these regions' terms-of-trade gains and why the scale gain for NAmerica outweighs its terms-of-trade loss.

The problem with Table 2 is that the reasons for the differences between the Armington and Melitz welfare results are obscured by offsetting overlap between the scale and terms-of-trade contributions in Melitz. Negative scale contributions in Melitz lead to positive terms-of-trade effects and vice versa. Eliminating the overlap helps us to find the fundamental factors underlying the net welfare differences.

In Table 2, the Melitz scale contributions for SE Asia and South Asia embrace the whole of the losses of wap total factor productivity while the Melitz terms-of-trade contributions include the offsetting higher wap export prices. In thinking through how to remove this overlap, we recognized that scale-related losses for SE Asia and South Asia in the Melitz simulation reduce the welfare of these regions only through increasing the cost of supplying domestically produced quality-adjusted wap to domestic households. From a welfare point of view, the increase in the cost of supplying exports [part of the scale effect in column (4) of Table 2] is offset by higher prices paid by foreigners [part of the terms-of-trade effect in column (3)]. The story for NAmerica is similar but in reverse: scale-related gains for NAmerica increase its welfare only through reducing the cost of supplying domestically produced quality-adjusted wap in NAmerica.

On the basis of these ideas, we constructed Table 3. We start the explanation with column (5) which is concerned with scale-related changes in the cost to domestic consumers of domestically produced quality-adjusted units of wap. Referring to detailed results not shown in this paper, we find under Melitz assumptions that the NAmerican tariff causes increases of 8.32 per cent and 9.27 per cent in the cost to SE Asian and South Asian households of quality-adjusted units of domestic wap. These cost increases reflect output contraction with associated loss of scale economies. In the two regions, domestic wap accounts for 2.25 and 0.60 per cent of aggregate consumption. Thus, the cost increases in supplying domestic wap generate welfare losses for the two regions of 0.19 per cent ($= 0.0225 \times 8.32$) and 0.06 per cent ($= 0.0060 \times 9.27$). These are welfare losses

under Melitz that do not occur under Armington. They are entered in column (5) of Table 3 as explanators of the difference between Melitz and Armington welfare results.

For NAmerica, scale gains under Melitz reduce the cost to households of quality-adjusted units of domestic wap by 8.11 per cent. For NAmerican households, domestic wap accounts for 0.92 per cent of aggregate consumption. Thus, as shown in column (5) of Table 3, the cost reduction under Melitz in supplying domestic wap generates a 0.07 per cent increase in NAmerican welfare ($= 0.0092 \times 8.11$).

What about terms-of-trade effects? In generating the entries for SE Asia and South Asia in column (6) of Table 3 we considered the terms-of-trade effects for these regions of rebalancing their trade to compensate for their losses of wap export revenue. As indicated by the extent of real devaluation, the rebalancing task is considerably more onerous under Melitz than under Armington. Real devaluation for SE Asia, measured by the reduction in its factor price index relative to the world price level, is 43 per cent greater under Melitz than under Armington. For South Asia, real devaluation is 87 per cent greater under Melitz than under Armington. Melitz requires extra real devaluation because NAmerica's tariffs are considerably more damaging to SE Asia's and South Asia's wap exports under Melitz than under Armington. Under Melitz, increased costs reduce the ability of SE Asia and South Asia to export wap to third markets. By contrast, under Armington real devaluation allows SE Asia and South Asia to increase their wap exports to third markets. As can be seen in Table 2, under Armington assumptions the terms-of-trade contributions for SE Asia and South Asia are -0.21 and -0.07 per cent. On the basis of the extra real devaluations (and extra stimulation of non-wap exports) required under Melitz relative to Armington, we estimate that Melitz imposes extra rebalancing terms-of-trade losses for the two regions of 0.09 per cent ($= 0.43 \times 0.21$) and 0.06 per cent ($= 0.87 \times 0.07$). These are the numbers shown for SE Asia and South Asia in column (6) of Table 3.

For NAmerica, the Melitz simulation gives a slightly smaller real appreciation than the Armington simulation (increases in factor prices of 0.246 and 0.278 per cent). On the one-hand, the NAmerican rebalancing adjustment under Melitz requires less stimulation of non-wap exports than under Armington because Melitz scale economies stimulate NAmerica's wap exports. But this is a negligible effect: as mentioned earlier, NAmerica has very little wap exports. On the other hand, NAmerica's rebalancing requires more stimulation of non-wap exports under Melitz because scale diseconomies in wap-supplying regions increase the cost of NAmerica's wap imports. Applying the ratio of the Melitz to Armington real exchange rate results to the Armington terms-of-trade result suggests extra terms-of-trade decline for NAmerica in the Melitz simulation of 0.005 per cent [$= 0.04 \times (0.246 / 0.278 - 1)$].

Column (7) of Table 3 simply records the differences between the Melitz and Armington other-tax-carrying-flow contributions in column (5) of Table 2. It follows from the earlier discussion that the differences for SE Asia and South Asia are mainly a reflection of differences in total consumption effects caused by other factors. For NAmerica, tax-carrying flows make a slightly more negative contribution under Melitz than under Armington, but the difference is very small.

The tariff-efficiency contributions for the three regions are the same in both simulations (-0.03 for NAmerica and 0 for the other two regions). Consequently there is no tariff contribution column in Table 3 which is concerned with differences between the two simulations.

Summing up

Comparison of columns (3) and (4) in Table 3 indicates that the back-of-the-envelope calculations for the three effects in columns (5), (6) and (7) do not provide a completely accurate decomposition of the difference between the Melitz and Armington welfare effects in our two simulations. Nevertheless they tell most of the story.

They tell us that the principal losers, SE Asia and South Asia, lose more welfare under Melitz than under Armington for two reasons. First, Melitz recognizes that the contraction of SE Asia's and South Asia's wap industries generates an increase in the cost to households in these two regions of domestically-supplied wap products [column (5)]. This welfare-reducing effect is not present in Armington. Second, the NAmerican tariff does more damage to SE Asia's and South Asia's wap industries under Melitz than under Armington, even when we normalize so that the contraction in NAmerica's wap imports is the same under both sets of assumptions. This is because in Melitz (but not in Armington) cost increases in SE Asia's and South Asia's wap industries caused by loss-of-scale economies reduce the two regions' wap export revenue from third markets. Consequently, greater stimulation of SE Asia's and South Asia's non-wap exports is required under Melitz than under Armington. Greater stimulation of these non-wap exports causes a larger welfare-reducing terms-of-trade reduction under Melitz than under Armington [column (6)].

For NAmerica, the story from column (5) of Table 3 is that welfare under Melitz (but not Armington) is increased by the expansion of the domestic wap industry which reduces the cost to NAmerican households of domestically-produced wap. On the other hand, welfare under Melitz (but not Armington) is reduced by increases in the cost to NAmerica of imported wap caused by loss of scale economies in supplying regions. This effect is reflected in the negative but small entry in column (6) of Table 3. *A priori* it was not clear that the positive entry for NAmerica in column (5) would outweigh the negative entry in column (6). The balance depends on the data. For example, if we considered a situation in which the domestic/import wap ratio in the NAmerican market is much smaller than

that in the GTAP database, then we might generate a negative entry in column (6) that outweighs the positive entry in column (5).

4. Other Armington to Melitz conversion approaches

In creating the GTAP-A2M system our main objective was to give users of the GTAP model the option of switching from Armington to Melitz for industries of their choice. This was also the objective of Akgul *et al.* (2016) in creating GTAP-HET.

What are the advantages and disadvantages of GTAP-HET versus GTAP-A2M? The advantage of GTAP-HET is that it allows for the composition of inputs to current production in an industry to be different from that used in establishing firms and setting up on trade links. In GTAP-A2M we assume no difference between the composition of input bundles used by industry (c,s) in current production from that of bundles used in meeting fixed-cost requirements. In the 2016 version of GTAP-HET this advantage hadn't been pushed very far. It was assumed that all fixed cost requirements in industry (c,s) are met with input-bundles consisting entirely of primary factors with the same mix and substitution possibilities as the primary factors used by (c,s) in current production. In Akgul *et al.* (2017) a serious effort was made to identify the occupational composition of the labor input to fixed costs. Work along these lines has the potential to make GTAP-HET's distinction between current and fixed input composition worthwhile. However, a disadvantage of the distinction is that it requires extensive changes in GTAP's input demand equations to separate demands for inputs to current production from demands for inputs to fixed costs.¹⁹ By contrast, GTAP-A2M leaves the original GTAP model almost untouched. We simply add some rather trivial equations on the end and implement them with closure swaps.

As demonstrated in DJR (2018, Appendix 7.3), a problem with GTAP-HET is that it fails to incorporate Melitz pricing to market. Instead of equation (5), in the notation of this paper GTAP-HET specifies that Melitz industries price according to

$$p_{market}(c,s,d) = p_{bundle}(c,s) - ptivity_{ave}(c,s) \quad (18)$$

Notice that there is no d argument on the right hand side of (18). The same market price applies to all destinations. In (18), GTAP-HET alters (5) by replacing the percentage change in the marginal productivity of the typical c-firm on the s,d-link with the percentage change in marginal productivity averaged over all (c,s) firms.

¹⁹ Akgul *et al.* (2016) also introduce explicit modifications to standard GTAP to eliminate the two-tier import/domestic substitution specification. We don't think this is necessary for introducing Melitz to GTAP. As mentioned in Section 2, the two-tier structure implicitly disappears if we adopt the same value for the substitution elasticity at each level.

Equation (18) is inconsistent with Melitz theory. Akgul *et al.* do not produce any other theory to justify it. In our view, (18) is a mistake. But does it matter? Simulations with a stylized model reported in DJR (2018) suggest that the answer is yes. In any case, inclusion of (18) along with a tortuous but *ad hoc* definition of $ptivity_{ave}(c,s)$ certainly impedes understanding of GTAP-HET and interpretation of its results.²⁰

Working independently of us, Bekkers and Francois (BF, 2018) have produced what they call the parsimonious approach for converting an Armington model into a Melitz model.²¹ In spirit, their conversion method is similar to ours. Like us, they implement their method in GEMPACK and test it by converting standard GTAP. To make the conversion from Armington to Melitz, BF introduce three shift variables that they refer to as demand, supply and generalized iceberg trade cost shifters. The details of the BF specifications of these shifters are complex, but in broad terms they correspond to our use of $as(c,s,d)$, $aoMel(c,s)$ and $txMel(c,s,d)$. What is necessary in going from Armington to Melitz is the introduction of three effects: demand changes in destination markets to reflect love of variety; total factor productivity changes in industries to reflect economies of scale; and pricing to market to reflect variations in the marginal productivities of the groups of firms that service different links. Comparison of our method with the BF method demonstrates that there are different ways of splitting the required effects between three variables. Not surprisingly, we prefer our split. It has helped us to interpret Melitz results, to explain them in terms of familiar mechanisms, and to understand why they are different from Armington results. But in the end, we must leave it to readers to decide what they prefer.

5. Concluding remarks

The GTAP-A2M system makes it possible for users of the GTAP model to explore the implications of Melitz theory in a computing system with which they are familiar. There are no new data requirements, beyond assigning values for the Pareto shape parameter (α) and revising the Armington substitution elasticities (σ) so that they take on values suitable for substitution elasticities between varieties rather than regions. There are no new dimensional limitations. In the GTAP-A2M system computed with GEMPACK, an r-region, n-industry model incorporating Melitz industries takes barely any longer to solve than if all the industries were Armington.

²⁰ Bekkers and Francois (2018, page 7) imply that we did not allow for free entry and exit in our test of the GTAP-HET model. This is incorrect. As can be seen from DJR (2018, p. 183), our stylized test of GTAP-HET does allow for entry and exit.

²¹ Like DJR (2018), Bekkers and Francois also explain how to convert from Armington to models with the theoretical structure developed by Krugman (1980). The Krugman model can be considered a special case of Melitz.

In our earlier paper in this journal, Dixon *et al.* (2016), we sided cautiously with Arkolakis *et al.* (2012)'s "same-old-gains" paper which argues on the basis of theory that welfare results from Melitz models are unlikely to be much different from those of Armington models. Our opinion was coloured by results from a stylized numerical model in which we inadvertently adopted simplifying assumptions similar to those in Arkolakis *et al.*'s basic theoretical model. Our assumptions included: (a) one factor of production in each region; (b) no intermediate inputs; (c) all industries being Melitz or all being Armington; and (d) the same parameter values in each industry. The basic theoretical model of Arkolakis *et al.* explicitly adopts (a) and (b). Because Arkolakis *et al.* allowed for only one industry in each region, they implicitly adopted (c) and (d). Not surprisingly, the numerical results from Dixon *et al.* (2016) supported the theoretical results from Arkolakis *et al.* (2012).

What we found from the GTAP-A2M system is that conclusions derived in a model embracing (a) to (d) are not readily generalizable.²² If some industries are modelled as Melitz and others as Armington, then the simulated welfare effects of trade policies can be quite different from those obtained when all industries are Armington. In our simulations of a tariff increase by NAmerica on wap imports we compared GTAP-A2M results in which wap was modelled as Melitz and all other industries were Armington with results in which all industries were Armington. In these simulations, resources in SE Asia and South Asia are transferred out of wap and into other industries, while the opposite happens for NAmerica. We found that when wap was treated as a Melitz industry (with increasing-returns-to-scale) the transfer of resources in SE Asia and South Asia into Armington industries (with constant returns to scale) generated significant additional welfare losses for these regions compared with the situation in which all industries were Armington. Using a decomposition method, we found that these extra losses in Melitz for SE Asia and South Asia could be explained by extra costs in these regions of supplying their own households with domestically produced wap and by extra terms-of-trade losses associated with extra stimulation of non-wap exports required to restore trade balance. Expansion of the wap industry in NAmerica generated an additional welfare gain for NAmerica when wap was Melitz that was not present when wap was Armington through reduction in the cost of supplying domestically produced wap to domestic households.

Armington has been the workhorse trade specification in CGE modelling since the 1970s. Should it be replaced by mixed-Armington-Melitz specifications? We hope that the GTAP-A2M system facilitates the research necessary to continue

²² This point has been made by several authors including Balistreri *et al.* (2011), Melitz and Redding (2015) and Akgul (2017). Perhaps our contribution is to sharply identify and quantify a particular mechanism that causes Melitz results to diverge from Armington results in models in which (a) to (d) do not hold.

teasing out the implications of Melitz assumptions and the realism of the results. For example, using GTAP-A2M we analysed the stability problem mentioned in section 3 associated with inputs of domestically produced own product to Melitz industries.

One final hint. In understanding Melitz results via explanatory back-of-the-envelope calculations, it is helpful to work with quality-adjusted prices and quantities. Working with these concepts allows us to think in terms of demand and supply curves which are not subject to shifts caused by changes in variety variables.

Acknowledgements

We thank Tom Hertel, Zeynep Akgul and two anonymous referees for comments on an earlier draft. We also thank the Australian Research Council for financial support under DP140100476.

References

- Akgul, Z., N.B. Villoria and T.W. Hertel (2016), "GTAP-HET: Introducing Firm Heterogeneity into the GTAP Model", *Journal of Global Economic Analysis*, vol. 1(1), pp. 111-180.
- Akgul, Z., C. Carrico and M. Tsigas (2017), "Does the labor composition of fixed business costs matter?", paper presented at the 20th Annual GTAP Conference, held at Purdue University, West Lafayette, Indiana, June, pp. 19.
- Akgul, Z., (2017), "One Model to Rule Them All: The Importance of Firm Heterogeneity in CGE Modeling of the Gains from Trade," *USITC Working Paper*, 2017-3-B.
- Arkolakis, C., A. Costinot, A. Rodriguez-Clare (2012), "New trade models, same old gains?", *American Economic Review*, 102(1), pp. 94-130.
- Armington, P. (1969), "A theory of demand for products distinguished by place of production", *Staff Papers-International Monetary Fund*, 16(1), pp. 159-178.
- Armington, P. (1970), "Adjustment of trade balances: some experiments with a model of trade among many countries", *Staff Papers-International Monetary Fund*, 17(3), pp. 488-523.
- Balistreri, E., R. Hillberry, T. Rutherford (2010), "Trade and welfare: does industrial organization matter?", *Economics Letters*, 109(2), pp. 85-87.
- Balistreri, E., R. Hillberry and T. Rutherford (2011). "Structural estimation and solution of international trade models with heterogeneous firms", *Journal of International Economics* 83(1), pp. 95-108.
- Balistreri, E. and T. Rutherford (2013), "Computing general equilibrium theories of monopolistic competition and heterogeneous firms", chapter 23 in P.B. Dixon and D.W. Jorgenson (editors), *Handbook of Computable General Equilibrium Modeling*, Elsevier, pp. 1513-1570.

- Bekkers, E. and J. Francois (BF, 2018), "A parsimonious approach to incorporate firm heterogeneity in CGE models", *Journal of Global Economic Analysis*, vol 3(2), pp. 1-68.
- Corong, E., T. Hertel, R. McDougall, M. Tsigas and D. van der Mensbrugghe (2017), "The standard GTAP model, Version 7", *Journal of Global Economic Analysis*, vol. 2(1), pp. 1-119.
- Deardorff, A.V., R.M. Stern and C.F. Baum (1977), "A multi-country simulation of the employment and exchange-rate effects of post-Kennedy round tariff reductions", Chapter 3, pp. 36-72 in N. Akrasanee, S. Naya and V.Vichit-Vadakan (editors), *Trade and Employment in Asia and the Pacific*, the University Press of Hawaii, Honolulu.
- Dixon, P.B., B.R. Parmenter, G.J. Ryland and J. Sutton (1977), *ORANI, A General Equilibrium Model of the Australian Economy: Current Specification and Illustrations of Use for Policy Analysis*, Vol. 2 of the First Progress Report of the IMPACT Project, Australian Government Publishing Service, Canberra, pp. xii + 297.
- Dixon, P.B., B.R. Parmenter, John Sutton and D.P. Vincent (1982), *ORANI: A Multisectoral model of the Australian economy*, Contributions to Economic Analysis 142, North-Holland Publishing Company, pp. xviii + 372.
- Dixon, P.B., M. Jerie and M.T. Rimmer (2016), "Modern trade theory for CGE modelling: the Armington, Krugman and Melitz models", *Journal of Global Economic Analysis*, vol. 1(1), pp. 1-110.
- Dixon, P.B., M. Jerie and M.T. Rimmer (DJR, 2018), *Trade theory in computable general equilibrium models: Armington, Krugman and Melitz*, Advances in Applied General Equilibrium Modelling, Springer Nature, Singapore, pp. xi + 189
- Evans, H.D. (1972), *A General Equilibrium Analysis of Protection: the effects of protection in Australia*, Contributions to Economic Analysis 76, North-Holland Publishing Company, Amsterdam.
- Feenstra, R.C., P. Luck, M. Obstfeld and K. Russ (2018), "In search of the Armington elasticity", *The Review of Economics and Statistics*, 100(1), March, pp. 135-150.
- Harrison, J., J.M. Horridge, M. Jerie and K.R. Pearson (2014), *GEMPACK manual*, GEMPACK Software, ISBN 978-1-921654-34-3, available at <http://www.copsmodels.com/gpmanual.htm>.
- Hertel, T. W., editor, (1997), *Global trade analysis: modeling and applications*, Cambridge University Press, Cambridge, UK, pp. xvii + 403.
- Horridge, M., A Meeraus, K. Pearson and T. Rutherford (2013), "Software platforms: GAMS and GEMPACK", chapter 20, pp. 1331-82, in P.B. Dixon and D.W. Jorgenson (editors) *Handbook of Computable General Equilibrium Modeling*, Elsevier, Amsterdam.
- Johansen, L. (1960), *A multi-sectoral study of economic growth*, Contributions to Economic Analysis 21, North-Holland Publishing Company, pp. ix + 177.

- Krugman, P. (1980), "Scale economies, product differentiation, and the pattern of trade", *American Economic Review*, 70(5), pp. 950-959.
- Melitz, M.J. (2003), "The impact of trade on intra-industry reallocations and aggregate industry productivity", *Econometrica*, 71(6), pp. 1695-1725.
- Melitz, M. J., and Redding, S. J. (2015). "New Trade Models, New Welfare Implications," *American Economic Review*, 105(3), pp. 1105-1146.
- Zhai, F. (2008), "Armington meets Melitz: introducing firm heterogeneity in a global CGE model of trade", *Journal of Economic Integration*, Vol. 23(3), September, 2008, pp. 575-604.