Exploiting complementarity in applied general-equilibrium models: endogenizing zeros, firm and mode types, capacity constraints

James R. Markusen
Department of Economics
University of Colorado, Boulder
http://spot.colorado.edu/~markusen

Abstract

Applied general-equilibrium (AGE) models have often made compromises to deal with or circumvent difficult modeling problems. One is how to model or avoid endogenous zeros. Perfect competition models: when do technologies or trade links switch from active to inactive or vice versa? Heterogeneous firms and multinational production: what types of firms are active in equilibrium and when do firms switch from exporting to foreign production? Capacity constraints: could trade links or production sectors hit capacity limits? Here I exploit the complementarity approach to general equilibrium, focusing on modeling heterogeneous firms and endogenous multinational production. Instead of the traditional continuum formulation, there is a discrete and finite set of firm types, differing in marginal costs across but not within types. There is an upper bound on the number of firms that can enter in each firm type. Formulated as a non-linear complementarity problem, we can solve for the set of active firm types in relation to characteristics of the economy such as size or trade costs and their modes of operation: no entry, domestic, exporting, multinational. The analysis incorporates endogenous markups, positive aggregate profits, and slots directly into conventional AGE models and data sets: no integrals, integration, parametric distributions or probabilistic production required.

February 2025

Additional computer code and all detailed numerical results are available from the author. A detailed literature review, derivations of endogenous markups and alternative calibration subtleties are found in my recent pedagogic paper (Markusen 2023) in the open-access *Journal of Global Economic Analysis*, so I did not reproduce those here.

Special thanks to Michael Ferris at the University of Wisconsin, Madison, Computer Science. Ferris and colleagues are the authors of PATH, the non-linear complementarity or MCP (mixed complementarity problem) solver that can accessed in GAMS, or at the website below. Michael was generous with his time in walking me through how PATH converts weak inequalities to equations with the use of slack variable following the intuition from the Karush-Kuhn-Tucker theorem.
https://pages.cs.wisc.edu/~ferris/path.html

1. Introduction

The formulation of applied (numerical) general-equilibrium models has made considerable progress since they first appeared in the 1970s. Early analyses were restricted to perfect competition, constant returns to scale, ad valorem taxes, and assumptions guaranteeing interior solutions to systems of equations. They were generally constructed and calibrated from a one-year cross-section data set. The data were essentially a one-observation set of numbers, although that one observation generally involved a great deal of work: constructing micro-consistent data for many countries, many sectors, and intermediate use matrices, with many tariff and tax wedges in the benchmark data.

Calibration of the data to a numerical model left multiple degrees of freedom over parameters such as elasticities, long a criticism made by applied econometricians. This tension has at least partially abated in recent years, with the estimation of structural models to select parameter values, with the model then being used in counterfactual analyses in a manner quite similar to that in the traditional calibrated models. A more thorough review of the history of GE modeling focusing on the underlying mathematical formulations is found in my pedagogic article: Markusen, JGEA-open-access journal (2021), so I won't repeat that here. But I will note Harris (1984) as a early example of expanding the range of model sophistication by introducing imperfect competition and increasing returns to scale into traditional models.

The purpose of this paper is to offer an alternative approach to one nagging issue of general-equilibrium modeling that has generally persisted throughout its history. This is the difficulty of endogenizing corner solutions and regime switches. These usually involve the fact that most economic variables such and prices and quantities are by nature non-negative. But with existing older mathematical formulations, it was difficult to allow for variables to take on either positive or zero values. If assumptions could be made such that positive benchmark values stayed positive (or constrained to remained zero), then the model could be formulated as a set of equations and we knew how to solve large systems of equations. Allowing for endogenous zeros required formulating a model as a system of weak inequalities.

An early and still-used tactic for ruling out zeros is referred to as the Armington assumption: goods in a sector are differentiated by country of origin and, with CES preferences, every country will always produce and export in all sectors (unless some sector or trade link is constrained to be inactive throughout). If I understand it correctly, restrictions ensuring no zeros persists today in the popular and highly cited Eaton and Kortum (2002) model. But the world trade matrix, exports by country i to country j in industry k, is full of zeros, as much as 80 percent depending on the degree of aggregation. An important step forward was provided by Eaton, Kortum and Sotelo (2013) which introduced discrete numbers of firms by type and which allowed for endogenous zeros in trade flows. A very different concept is the "balls-and-bins" model of Armenter and Koren (2014). Both Eaton-Kortum-Sotelo and Armenter-Koren exploit probabilistic production in their approaches. While the objectives of my paper and that of Eaton-Kortum-Sotelo are quite similar, I will adopt a deterministic formulation as a simple alternative to the probabilistic-production approaches.

My approach is to model general equilibrium as a complementarity problem. This approach was first formulated by Mathiesen (1985) and implemented in numerical models by Rutherford (1985, 1995). Deriving its intuition from the Karush-Kuhn-Tucker theorem (KKT: Karush 1939, Kuhn and Tucker (1951)), a model is specified as a system of weak inequalities, each with a complementary non-negative variable. If a weak inequality holds as an equation (e.g., marginal cost greater-than-or-equal to price), then the complementary variable is positive (output). If it holds as a strict inequality, then the complementary variable is zero. The insertion of "slack" variables converts weak inequalities to equations, which can be solved by algorithms such as the Newton method. This will be illustrated later in the paper. More detail is provided in Markusen (2021).

I will not attempt a general presentation here, but rather focus on using the tools of complementarity to offer an alternative way to incorporate heterogeneous firms and their mode choices (domestic, exporting, multinational) into AGE models. But first, a little background and motivaton. The late '90s and early '00s brought a major development lead by mutually reinforcing developments in empirics and theory. Empirical analysis, made possible by the availability of firm-level data, showed that exports were concentrated among a very small number of large, productive firms, as well as documenting the role of entry and exit following liberalizations (e.g., Bernard and Jensen 1999, Bernard, Eaton, Jensen and Kortum 2003). The theoretical approach that fit so perfectly with the data was Melitz (2003). Firms in a sector are heterogeneous in their marginal costs (or inversely productivity), and reductions in trade costs generally lead to a sorting among firms in which the most productive expand and begin exporting while the least productive firms exit.

The great attention and research devoted to heterogeneous firm models follows from the intersection of their theoretical appeal and their empirical relevance as revealed by firm-level data. But the analysis is complex and likely difficult for modelers to incorporate into high-dimension applied general-equilibrium (AGE) models. A new article by Balistreri and Tarr (2023) makes good progress on this, but it is clearly not a simple matter. Further, even to get to this level of complexity, a number of restrictive assumptions are typically (but not universally) made. Let me list a few of these, without in any way demeaning the great work that has been done.

First, much of the literature I am aware of continues to use "large-group" monopolistic competition (LGMC) in which it is assumed that firms are too small to affect the price index in their industries, leading to constant markups.[1] This removes a difficult endogeneity from the models, but leads to counter-empirical results such as all firms having the same price to marginal cost ratios. Second, the pattern of productivities across firms must follow a narrow class of parametric distribution functions so as to permit tractable integration. But the actual values of

[1]Earlier papers with some form of endogenous markup include Horstmann and Markusen (1992), Levinsohn (1993), Bernard et. al. (2003), Melitz and Ottaviano (2008), and Atkeson and Burstein (2008). A detailed literature review including more recent work is found in Markusen (2023). A paper that does feature positive profit income, the same Bertrand markup rule I use here, and a finite number of firms in a sector is Gaubert and Itskhoki (2021). Eaton-Kortum-Sotelo (2013) also use Bertrand conjectures.

firm productivities/sizes may depart substantially from any parametric distribution. Third, aggregate profits are disposed of by the assumption that firms must pay for "draws" to learn their productivity. This removes another awkward endogeneity in general equilibrium, allowing total income to be independent of profits.[2]

The alternative way to model heterogeneous firms in an industry offered here avoids all of the problems just mentioned and has the further advantage of a much simpler algebraic formulation. The basic concept is to break the data on the ranking of firms (e.g., by sales) into a discrete set of firm "types". There could, for example, be five firm types, the number I will use here, and these different in their marginal cost of production. There is free entry into firm types, but only up to a limited number for each type. This limit per type and the pattern of cost differences across types is the discrete equivalent of the continuous parametric distribution used in the literature. The model will solve for the set of active firm types depending on parameters such as the economy's size or trade costs. The lowest cost type(s) will earn positive profits in equilibrium and these will be added to the economy's total income.

This is a difficult problem for traditional analytical methods in economics. We have endogenous variables, the number of active firms of each type, which have not only a lower bound of zero but also an upper bound specified by the modeler. The equilibrium number of firms of a type must lie in a closed interval given by weak inequalities at both the upper and lower end. Second, it may be that the "cutoff" firm type, the most costly type that is active in equilibrium, is not given by a zero-profit condition. Because of the discrete differences in firm-type costs, the most costly firm type active in equilibrium may earn positive profits (the next more costly type would earn strictly negative profits by entering). This means that solving for the for the cutoff type using a zero profit condition, a key property in existing models, won't work.

Formulating a heterogeneous-firm model exploiting complementarity, KKT and using the PATH MCP (mixed complementarity problem) solver in GAMS and available in other software has a number of advantages, more or less the opposite of the disadvantages of the traditional continuous approach noted above. First, no integrals or integration is needed. Second, there is no need to impose any parametric distribution function on firm productivity or cost. These can be calculated directly from data once those are divided into firm classes or type. Third, there is no need to assume LGMC. Nash Cournot and Nash Bertrand, used here, can allow for different market shares and thus different markups across firm types, the added endogeneity in general equilibrium being of no consequence. Fourth, there is no need for "draws": the added endogeneity (of income to profits, firm numbers, cutoffs etc.) of positive aggregate profit income is easily incorporated.

In what follows, I first review definition of a complementarity problem, and then describe how double-sided inequalities (upper and lower bounded variables) are formulated in PATH and

[2]Another way around this endogeneity of income and profits is to assume quasi-linear preferences as in Melitz and Ottaviano (2008). But this assumes that the industries in question have zero income elasticities of demand, borderline inferior goods.

GAMS. Then I specify a basic general-equilibrium model, which I hope permits maximum clarity: a two-sector, one-factor closed-economy model. There are five firm types in one increasing-returns sector, free entry up to an upper bound on each firm type. A simple experiment is used, which is growth in the economy, a parable for adding together identical economies first exploited by Krugman (1979). I compare results under small group Nash Cournot (SGC), Nash Bertrand (SGB) and LGMC.

The second model extends the first to a two-country trade model, with a specification that closely follows that of Melitz (2003). Firms may export, but there is an added fixed cost to entering exporting. Unlike Melitz, I allow for endogenous markups and compare SGB to LGMC (the Melitz original). Paying for productivity draws is unnecessary and the aggregate profits of the active firms are added to the income of the representative consumer.

The third model allows firms to establish a foreign plant to serve that market. Firms incur a fixed cost larger than that for exporting, but do not pay a (unit) trade cost. This is close to the Helpman, Melitz and Yeaple (2004) model which adds heterogeneous firms to the earlier horizontal multinationals models of Horstmann and Markusen (1992) and Markusen (2002).[3]

The second and third models yield simulation results that resemble those in Melitz (2003) and Helpman, Melitz and Yeaple. At moderate to high trade costs, the most productive firm type(s) will choose the multinational mode (when allowed in model 3), the middle-productivity firms will choose exporting, the less productive yet serve only their domestic market and the least productive firms may not enter. The simulations trace through how the configuration of modes across firm types changes as trade costs fall from a very high level to costless trade. There are generally mode and entry switches as trade costs fall; e.g., middle-productivity firms start exporting and low-productivity types exit.

A short appendix to the paper notes that the same innovation of upper-bounded variables (or non-zero lower bounds) can be applied to compute short-run scenarios in which there are firm/industry-level capacity constraints or exit constraints. For example, port or airport capacity limits impose constraints on trade links, similarly for energy production. For two years, the business press has been full of stories about "bottlenecks" and supply constraints arising in the post-covid recovery. There is a subtlety, which is that hitting an upper bound or non-zero lower bound generates some sort of profit or loss, and this must be endogenized into income to compute valid short-run equilibria.[4]

---

[3]It is worth noting that non-linear complementarity approach derived from KKT is quite different from an integer programming approach and the continuum approach in Melitz and Helpman et. al. In KKT, variables are continuous but subject to boundaries. Markusen (2002) applies complementarity to solving for the mode choices of firms active in equilibrium: domestic, exporting, horizontal and vertical multinationals. Helpman, Melitz and Yeaple (2004) incorporate heterogeneity in sorting firms in the continuum by productivity using zero-profit or equal-profit equations to establish cutoffs.

[4]There is an important new paper by Arkolakis, Eckert and Shi (2023) which is quite different from my approach but perhaps related in spirit. Both papers are trying to capture some empirically important aspects of discreteness. Arkolakis et. al. are analyzing a firm's decision on the number and

The fact that the results of the open economy / mode choice models look familiar is, while valuable, not my only or even main selling point. It is rather that the techniques developed here are entirely straightforward extensions of standard general-equilibrium complementarity modeling frameworks (e.g., Rutherford's MPS/GE), allowing the insertion of heterogeneous firms, mode choices, endogenous markups and capacity constraints directly into existing programs and data sets.

2. Models with inequality side constraints, complementarity conditions

In this section, I review the complementarity formulation of economic equilibrium models. Suppose we have an equation or set of equations in implicit form $F(Z) = 0$, that we wish to solve. But there is a complication in that there are inequality side constraints $G(Z) \geq 0$. It may be or likely be that any solution to $F(Z) = 0$ violates one (or more) of the inequality constraints $G(Z) > 0$. A typical complementarity problem reformulates the set of equations as weak inequalities, $F(Z) \geq 0$. Each of these is matched with a specific weak inequality side constraint. The complementarity condition is that the product of the each weak inequality F times the matched weak inequality G equals zero. A complementarity problem is thus written as

$$\text{Solve} \quad F(Z) \geq 0 \qquad s.t \quad G(Z) \geq 0 \qquad F(Z)^T G(Z) = 0 \tag{1}$$

In economics, the archetype situation is that the Z are constrained to be non-negative. In such cases, $G(Z) \geq 0$ reduces to $Z \geq 0$. Later, I introduce a heterogeneous firm model in which there are two-sided bounds on Z, but for now let's stick with $Z \geq 0$. The formulation in (1) looks simple, but it does require the correct direction of the inequalities and the correct association of each member of the set $F(Z)$ with a member of $G(Z)$. These must come from the modeler's knowledge of the underlying problem.

We can develop the intuition with a simple supply-demand model, good X with price p_x, an intuition that carries through to much more complex general-equilibrium models. First, there is the optimization condition that price equals marginal cost, $mc(X)$. But a non-negativity restriction on X may mean that the solution is that marginal cost is greater than the demand price p_x at $X = 0$ and the good is not produced. Our economics tells us that the inequality cannot run the other way, price greater than marginal cost leaves an unexploited profit opportunity. So the correct formulation of the problem is

$$mc(X) - p_x \geq 0 \qquad X \geq 0 \qquad (mc(X) - p_x)X = 0 \tag{2}$$

If $mc > p$ in equilibrium, X is not produced. There are many examples of this in economics. One is that there are multiple technologies for producing X (electricity), but some are

location of foreign production plants. The problem is complex because a plant in any one location influences the profitability of all other possible plants. A brute-force algorithm would compute the firm-wide profitability of all possible combinations of plants, but this quickly because computationally huge as the number of countries (potential locations) increases. They develop instead a "shriking" algorithm which greatly reduces the computational exercise.

unprofitable under current economic conditions ((2) refers to a specific one of those technologies). Second, the world trade matrix is full of zeros: it is unprofitable at the equilibrium to export good i from country j to country k.

The other type of situation that requires a complementarity condition due to a non-negativity constraint is a market clearing equation, supply equal demand, $D(p_x)$. But it may be that the solution involves excess supply in equilibrium, implying that the price is zero.

$$X - D(p_x) \geq 0 \qquad p_x \geq 0 \qquad (X - D(p_x))p_x = 0 \tag{3}$$

If supply > demand then $p_x = 0$ and X is a free good. This outcome is less common than the optimization condition example, but situations include cases where licenses or permits are in excess supply, a quota is non-binding in equilibrium (X is the suuply of licenses or quota level), or a factor of production in excess supply with fixed-coefficient technologies.

Figure 1gives the three possible outcomes of this simple partial-equilibrium, supply-demand problem. The top panel is the classic undergraduate textbook case of an interior solution. If we knew this was going to be the case, complementarity is not needed; that is, there is no need to associate each equation to a specific variable. The middle panel shows a situation in which the supply curve lies everywhere above the demand curve. This is a case where finding the solution does require complementarity: good X will not be produced in equilibrium. The bottom panel shows the third possibility, where the supply curve lies everywhere to the right of the demand curve (in the case of X being permits or licenses, the supply curve is generally vertical). The complementarity formulation yields a solution in which the price is zero and X is a free good.

How does the modeler know which of these situations will be the solution, especially in a complex general-equilibrium model; e.g., the positions of the supply and demand curves depend on all the other variables in the model? The Karush (1939) and Kuhn and Tucker (1951) theorems come to the rescue. Their formulations are to introduce non-negative "slack" variables into the problem, which explicitly incorporate the complementarity restriction into the model and convert the weak inequalities to equations. Denote the non-negative "slack" variables as s_x, and $s_p \geq 0$. The addition of two variables requires two added equations, and these are the complementarity conditions. The problem is written as

$$mc(X) - p_x - s_x = 0 \qquad s_x = \text{unprofitability at } X = 0 \tag{4}$$

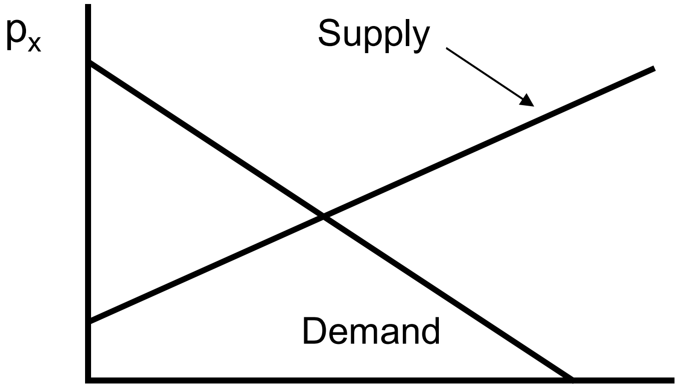
$$s_x X = 0 \tag{5}$$

$$X - D(p_x) - s_p = 0 \qquad s_p = \text{excess supply at } p_x = 0 \tag{6}$$

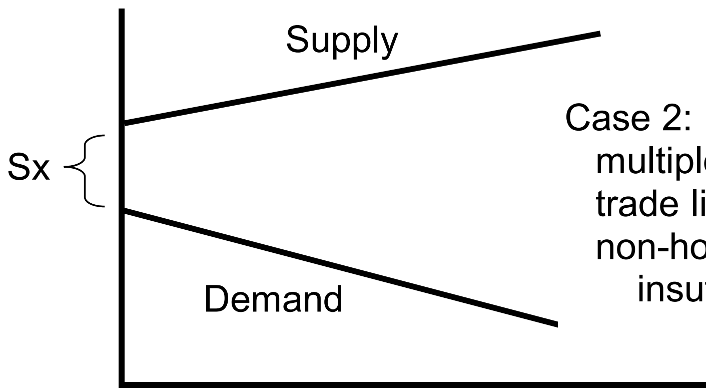
$$s_p p = 0 \tag{7}$$

We now have a model of four equations in four unknowns. This can be solved by an iterative procedure such as the Newton method, and that is done for the modeler in the PATH solver. There is no need to write a brute force procedure in which different solutions are all tried

Figure 1: Three equilibrium outcomes

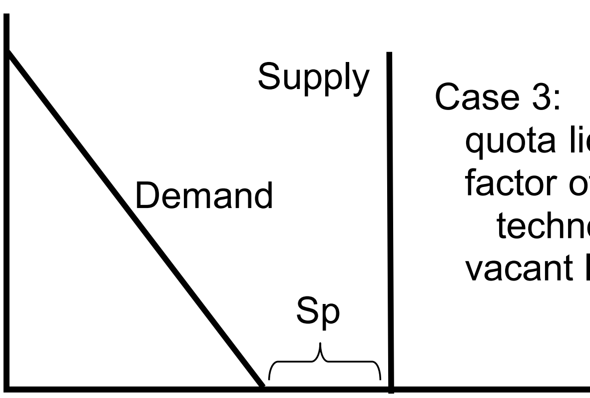


Case 1: interior solution X, $p_x > 0$



Case 2:
multiple technologies - one too costly
trade link too costly - exports zero
non-homoethic preferences - income insufficient

Case 2: X not produced (X = 0)



Case 3:
quota licenses, permits in excess supply
factor of production with fixed-coefficient technology
vacant land - beyond margin of cultivation

Case 3: excess supply, X is a free good ($p_x = 0$)

out: there are only three in Figure 1, but full GE models made have dozens to hundreds. The values of the slack variables at the solution (equal zero in an interior solution) give the unprofitability of the technology/trade link and the excess supply and correspond to the measures with the same notation in panels two and three of Figure 1.[5]

3. Complementarity and KKT with bounded variables

I will lay out a full but simple general equilibrium model in the the next section. But first I want to introduce how the PATH MCP solver handles a two-sided weak inequality (a variable has both an upper and lower bound). It is well known that when using the standard technology with constant marginal costs and a fixed cost (at constant factor prices), the zero-profit condition for a firm simplifies to markup revenues equal fixed costs. In a complementarity formulation with identical firms, the variable associated with the weak inequality is the number of firms in equilibrium: firms enter until profits for a representative firm are zero. If fixed costs are greater than (potential) markup revenue at the solution, then the number of firms is zero.

Here, we are also going to specify an upper bound to the number of firms of a given type, and in PATH, this is handled in the same way. Let *i* index firm types, differing by marginal costs. Let *FC(i)* denote the *value* of fixed costs of firm type i, and *MKR(i)* the markup *revenues* of firm type i, calculated regardless of whether or not the firm actually enters. Let *N(i)* equal the number of firms active at the solution, and let *N.up(i)* and *N.lo(i)* be parameters giving the upper and lower bounds on *N(i)* (this is actual GAMS notation). Firm profits may be positive in equilibrium, this occurring when the number of firms hits the upper bound. The lower bound will be set at *N.lo(i)* = 0 throughout.

The profit (entry) condition for type *i* can have three solutions in general equilibrium.[6]

$$FC(i) = MKR(i) \quad \Rightarrow \quad N.up(i) > N(i) > N.lo(i) \quad \text{zero profits} \tag{8}$$

$$FC(i) > MKR(i) \quad \Rightarrow \quad N(i) = N.lo(i) \quad \text{losses, no entry} \tag{9}$$

$$FC(i) < MKR(i) \quad \Rightarrow \quad N(i) = N.up(i) \quad \text{positive profits} \tag{10}$$

The key to understanding the importance of KKT is that it turns the weak *inequalities* into a formulation with three *equations*, by introducing two non-negative "slack" variables into the problem, one for each bound on the variable *N(i)*: denote *w(i)* for the lower bound, and *v(i)* for the upper bound slack variables. The way this is formulated in PATH is the following:

---

[5]KKT are the necessary conditions for a solution to a non-linear programming problem with inequality side constraints. Strictly speaking, a market clearing equation is not a KKT condition. But the logic of KKT was extended to economic equilibrium problems by Mathiesen (1985) and Rutherford (1985, 1995). This in turn became the foundation for the PATH MCP solver in GAMS. The values of the slack variables are given in the solution file in GAMS where they are called "marginals".

[6]I'll ignore knife-edge solutions, such as profits equal zero at *N(i)* = *N.lo(i)* = 0.

$$FC(i) = MKR(i) + w(i) - v(i) \tag{11}$$

$$w(i)(N(i) - N.lo(i)) = 0 \qquad w(i) = \text{(potential) losses if } i \text{ enters (defined} \geq 0) \tag{12}$$

$$v(i)(N.up(i) - N(i)) = 0 \qquad v(i) = \text{profits in equilibrium} \tag{13}$$

Equations (12) and (13) are the complementary conditions. Both *w(i)* and *v(i)* are zero in an interior (zero profit) solution. As just note, we let *N.lo(i)* = 0 (no entry). *w(i)* is positive if no entry occurs, and its value gives the (negative of) potential profits from entering. *v(i)* is positive if *N(i)* hits its upper bound, and its value gives the positive profits earned in equilibrium. If *w(i)* is positive then *v(i)* is zero and vice versa. Having converted the weak inequalities to equations, PATH then solves the (full) model by a Newton-type algorithm.

4. Incorporating endogenous markups

The next task of this section is to specify imperfect competition behavior. I'll be brief here since this is all derived in earlier paper, Markusen (2023). The representative consumer's welfare is a simple, two-level CES. The upper nest between *X*, our sector of interest and *Y*, a homogeneous good with constant returns and imperfect competition is Cobb-Douglas. The lower nest is CES with all *X* goods of all firm types being symmetric but imperfect substitutes. *X(i)* will denote the output of a representative firm of type *i*, with all firms of type *i* having identical technologies and thus producing the same amounts at the same prices in equilibrium.. *N(i)* is the number of active type *i* firms, σ is the elasticity of substitution among the *X* goods.

Utility or welfare (*W*) of the representative consumer, and the symmetry of varieties within the *X* goods allows us to write utility as follows ($0 < \alpha,\ \beta < 1$).

$$W = X_c^{\beta} Y^{1-\beta}, \qquad X_c \equiv \left[\sum_{i=1}^{I} N(i)\,(X(i))^{\alpha}\right]^{1/\alpha} \qquad \sigma = \frac{1}{1-\alpha} > 1 \tag{14}$$

where X_c is often referred to as a composite commodity or sub-utility. Let *s(i)* be the market share of a representative firm of type *i* in the total output of all *X* firms of all types. $p_x(i)$ denotes the price of all/any *X* firms of type *i*, with these prices varying across firm types.

The market share of an individual firm of type *i* is given by the first equation in (15).

$$s(i) = \frac{p_x(i)X(i)}{\sum_i p_x(i)N(i)X(i)} \qquad p_x(i)(1 - 1/\eta(i)) = mc(i) \qquad mk_i \equiv \frac{1}{\eta(i)} \tag{15}$$

The perceived price elasticity of demand for firm type i is given by η*(i)*, with the marginal revenue = marginal cost (*mc*) optimization condition given by the second equation in (15). The third equation in (15) is how I will define the markup rate (not revenue) *mk(i)* in this paper, although it is often flipped around to the other side with the markup defined as *p/mc* > 1.

In my earlier paper (Markusen 2023), I consider and compare three types of imperfect competition and will do the same here. These are large-group monopolistic competition (LGMC), the usual assumption that yields a constant perceived elasticity and markup; small-group Cournot (SGC); and small-group Bertrand (SGB). I derive the three perceived elasticities, denoted as η_c or η_b for the small group cases, and therefore their markups in the earlier paper. The Cournot formula is derived under the assumption that the firm makes a best response in quantity holding the quantities of other X firms of all types constant. The Bertrand formula holds the price of all other X firms constant. Both also hold expenditure on X goods constant. In LGMC, the perceive elasticity of firm demand is just $\eta = \sigma$. The perceived elasticities of firm demand for Cournot and Bertand are, however, more complex. These are given as follows for firms of type i.

$$\eta_b(i) = \sigma - s(i)(\sigma - 1) \qquad \text{Bertrand} \qquad (16)$$

$$\eta_c(i) = \frac{\sigma}{\sigma s(i) + (1 - s(i))} = \frac{1}{s(i) + (1 - s(i))\frac{1}{\sigma}} \qquad \text{Cournot} \qquad (17)$$

Both elasticities converge to the LGMC case of $\eta = \sigma$ as the market share of an individual firm goes to zero: LGMC is Nash if and only if $s(i) = 0$; Although the Cournot elasticity seems quite different from the Bertrand formula in (16), they have the same values at the extremes $s_i = 0$ and $s_i = 1$. Here is the comparison of (16) and (17) for a given firm type (given i):

At $s = 0$, $\eta_c = \eta_b = \sigma$ (= LGMC) (18)

At $s = 1$, $\eta_c = \eta_b = 1$ (monopoly)

For $0 < s < 1$, $\eta_c < \eta_b < \sigma$ (Cournot is less elastic)

For $0 < s < 1$ and $\sigma = \infty$, $\eta_c = \frac{1}{s}$, $\eta_b = \infty$ (perfect substitutes)

The markups as defined in (15) are just the inverse of (16) and (17) and are given by

$$mk_b(i) = \frac{1}{\sigma - s(i)(\sigma - 1)} \qquad mk_c(i) = s(i) + (1 - s(i))\frac{1}{\sigma} \qquad (19)$$

As noted in the third line of (18), the markup for Cournot will be higher than either LGMC or SGB for the same value of $s(i)$ strictly between zero and 1. In that sense, we can say that Cournot is “less competitive” than the other two.[7]

[7]The Bertrand markups formula is the same as that found in Gaubert and Itskhoki (2021).

5. Modeling entry and accounting for positive profit income

Let *CONS* be the income of the representative consumer. Firm "entrepreneurs" will be treated much like a consumer: they receive markup revenues as income and demand fixed costs. Added demand for fixed costs constitutes entry, so the variable complementary to the entrepreneur's budget balance equation is the number of firm in equilibrium. The income (from markup revenues) of a entrepreneur of firm type i is denoted *ENTRE(i)*. The markup rate (not markup revenue as in (8)) for a firm of type is denoted *mk(i)*. In this section and the subsequent one, I will use small-group Bertrand as the example in this section and the following one. I will provide the code for all three versions on request.

Let $\perp$ denoted the usual math programming symbol for complementarity: if the weak inequality holds as an equation, the complementary inequality is strict and vice versa. The key entry inequalities for a firm of type *i* under SGB are as follows, where market share *s(i)* is given in (15) above.

$$mk(i) \geq \frac{1}{\sigma - s(i)(\sigma - 1)} \qquad \perp \qquad mk(i) \geq 0 \tag{20}$$

$$ENTRE(i) \geq mk(i)p_x(i)X(i) \qquad \perp \qquad ENTRE(i) \geq 0 \tag{21}$$

$$ENTRE(i) \geq p_l fc \geq ENTRE(i) \qquad \perp \qquad N.up(i) \geq N(i) \geq 0 \tag{22}$$

where p_l is the price of labor, the single factor of production in the model to follow, and *fc* is in units of labor, not in value as in (8) (identical for all firm types). Adding the upper bound on the number of firms of a type generates an equation with two slack variables for (22) as described in the previous section.

As noted above, it is relatively easy to incorporate positive profits in equilibrium into the model. There is no need to introduce "draws" in order to eliminate aggregate profits from the model. In my version here, all active firm types earn positive profits, except in the case where the least productive active firm type just breaks even. Profits are redistributed to the representative consumer (*CONS*), so that the budget constraint for the consumer is

$$CONS \geq p_l ENDOW + \sum_i [N(i)ENTRE(i) - N(i)p_l fc] \qquad \perp \qquad CONS \tag{23}$$

where the summation term is the profits of the active firm types (*N(i)* > 0).[8] *ENDOW* is a parameter giving the economy's endowment of labor *L*, which we will vary in our experiments.

The way I have set up the model lets it compute what the optimal price $p_x(i)$ and output

[8]Gaubert and Itskhoki (2021) allow for positive aggregate profit income as noted earlier. This is distributed to the representative consumer in the same fashion as I do here.

X(i) would be for an inactive firm type, and also calculate what profit income would be (summation term in (23)), but they don't affect the solution. Note in (8), for example, that the sum of the markets shares of all firms is given by multiplying both sides by *N(i)*, since *s(i)* is the market share of an individual firm of type *i*. The sum of all market shares is then equal to one. Similarly, the bracketed term in (23) is not affected by the fact that inactive firm types (*N(i)* = 0) would earn negative profits. This way of constructing the model has the advantage that we can see just how unprofitable inactive firm types are at the solution.[9]

6. The single-economy general-equilibrium model

Both the single-economy and two-country general-equilibrium models have two goods, *X* and *Y*, and one factor of production *L*. For variables indexed by *i*, there are multiple variables and equations. There are five values of *i* in the simulations to follow in the next section, ordered by marginal costs of the firm type, with *i* = 1 being the lowest cost (most productive) firm type. The variables and parameters of the model for the Bertrand case are listed in Table 1. With five firm types, there are 37 non-negative variables in the model, not including the slack variables added by KKT.

Table 2 specifies the equations of the model, each with its complementary variable assignment. Formulating the model using GAMS and the PATH MCP solver does not require the modeler to add the slack variables and complementary-slackness equations (MCP: mixed complementarity problem). This is done by the solver: the modeler only needs to specify the weak inequalities and correct complementary variable assignments.

Units are chosen such that one unit of *Y* requires one unit of *L*, and p_y is arbitrarily used as numeraire so $p_y = p_l = 1$, with p_l then the marginal cost of *Y*. Parameter *fc* is the number of units of *L* require for one *N* (for all firm types). p_e is the standard CES price index for the composite or sub-utility X_c good, and p_w is the consumer price index (unit expenditure function) for welfare *W*. *Y* and X_c have Cobb-Douglas shares 0.5 in welfare.

mc(i) is the marginal cost in units of *L* of producing an *X* good of type *i*. These are arbitrarily chosen in the simulation model to follow, but in practice can be calculated from data

---

[9]An issue arises concerning multiple equilibria: could there be equilibria with less efficient types only active, blockading the entry of the most efficient type? I argue that this can only occur here under strong and restrictive additional assumptions, and I present this in detail in Appendix 3 below. But a brief comment here is in order. Atkeson and Burstein (2008), Eaton, Kortum and Sotelo (2013) and Gaubert and Itskhoki (2012) rule this out by assuming sequential entry, with the most efficient firm type allowed to enter first. This type of multiple equilibria doesn't occur in my model because a "fractional firm" (size between zero and one) is permitted: *N(i)* is continuous between bounds and is not constrained to integer values. This in turn sets off adjustment iterations that arrives at a solution that is independent of starting values. If only full size (integer) firms can enter, then there can exist multiple equilibria as shown in Horstmann and Markusen (1992); Markusen (2002, chapter 4) provides examples of blockaded entry (full-sized firms) but these in effect imply that first-mover advantage for the more efficient firm would eliminate the multiple equilibria exactly as in the three papers just noted.

has I have indicated above. These replace some distribution function and eliminates the need for any integration found in the conventional approach. The values of *mc(i)* and *N.up(i)* across type constitute the discrete equivalent of the continuous parametric distributions in the conventional approach. Throughout, I give *N.up(i)* the same value for all i, but these can be set independently in the code. [10]

Table 1: Variables and parameters of the single-economy simulation model

Variables:	
X(i)	quantity of one *X* good (variety) of firm type i
N(i)	number of active *X* sector firms of firm type i
Y	total output of *Y*
W	welfare
$p_x(i)$	price of one *X* variety of firm type i
p_y	price of one unit of *Y*
p_w	price of one unit of welfare *W* (consumer price index)
p_l	price of one unit of labor *L*
CONS	income
ENTRE(i)	entrepreneur revenue
sh(i)	market share of an individual firm of type i
mk(i)	markup on an *X* good of type i
p_e	price index for composite *X* goods (sub-utility)
Parameters:	
ENDOW	labor endowment
σ	elasticity of substitution among *X* goods
fc	fixed cost of *X* in units of labor *L*
mc(i)	marginal cost of *X* for firm type i in units of labor

With 5 firm types, 37 weak inequalities in 37 non-negative variables

[10]A practical issue arises in assigning marginal cost values to firm types that are inactive in benchmark data. This is done automatically in the Melitz tradition: data on active firm types is used to calibrate the parameters of a Pareto distribution, and then that calibrated distribution is in turn used to assign values to inactive firm types in the continuum. A similar procedure could be used here: active firms types are divided into bins, with costs calculated from their sales and market shares. Then inferences are made about inactive firm types (which could be done by fitting a functional form to the active types). But in either case, assigning values to latent types (or inactive trade links) is guess work.

Table 2: Simulation model formulated as a non-linear complementarity problem

Pricing inequalities		Complementary non-negative variables	
$p_l mc(i) \geq p_x(i)(1 - mk(i))$	$\perp$	X(i)	(24)
$ENTRE(i) \geq p_l fc \geq ENTRE(i)$	$\perp$	$N.up(i) \geq N(i) \geq 0$	(25)
$p_l \geq p_y$	$\perp$	Y	(26)
$p_e^{0.5} p_y^{0.5} \geq p_w$	$\perp$	W	(27)
Market-clearing inequalities			
$X(i) \geq p_x(i)^{-\sigma}(p_e^{(\sigma - 1)})\, CONS/2$	$\perp$	$p_x(i)$	(28)
$Y \geq CONS/(2p_y)$	$\perp$	p_y	(29)
$W \geq CONS/p_w$	$\perp$	p_w	(30)
$ENDOW \geq Y + \sum_i N(i)(mc(i)X(i) + fc)$	$\perp$	p_l	(31)
Income balance			
$CONS \geq p_l L + \sum_i (ENTRE_i - p_l fc)N_i$	$\perp$	$CONS$	(32)
$ENTRE(i) \geq p_x(i)\, mk(i) X(i)$	$\perp$	$ENTRE(i)$	(33)
Behavioral and definitional inequalities			
$s(i) \geq \dfrac{p_x(i)X(i)}{\sum_i p(i)X(i)N(i)}$	$\perp$	$s(i)$	(34)
$mk(i) \geq [\sigma - s(i)(\sigma - 1)]^{-1}$	$\perp$	$mk(i)$	(35)
$p_e \geq [\sum_i N(i)p_x(i)^{1-\sigma}]^{(1/(1-\sigma))}$	$\perp$	p_e	(36)

7. Economy Size, maximum number of firms grow in proportion

Before presenting simulations, a world about calibration when comparing differing formulations is in order. I initially calibrated the model above to an economy with an endowment of 400 units of labor and markup revenues of 20 percent of *X* sector sales. The calibration assumed no constraint on the number of firm of each type, so in the replication check only the most productive firms are active in equilibrium. Three versions of the model are calibrated to the same data. Three version of the model are Cournot (SGC), Bertrand (SGB) and large-group monopolistic competition (LGMC). I include the code for the SGB two-country case with and without mnes allowed at the end of the paper. The spread in marginal costs for the five firm type *mc(i)* are given just below the declaration of parameters

I will try to be brief here, since I discussed the micro-consistency requirements of the data matrix, and calibration subtleties and trade-offs in my earlier paper (Markusen 2023). As AGE modelers know well, if we calibrate the same data to different underlying theoretical formulations, changing one thing such as the markup rule, requires balancing that change by altering at least one other parameter. Referring back to the markup rules in (19), changing from one to the other requires changing the number of firms (the market share s) and/or changing the elasticity of substitution σ. If the initial number of firms is to be held roughly constant, then a large change in σ may be required or vice versa. Large calibrated numbers of firms in SGC and SGB will mean very weak pro-competitive and firm-scale effects in counter-factuals. But a change in σ instead mean that the counter-factuals are comparing economies with different preferences. Here, I will present results based on one of the calibrations in my earlier paper.

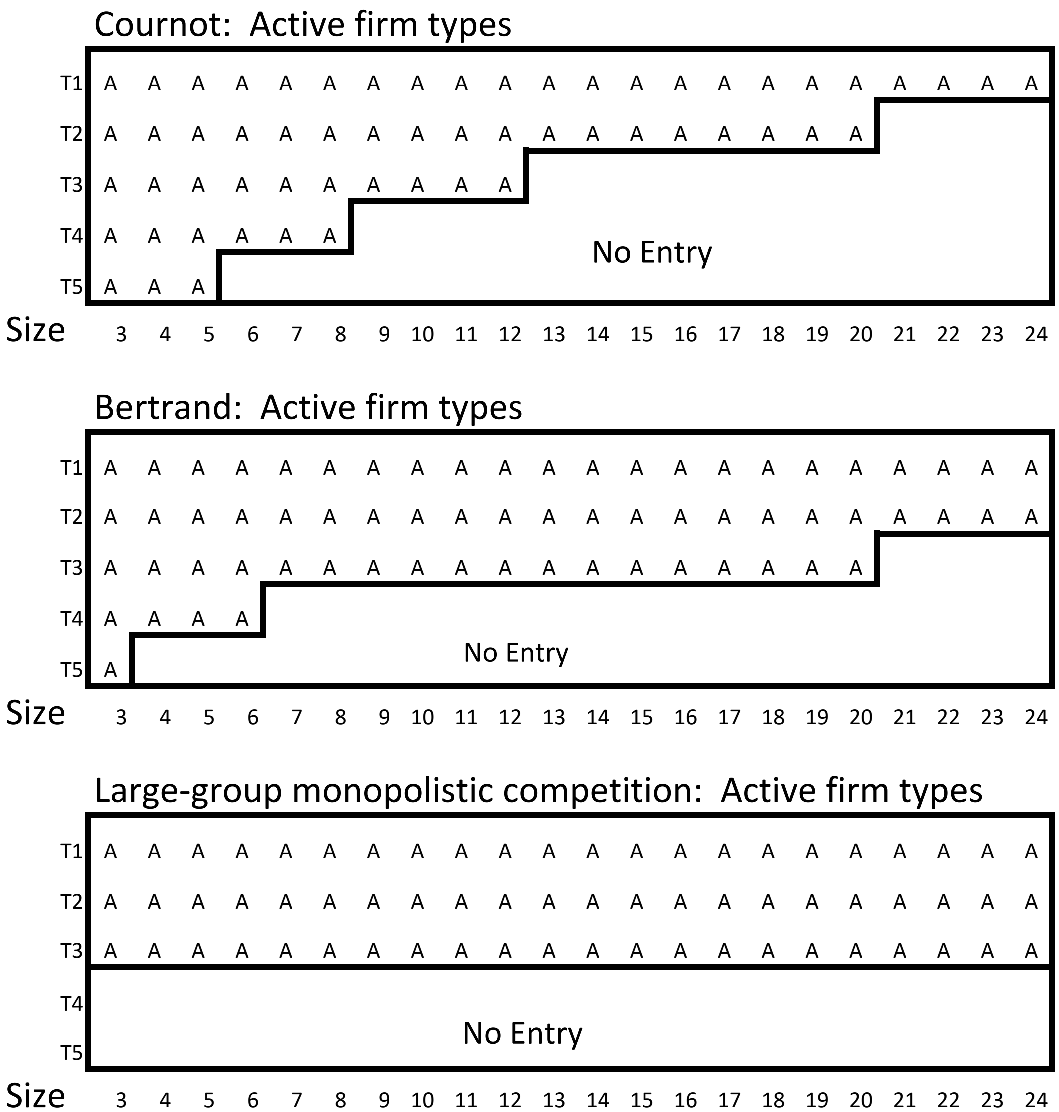
SGC:	$\sigma = \infty$	$s = 0.20$	$(N = 5)$	
SGB:	$\sigma = 6.3333$	$s = 0.25$	$(N = 4)$	(37)
LGMC:	$\sigma = 5$			

All three values in (37) yield a markup $1/\eta_c = 1/\eta_b = 1/\eta_{lg} = 0.2$. Assuming perfect substitutes in the Cournot case is conceptually useful, because then the Cournot example contrasts sharply with LGMC: in the former, welfare gains from a larger economy are purely in the form of increase firms scale (productivity) and lower markups. In LGMC, gains are purely in the form of more variety. An arbitrary number of firms will be used in LGMC version of the model (scaled to equal the total benchmark *X* output), but that doesn't affect the solution since *s* does not enter the markup formula: $mk = 1/\sigma$.

Figure 2 presents results. I am keeping the experiments very simple in order to show how my formulation produces clear and intuitive results for basic questions. For Figure 2, each of the three cases is run in a loop over the size of the economy (ENDOW) in unit of labor, giving the horizontal axis values. The vertical axis indexes the five firm types, T1 being the lowest cost (most productive) The maximum number of (actual or potential) firms of each type that can enter (*N(i).up*) is assumed to grow in strict proportion to the size of the economy. As mentioned earlier, this is Krugman's (1979) parable for adding together identical economies.

The results in Figure 2 for the Cournot and Bertrand cases show the importance of the

Figure 2: Adding together identical economies:
max firm number N.up(i) increases in proportion to size
size in '000 units of labor A = active firm type



variable markup assumption. As the economy grows (or identical economies added), the number of active firm types shrinks, with the most costly firm types exiting one after another. As the economy and *N.up*(i) grow together, there is entry of more firms of the most productive firm types. The market shares of these firms decrease, their markups fall and firm scale increases. This forces down the prices of the *X* goods and leads to losses for the remaining least productive firm types, which then exit.

However, the LGMC case in Figure 2 is a stark comparison and highlights the limitations of this traditional approach. Because markups are fixed at 1/σ for all firm types, growing the economy and *N.up*(i) in proportion to it just replicates quantities and prices, and active firm numbers for each firm type grow in the same proportion. Welfare increases, but there are no firm scale effects, no pro-competitive effects, no firm-type-selection effects, and markups and *p/mc* ratios for all firms are the same. I believe that all of these things are counter empirical. Figure 2 also suggests caution in making simple general statements such as trade causes sorting and exit: this is not true in the Krugman experiment under LMGC.

Simulation results behind Figure 2, shown in Table 3, emphasize the differences in the composition of the effects contributing to welfare changes. In particular, these numbers emphasize that welfare gains from LGMC are purely in the form of increased variety. Cournot, with the goods being perfect substitutes, is the oppose extreme. There is a small increase in firm numbers, but this has no variety effect on welfare. Instead, the welfare increases due to a strong increase in firm scale (lower average cost) and a large fall in markups. Bertrand is in between, with added variety, higher firm scale, and lower markups all contributing to welfare.

Table 3: Moving left (small economy) to right (large economy) in Figure 2

	LGMC	Bertrand	Cournot
Economy size increases by a factor of:	8.00	8.00	8.00
Firm numbers (variety) increase by a factor of:	8.00	3.43	1.60
Total output of X goods increases by a factor of:	8.00	13.90	18.95
Markup of most productive firm change (percent):	0.00	-48.25	-53.57
Percent change, welfare/capita:	2.97	2.64	2.22

Proportional changes in welfare per capita over the whole size range don't show the path. Cournot welfare increases quickly when size is initially small, as firms move down the steep section of their average cost curve to higher productivity. But this effect diminishes when size becomes large. There is less increase in firm scale and the average cost curve is flatter when the economy is already large, and there is no variety gain from more firms with perfect substitutes. But a (often disparaged) property of CES preferences means that added varieties keep on contributing to welfare under LGMC even when economies are initially large. I discuss this in Markusen (2023). The implication for modeling is that initially small economies benefit greatly from trade under SGC or SGB, but this effect diminishes relative to LGMC when economies grow large.

8. Two-country trade model

Consider next a two-country trade model, where each country is identical to the closed-economy version presented in the previous sections and therefore identical to one another. There continues to be a single factor L, and no pattern of comparative advantage between the two countries. There is no entry cost ("draws") so there are positive aggregate profits in equilibrium which are added to income. There is a fixed cost to exporting in addition to an iceberg trade cost. The experiment considered is changes in (symmetric) trade costs between the two countries. The GAMS model for the Bertrand case, with or without the multinational option added, is included at the end of the paper.[11]

For the differentiated-goods Bertrand case used in this section, I will assume that firms can price exports and domestic sales independently (segmented markets) which has two consequences. First, firms will charge different markups on domestic sales and export sales in the presence of trade costs.[12] This works through the market share variables in the markup equations for Bertrand. Second, within a country, imported varieties and domestic varieties will sell for different prices. Trade costs are modeled as a iceberg cost rather than as a tariff. *tc* is the gross trade costs (one plus the rate). If the home marginal cost is *mc*, the cost of an exported unit is *mc*tc*: the trade cost enters the pricing equation in the same way a tax does. But the trade cost shrinks the amount received rather than creating an income stream. If the amount exported is X, the amount received is X/tc. The revenue receive by the exporter then equals the expenditure by the importer, a condition for a valid equilibrium. I continue to use the same five-firm-type formulation (set i), although the numerical values of marginal cost have been adjusted.

The two countries are denoted with subscripts h and f as in home and foreign. In addition to firm types indexed by marginal costs, for each of these types the are two -sub-types (call them modes): domestic sales only and domestic plus exports. Subscript d will denote a domestic-sales-only firm, and subscript x will denote a firm that also exports. There cannot exist a firm that only exports with symmetric countries. For a firm of a given type, domestic sales will be the same whether or not it also exports, which economizes a bit on notation. Output of a firm of type i is given by $X_{jk}(i)$ which is production in country j for sale in country k. Domestic sales only firms will have $X_{hf}(i)$ or $X_{fh}(i)$ equal to zero.

There is free entry into what are now five firm types each with two modes. With a fixed exporting cost, more productive firms will export and less productive but active types will only

---

[11] The model can be written more compactly using more sets for goods and countries, but I think the longer version I include may be more transparent. I also assume no trade costs for Y in this section, which allows for a simpler representation of Y production and trade, but I have a fuller version that allows for tariffs and/or trade costs on Y and which is more useful when considering asymmetric countries. The code is written so that making countries asymmetric in technologies, size, etc. is trivial.

[12] With the countries symmetric in this section, an arbitrage constraint is never binding. The price of a variety in the importing country is higher than in the exporting country. Though pass through is incomplete, it is not profitable to re-export an imported good.

sell domestically as in the Melitz tradition. The difficulty is that in a different sense the two firm modes are the same firms and the limit on each firm type must be on the sum of the exporting and strictly domestic modes of each type. Let $\bar{N}(i)$ denote the maximum number of a type that can enter (ignoring country subscripts), which will be set equal to one for all firm types in the simulations below.[13] Denoting the number of exporting firms and the number of domestic-sales-only firms as $N_x(i)$ and $N_d(i)$, the constraints for the five types are now weak inequalities

$$\bar{N}(i) \geq N_x(i) + N_d(i) \quad \perp \quad \lambda(i) \qquad (\bar{N}(i) \text{ is a parameter, } N(i) \text{ are variables}) \tag{38}$$

This added weak inequality requires a complementary variable, denoted $\lambda(i)$ in (38), which must appear somewhere else in the model. The approach I am implementing is to introduce what is in effect a tax on fixed costs (one for each country and firm type in units of labor) that raises the fixed costs for both domestic and exporting firms if the unrestricted number of firms violates (38). Let $\lambda(i)$ be the shadow tax on fixed costs for a particular country, with $\lambda(i)$ complementary to (38). Let fc_d and fc_x denote the fixed costs of a domestic and exporting firm (same for all firm types) respectively, with subscripts d and x on entrepreneur's markup income for the two modes. In the code, the equations complementary to the number of firms of each type (fixed costs greater-than-or-equal-to entrepreneur markup revenues) are then given by

$$p_l fc_d + \lambda(i) \geq ENTRE_d(i) \quad \perp \quad N_d(i) \tag{39}$$

$$p_l fc_x + \lambda(i) \geq ENTRE_x(i) \quad \perp \quad N_x(i) \tag{40}$$

Consistent and valid general-equilibrium solutions require that this $\lambda(i)$ be accounted for somewhere else in the model, otherwise there will be a residual imbalance which invalidates any "solution". So I have modeled it as a virtual tax, with the tax revenue being returned to the representative consumer. Or we can just say that profits are the property or the representative consumer. Firm (retained) profits under this scheme are zero: $\lambda(i)$ and profits are both zero if the number of firms is less than $\bar{N}(i)$ in equilibrium, and $\lambda(i)$ is endogenously set to make (39) and (40) equations when the number of firms hits $\bar{N}(i)$. With tax revenue returned to or profits assigned to the consumer, the income balance constraint for country h is given as follows.

$$CONS_h = p_{lh} ENDOWH + \sum_i N_{hd}(i) p_{lh} \lambda_h(i) + \sum_i N_{hx}(i) p_{lh} \lambda_h(i) \tag{41}$$

Table 4 gives the variables of the model. Trade costs on X are the same in both directions and there is no trade cost for Y without loss of generality. Again, this model is a straightforward extension of the single-economy model of the previous section, differing

[13] $\bar{N}(i)$ could be declared as a variable and then follow the same procedure as used above and specify and upper bound on $\bar{N}(i)$. Here I use a different tactic, which I and others developed in the early '90s for modeling quantitative restrictions such as quotas by using an endogenous tax rate.

primarily in number of dimensions. These are due to the ability of the firms to discriminated on markups and outputs to the two markets, added firm types by country and modes, markups, and entrepreneurs The full GAMS model is given at the end of the paper.[14]

Table 5 gives the full specification of the model. This specification of a general-equilibrium model follows the format of Mathiesen (1985) and Rutherford (1995), which treats general equilibrium as a sequence of complementarity problems. For those who have seen this formulation before, an important characteristic of my approach is that the inequality set in Table 4 adds in heterogeneous firms in a simple and straight forward way. The only additions required to a simple single-firm-type model is the set dimension of i, the constraint equation on the sum of mode d and x firms, and the virtual tax revenue added to the consumer income definition. Put differently, this methodology allows heterogeneous firms to be added to large-dimension general-equilibrium models with minimal complexity.

Pushing the ideal of simplicity a little further, I note that the model in Table 5 converts to a single firm-type model by simply setting the dimension of set i to a singleton. The model converts to LGMC by fixing the markup variables to σ (GAMS then drops the complementary equations from the model) and changing σ to the value 5 as shown in (37). Converting the model to Cournot is only slightly more complicated if the X goods are perfect substitutes: the demand equations have to be modified and the markup equations simplified as noted earlier.

[14] I report only identical country results here, but the model can handle most any type of asymmetries such as country size or marginal costs (comparative advantage) as noted earlier, and also tariffs on X or Y.

Table 4: Variables and parameters of the two-country Bertrand model

Variables complementary to pricing inequalities

$X_{hh}(i)$	X good produced in h and sold in h, firm type i
$X_{hf}(i)$	X good produced in h and sold in f, firm type i
$X_{ff}(i)$	X good produced in f and sold in f, firm type i
$X_{fh}(i)$	X good produced in f and sold in h, firm type i
$N_{hd}(i)$	number of country h domestic X sector firms of firm type i
$N_{hx}(i)$	number of country h exporting X sector firms of firm type i
$N_{fd}(i)$	number of country f domestic X sector firms of firm type i
$N_{fx}(i)$	number of country h exporting X sector firms of firm type i
Y_h, Y_f	total output of Y in countries h and f
W_h, W_f	welfare of countries h and f

Variables complementary to market clearing inequalities

$p_{xhh}(i)$	price of X good produced in h and sold in h, firm type i
$p_{xhf}(i)$	price of X good produced in h and sold in f, firm type i
$p_{xff}(i)$	price of X good produced in f and sold in f, firm type i
$p_{xfh}(i)$	price of X good produced in f and sold in h, firm type i
p_y	world price of Y (complementary to trade balance condition)
p_{wh}, p_{wf}	price of one unit of welfare W (consumer price index) in h and f
p_{lh}, p_{lf}	price of one unit of labor L in countries h and f

Variables complementary to income balance inequalities

$CONS_h$,	consumer income in h (includes profit income, virtual tax revenue)
$CONS_f$	consumer income in f (includes profit income, virtual tax revenue)
$ENTRE_{hd}(i)$	entrepreneur markup revenue country h domestic firm type i
$ENTRE_{hx}(i)$	entrepreneur markup revenue country h exporting firm type i
$ENTRE_{fd}(i)$	entrepreneur markup revenue country f domestic firm type i
$ENTRE_{fx}(i)$	entrepreneur markup revenue country f exporting firm type i

Variables complementary to definitional inequalities

p_{eh}, p_{ef}	price index for X goods in countries h and f
$mk_{xhh}(i)$	markup on X good produced in h and sold in h, firm type i
$mk_{xhf}(i)$	markup on X good produced in h and sold in f, firm type i
$mk_{xff}(i)$	markup on X good produced in f and sold in f, firm type i
$mk_{xfh}(i)$	markup on X good produced in f and sold in h, firm type i

Auxiliary variables (virtual taxes) to implement firm number constraint

$\lambda_h(i)$	virtual tax on entrepreneur income country h, firm type i
$\lambda_f(i)$	virtual tax on entrepreneur income country f, firm type i

Parameters:

$ENDOW_h$	labor endowment of countries h
$ENDOW_f$	labor endowment of countries f
σ	elasticity of substitution among X goods
fcd, fcx	fixed cost of X in units of labor L, domestic and exporting firms
$mc(i)$	marginal cost of X for firm type i in units of labor
tc	gross trade cost (1 + trade cost rate)
$\bar{X}(i)$	maximum number of type i firms, set = 1 for all types, both countries

With 5 firm types, 145 weak inequalities in 145 non-negative variables

Table 5: Inequalities of the two-country Bertrand model

Pricing inequalities, quantities complementary variables

$$p_{lh} mc(i) \;\geq\; p_{xhh}(i)(1 - mk_{hh}(i)) \qquad \perp \quad X_{hh}(i)$$
$$p_{lh} mc(i) tc \;\geq\; p_{xhf}(i)(1 - mk_{hf}(i)) \qquad \perp \quad X_{hf}(i)$$
$$p_{lf} mc(i) \;\geq\; p_{xff}(i)(1 - mk_{ff}(i)) \qquad \perp \quad X_{ff}(i)$$
$$p_{lf} mc(i) tc \;\geq\; p_{xfh}(i)(1 - mk_{fh}(i)) \qquad \perp \quad X_{fh}(i)$$

$$fc_d mc(i) + \lambda_h(i) \;\geq\; ENTRE_{hd}(I) \qquad \perp \quad N_{hd}(i)$$
$$fc_x mc(i) + \lambda_h(i) \;\geq\; ENTRE_{hx}(I) \qquad \perp \quad N_{hx}(i)$$
$$fc_d mc(i) + \lambda_f(i) \;\geq\; ENTRE_{fd}(I) \qquad \perp \quad N_{fd}(i)$$
$$fc_x mc(i) + \lambda_f(i) \;\geq\; ENTRE_{fx}(I) \qquad \perp \quad N_{fx}(i)$$

$$p_{lh} \;\geq\; p_y \qquad \perp \quad Y_h$$
$$p_{fl} \;\geq\; p_y \qquad \perp \quad Y_f$$
$$p_{eh}^{0.5} p_y^{0.5} \;\geq\; p_{wh} \qquad \perp \quad W_h$$
$$p_{ef}^{0.5} p_y^{0.5} \;\geq\; p_{wf} \qquad \perp \quad W_f$$

Market clearing inequalities, prices complementary variables

$$X_{hh}(i) \;\geq\; p_{xhh}(i)^{-\sigma} (p_{eh}^{(\sigma-1)})\, CONS_h/2 \qquad \perp \quad p_{xhh}(i)$$
$$X_{hf}(i)/tc \;\geq\; p_{xhf}(i)^{-\sigma} (p_{ef}^{(\sigma-1)})\, CONS_f/2 \qquad \perp \quad p_{xhf}(i)$$
$$X_{ff}(i) \;\geq\; p_{xff}(i)^{-\sigma} (p_{ef}^{(\sigma-1)})\, CONS_f/2 \qquad \perp \quad p_{xff}(i)$$
$$X_{fh}(i)/tc \;\geq\; p_{xfh}(i)^{-\sigma} (p_{eh}^{(\sigma-1)})\, CONS_h/2 \qquad \perp \quad p_{xfh}(i)$$

$$Y_h + Y_f \geq CONS_h/(2p_y) + CONS_f/(2p_y) \qquad \perp \quad p_y$$
$$W_h \;\geq\; CONS_h/p_{wh} \qquad \perp \quad p_{wh}$$
$$W_f \;\geq\; CONS_f/p_{wf} \qquad \perp \quad p_{wf}$$

$$ENDOW_h = Y_h + \sum_i [(N_{hd}(i) + N_{hx}(i))X_{hh}(i) + N_{hx}(i)X_{hf}(i)]mc(i) + \sum_i [N_{hd}(i)fc_d + N_{hx}(i)fc_x] \qquad \perp \quad p_{lh}$$
$$ENDOW_f = Y_f + \sum_i [(N_{fd}(i) + N_{fx}(i))X_{ff}(i) + N_{fx}(i)X_{fh}(i)]mc(i) + \sum_i [N_{fd}(i)fc_d + N_{fx}(i)fc_x] \qquad \perp \quad p_{lf}$$

Income balance inequalities, incomes complementary variables

$$CONS_h \;=\; p_{lh}ENDOW_h \;+\; \sum_i \,[N_{hd}(i) \;+\; N_{hx}]\,p_{lh}\lambda_h(i) \qquad \perp \quad CONS_h$$
$$CONS_f \;=\; p_{lh}ENDOW_f \;+\; \sum_i \,[N_{fd}(i) \;+\; N_{fx}]\,p_{lf}\lambda_f(i) \qquad \perp \quad CONS_f$$

$$ENTRE_{hd}(i) \;\geq p_{xhh}(i)\, mk_{hh}(i) X_{hh}(i) \qquad \perp \quad ENTRE_{hd}(i)$$
$$ENTRE_{hx}(i) \;\geq p_{xhh}(i)\, mk_{hh}(i) X_{hh}(i) \;+\; p_{xhf}(i)\, mk_{hf}(i) X_{hf}(i)/tc \qquad \perp \quad ENTRE_{hx}(i)$$
$$ENTRE_{fd}(i) \;\geq p_{xff}(i)\, mk_{ff}(i) X_{ff}(i) \qquad \perp \quad ENTRE_{fd}(i)$$
$$ENTRE_{fx}(i) \;\geq p_{xff}(i)\, mk_{ff}(i) X_{ff}(i) \;+\; p_{xfh}(i)\, mk_{fh}(i) X_{fh}(i)/tc \qquad \perp \quad ENTRE_{fx}(i)$$

Definitional inequalities, Bertrand markup inequalities, firm-number constraints

$$p_{eh} \;=\; \left[\sum_i \,(N_{hd}(i) + N_{hx}(i))p_{xhh}(i)^{(1-\sigma)} \;+\; N_{fx}(i)p_{xfh}(i)^{(1-\sigma)}\right]^{1/(1-\sigma)} \qquad \perp \quad p_{eh}$$

$$p_{ef} \;=\; \left[\sum_i \,(N_{fd}(i) + N_{fx}(i))p_{xff}(i)^{(1-\sigma)} \;+\; N_{hx}(i)p_{xhf}(i)^{(1-\sigma)}\right]^{1/(1-\sigma)} \qquad \perp \quad p_{ef}$$

$$mk_{hh}(i) \;\geq\; [\sigma \;-\; s_{hh}(i)(\sigma - 1)]^{-1} \qquad \perp \quad mk_{hh}(i)$$
$$mk_{hf}(i) \;\geq\; [\sigma \;-\; s_{hf}(i)(\sigma - 1)]^{-1} \qquad \perp \quad mk_{hf}(i)$$
$$mk_{ff}(i) \;\geq\; [\sigma \;-\; s_{ff}(i)(\sigma - 1)]^{-1} \qquad \perp \quad mk_{ff}(i)$$
$$mk_{fh}(i) \;\geq\; [\sigma \;-\; s_{fh}(i)(\sigma - 1)]^{-1} \qquad \perp \quad mk_{fh}(i)$$

(in the code, market share equations and substituted directly into markup equations)

$$s_{hh}(i) \;=\; p_{xhh}X_{hh}(i)\left[\sum_i \,(N_{hd}(i) \;+\; N_{hx}(i))p_{xhh}(i)X_{hh}(i) \;+\; N_{fx}(i)p_{xfh}X_{fh}(i)/tc\right]^{-1}$$
$$s_{hf}(i) \;=\; p_{xhf}X_{hf}(i)\left[\sum_i \,(N_{fd}(i) \;+\; N_{fx}(i))p_{xff}(i)X_{ff}(i) \;+\; N_{hx}(i)p_{xhf}X_{hf}(i)/tc\right]^{-1}$$
$$s_{ff}(i) \;=\; p_{xff}X_{ff}(i)\left[\sum_i \,(N_{fd}(i) \;+\; N_{fx}(i))p_{xff}(i)X_{ff}(i) \;+\; N_{hx}(i)p_{xhf}X_{hf}(i)/tc\right]^{-1}$$
$$s_{fh}(i) \;=\; p_{xfh}X_{fh}(i)\left[\sum_i \,(N_{hd}(if) \;+\; N_{hx}(i))p_{xhh}(i)X_{hh}(i) \;+\; N_{fx}(i)p_{xfh}X_{fh}(i)/tc\right]^{-1}$$

$$1 \;\geq\; N_{hd}(i) \;+\; N_{hx}(i) \qquad \perp \quad \lambda_h(i)$$

$$1 \;\geq\; N_{fd}(i) \;+\; N_{fx}(i) \qquad \perp \quad \lambda_f(i)$$

Figure 3 presents the results for a simulation, looping over trade costs *tc*. A value high enough to induce autarky is on the left side of the horizontal axis, and free trade is on the right edge.[15] The small-group Bertrand case is used, and countries are identical so only one is shown. The fixed costs of being an exporting firm are set at 75 percent higher than for a domestic sales only firm. This value is not based on any data, but rather chosen simply to produce interesting results. $\bar{N}(i)$ = 1 for all firm types for both countries, with type T1 in the top row of Figure 3. Results on active firm types in Figure 3 duplicate the finding we have come to know from Melitz (2003) and successors. As trade costs fall, the most productive type T1 begins exporting, while types 2 and 3 enter exporting successively as costs fall. This increases supply from the other country, which in turn reduces profits and leads to exit of less productive firms which cannot afford to export (types 4 and 5 exit).

Figure 4a shows the output per firm for the five firm types over the range of trade costs. Output jumps up when a firm begins exporting: domestic sales actually fall due to competition with imports from the other symmetric country, but this is more than offset by increased exports. At the same time, the increased imports from the other country reduce the markups and profits of the less productive firms on their home sales which cannot afford to export, and so first type T5 exits and then T4 as well as shown in Figure 3 and 4a.[16]

Figure 4b plots the markups for the two most productive firm types and shows both the domestic and export markups. The segmented market assumption (arbitrage constraint not binding) means that, for positive trade costs, the domestic markup exceeds the export markup when exports are positive. This is working through the market share variables in the domestic and export markup equations: With the countries symmetric, a firm's market share in the foreign market is less than its domestic share until trade costs go to zero. In alternative terminology, this can be called "partial pass through": the firm absorbs part of the trade cost through a lower markup. Second, the convergence of domestic and export markups in which the domestic markup falls and the export markup rises is also a reflection of the market share changes. In particular, the domestic markup falls due to the import competition reducing the firm's domestic market share. Third, the more productive firm has the higher markup, again a reflection of its larger market share. I believe that this is consistent with empirical evidence. Note that none of these results would be true under large-group monopolistic competition with factory-gate pricing. Every one of those markups would be the same, equal to the inverse elasticity of substitution.

---

[15]A very small trade cost tc = 1.0001 is needed to prevent model degeneracy, infinitely many solutions, all of which have the same *net* trade flows. With zero trade costs, there are no "sticky places", just "slippery spaces". Terminology borrowed from my sister Ann Markusen (1996).

[16]The curves in Figure 3a are actually discontinuous at the discrete change in trade costs that cause a jump to exporting or exit, but Excel plots them as continuous; that is, those steep bits are not actually plotting output, they are the space between two discrete values of *tc*.

Figure 3: Sequence of equilibria, autarky to free trade

Small-group Bertrand competition

D = domestic sales only E = export sales N = no entry

T1	D	D	D	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
T2	D	D	D	D	D	D	D	D	D	D	D	E	E	E	E	E	E	E	E	E	E	E	E	E	E
T3	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	E	E	E	E	E	E
T4	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	N	N	N	N	N	N	N
T4	D	D	D	D	D	D	D	D	D	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	1.28	1.26	1.25	1.24	1.23	1.22	1.21	1.19	1.18	1.17	1.16	1.15	1.14	1.13	1.11	1.1	1.09	1.08	1.07	1.06	1.05	1.03	1.02	1.01	1

Autarky Trade cost Free trade

Sequence of equilibria: autarky (high cost) to free trade

Autarky - all five types produce for dometic market

Type 1 begins exporting

Type 5 exits

Type 2 begins exporting

Type 4 exits

Type 3 begins exporting

Free trade - types 1-3 export, types 4-5 don't produce

Figure 4a: Output per firm by type
Small-group Bertrand competition

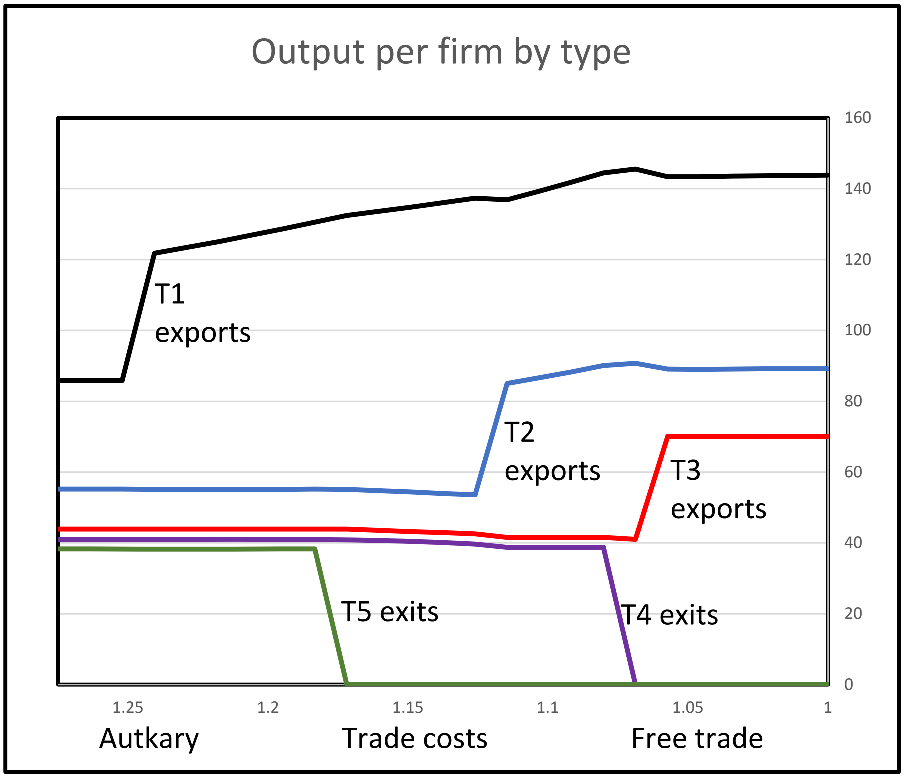
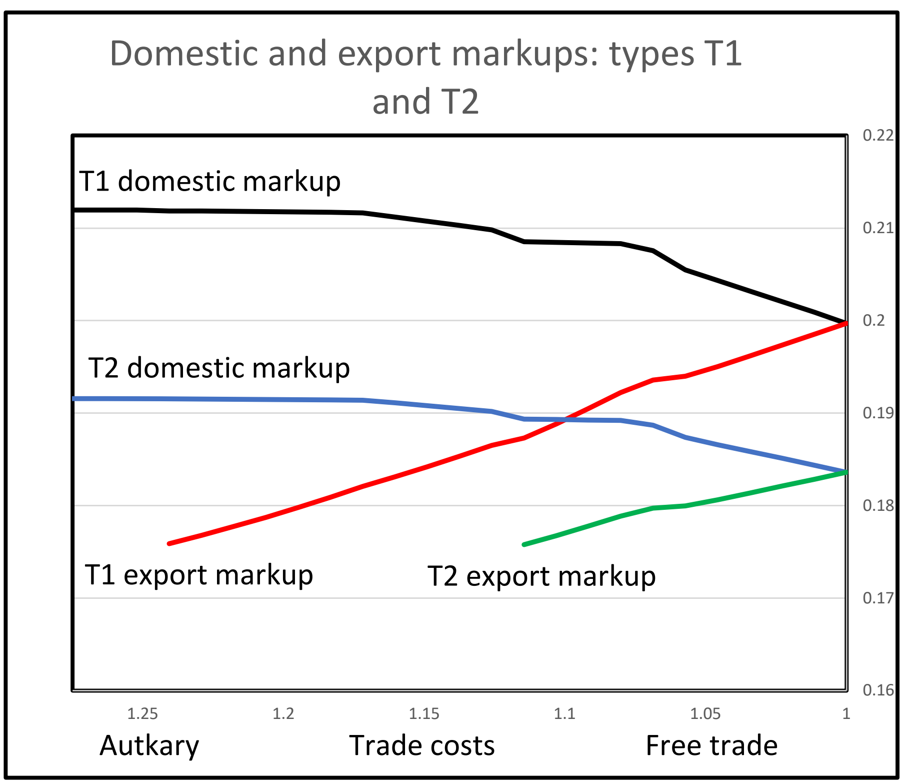


Figure 4b: Dometic and export markups
two most productive firm types: SGB



Figures 5a and 5b show some results for LGMC. Markups are fixed at 0.20 and the elasticity of substitution is changed from 6.333 to 5.0 as in (37) above in order to calibrate to the same benchmark data. Otherwise this model version is exactly the same. The point here is to check if the endogenous markups are creating some results substantially different from the traditional LGMC approach with respect to firm exit and/or entry into exporting. A broad qualitative answer, illustrated by Figures 5a,b is no. Figure 5a is quantitatively a little different from Figure 3, but qualitatively similar. I have left the scale on the horizontal axis the same for comparison purposes. Figure 5a is similar to Figure 3 in that types T2 and T3 begin exporting over the range of falling trade costs, and T5 and T4 exit.

Figure 5b completes the analysis by showing the pattern of firm output for the LGMC case. Figure 5b looks qualitatively similar to Figure 4a. In the case shown for LGMC in Figure 5a,b with type T1 already exporting on the left-hand edge, the total number of varieties available in each country equals 6 both at the left-hand edge (domestic and imported type T1 plus domestic T2-T5) and at the right-hand edge. Again, the point is that the traditional results about firm entry to exporting and firm exit under LGMC continue to hold up in my alternative formulation.

It is interesting to note that the welfare gains under LGMC (not shown) in Figures 5a, 5b are not due to increased product variety, but are due to the transfer of production from less efficient to more efficient firms and higher outputs per firm by the more efficient types. As noted to the right of Figure 5b, there are six varieties consumed in each country at both the left-hand edge and the right-hand side. The fact that the number (mass in a continuum) of firms/varieties may increase or decrease as competitive pressures increase is identified and explored systematically in Matsuyama and Ushchev (2023).[17] I should also note that welfare continues to increase on the flat sections of Figure 5b as trade costs fall even without changes in variety or firm scale. This is due to a more even consumption of domestic and foreign varieties: in the terminology and results of Arkolakis, Costinot and Rodríguez-Clare (2012), the share of expenditure on domestic varieties continues to fall with falling trade costs.

[17]We are generally conditioned to assume that a major source of gains from trade in differentiated goods models is increase variety. While I don't pursue this further here, in their important paper Matsuyama and Uschchev (2023) show that the impact on the mass of active firms (varieties) in the continuum case depends, often critically, on whether the elasticity of the distribution of the marginal cost is increasing or decreasing with Pareto-distributed productivity being the knife-edge case.

Figure 5a: Sequence of equilibria, autarky to free trade

Large group monopolistic competition

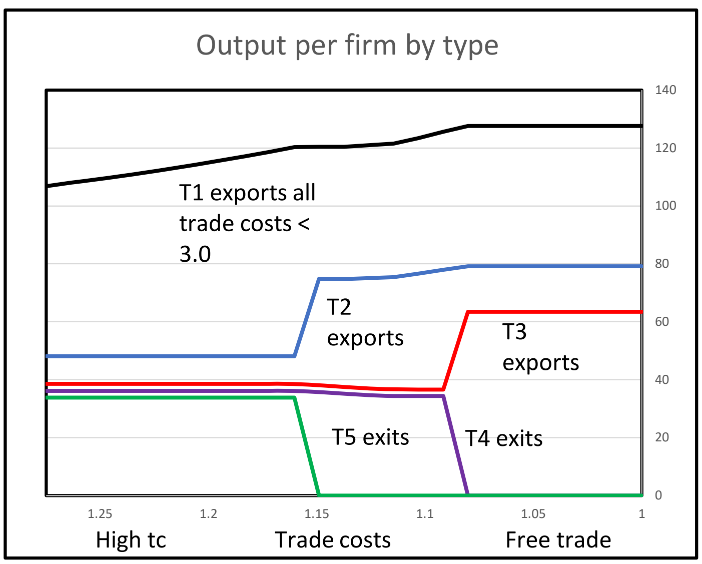
D = domestic sales only E = export sales N = no entry

T1	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
T2	D	D	D	D	D	D	D	D	D	D	D	E	E	E	E	E	E	E	E	E	E	E	E	E	E
T3	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	E	E	E	E	E	E	E	E
T4	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	N	N	N	N	N	N	N	N
T5	D	D	D	D	D	D	D	D	D	D	D	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	1.28	1.26	1.25	1.24	1.23	1.22	1.21	1.19	1.18	1.17	1.16	1.15	1.14	1.13	1.11	1.1	1.09	1.08	1.07	1.06	1.05	1.03	1.02	1.01	1

High tc Trade cost Free trade

Figure 5b: Output per firm by type

Large-group monopolistic competition



(1) There is no change in product variety moving left to right (6 firms/varieties)

(2) There are systematic results on this

Matsuyama and Ushchev (2023)

(3) All welfare gains are in

(a) replacement of inefficient firms

(b) increased output per firm

(c) more balance of domestic & foreign varieties (falling domestic share)

Arkolakis Costinot Rodríguez-Clare (2012)

9. Adding a horizontal multinational mode to serve the foreign country

The open economy model can be altered or extended to include an endogenous choice between exporting and foreign affiliate production by almost trivial changes to the formulation. I published a number of papers in the 1980s and 1990s with endogenous multinationals (many with Ignatius Horstmann, Anthony Venables or Keith Maskus), most of which are integrated and extended in Carr, Markusen and Maskus (2001) and in my MIT Press book (Markusen (2002)). But this was prior to the heterogeneous firm revolution. The "horizontal" approach was extended to heterogeneous firms by Helpman, Melitz and Yeaple (2004), using the continuum approach of Melitz (2003), solving for "cutoff conditions" to determine domestic versus exporting versus foreign production firm types.[18]

The model developed above using the fixed-cost of exporting formulation of Melitz, can be extended to a model with a foreign production option by simply adding an additional option as to the size of fixed costs and marginal costs for the multinational type and where they are incurred. I assume that the fixed cost of a foreign plant is greater than the fixed cost of exporting and that the foreign affiliate fixed cost is incurred in units of foreign labor. The marginal costs of foreign affiliate production are also in units of foreign (local) labor. The advantage of the foreign production mode is that it does not incur the unit trade costs of exporting. The multinational mode will be chosen if the added fixed cost of foreign production above that for exporting is less than the saving on trade costs. No firm will choose both foreign production and exporting in this formulation, which simplifies things a bit. As is generally true in models with horizontal multinationals, trade and foreign production are substitutes.

To include all three types of firms - domestic, exporting, and multinationals requires adding the third mode choice (MNEs) to the existing model I development above. This implies an increase in model dimensionality, but really no other complications. To simplify a little, the blocks of four inequalities in Table 5 become blocks of six. For example, there are now six markup equations instead of four with the added markups of firms choosing the multinational mode for their domestic and foreign sales. I'm guessing that the reader is not interested in seeing this extension in detail, so I will not present the equivalent of Tables 4 and 5 here. But the GAMS code is given in at the end of the paper. However, I will note that the constraint inequalities on the maximum number of firms of a type that can enter, the final inequalities in Table 5, are now given by

$$1 \geq N_{hd}(i) + N_{hx}(i) + N_{hm}(i) \qquad \perp \qquad \lambda_h(i) \qquad (42)$$

---

[18]By "horizontal" multinationals, we mean that foreign affiliates serve the host-country market, but exports back to the home country are not considered, neither are they in Helpman, Melitz, Yeaple. "Vertical" affiliates which produce abroad and export back to the parent country are included in my Knowledge Capital Model (Markusen (2002)). Vertical multinationals would never arise in the symmetric case here: with positive trade costs, a firm might have a single plant at home but would never use a foreign plant to export back to home as well or have only a foreign plant to serve home. Vertical firms only arise if countries are of different size or have different factor prices.

$$1 \geq N_{fd}(i) + N_{fx}(i) + N_{fm}(i) \qquad \perp \quad \lambda_f(i) \tag{43}$$

where the subscript *m* denotes the multinational mode. Redistributed profits for country *h* imply

$$CONS_h = p_{lh}ENDOWH + \sum_i (N_{hd}(i) + N_{hx}(i) + N_{hm})\, p_{lh}\lambda_h(i) \tag{44}$$

Results of a simulation are shown in Figure 6. I have adjusted only one parameter (fixed cost of exporting) a small amount from Figure 3 in order to show the effect of the multinational option more clearly. So the upper Panel A of Figure 6 gives the equivalent of Figure 3 (multinational mode suppressed) for the present case.

The introduction of the multinational option in Figure 6 Panel B leads to regime changes for high or moderate trade costs (left two-thirds of Panel B). At high trade costs, the two lowest cost (most productive) firm types switch to foreign production from exporting and domestic-only production for types1 and 2 respectively. The effect of this is to make these foreign firms more competitive (no trade costs) in each other's domestic markets, so type 5 never enters. At middle-level costs, type 2 switches to exporting. At slightly lower costs (further to the right), type 3 starts exporting and type 4 exits due to foreign-firm competition in its domestic market. Finally, when trade costs fall further, type 1 switch to exporting. From that point on, panel B is the same as panel A.

Figure 7 may help clarify by highlighting the cells where the introduction of multinationals causes regime switches; e.g., EM means a switch from exporting to the multinational mode. Two columns noted in the figure illustrate a possibility found in Helpman-Melitz-Yeaple (2004). These two columns correspond to the diagram in their paper in which all four modes are active over the firm types. The most productive type chooses the multinational mode, the second type chooses exporting. The third and four most production choose domestic sales only and the least productive type doesn't enter. So depending on parameter values, we can have an outcome that is qualitatively consistent with that of Helpman, Melitz and Yeaple.

Figure 6: Adding horizontal multinationals:
Sequence of equilibria, autarky to free trade

Panel A: multinational mode suppressed

T1	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
T2	D	D	D	D	D	D	D	E	E	E	E	E	E	E	E	E	E	E	E
T3	D	D	D	D	D	D	D	D	D	D	D	E	E	E	E	E	E	E	E
T4	D	D	D	D	D	D	D	D	D	D	D	N	N	N	N	N	N	N	N
T5	D	D	D	D	D	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	1.15	1.14	1.13	1.13	1.12	1.11	1.10	1.09	1.08	1.08	1.07	1.06	1.05	1.04	1.03	1.03	1.02	1.01	1.00

High tc Trade cost Free trade

Panel B: multinationals allowed

T1	M	M	M	M	M	M	M	M	M	M	M	M	M	E	E	E	E	E	E
T2	M	M	M	M	M	M	M	M	M	E	E	E	E	E	E	E	E	E	E
T3	D	D	D	D	D	D	D	D	D	D	D	E	E	E	E	E	E	E	E
T4	D	D	D	D	D	D	D	D	D	D	D	N	N	N	N	N	N	N	N
T5	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	1.15	1.14	1.13	1.13	1.12	1.11	1.10	1.09	1.08	1.08	1.07	1.06	1.05	1.04	1.03	1.03	1.02	1.01	1.00

High tc Trade cost Free trade

Both panels have the same parameter values

These figures have a slightly higher value of FCX than Figure 2 in order to show a rich set of outcomes with multinationals, but otherwise the code is the same as Figure 2

The two columns in bold, Panel B, have the full set of four modes active and correspond to the analysis of Helpman et. al. (2004)

Figure 7: Sequence of equilibria, high trade cost to free trade
No multinationals allowed to mnes permitted

Cells in bold / yellow: **regime shifts** following allowing mnes
EM means shift from export mode to mne mode, etc.

Helpman-Melitz-Yeaple (2004) region with all four modes noted

Horstmann-Markusen (1992), Markusen (2002) horizontal model

HMY region of four modes

T1	**EM**	**EM**	**EM**	**EM**	**EM**	**EM**	**EM**	**EM**	**EM**	**EM**	**EM**	**EM**	**EM**	E	E	E	E	E	E
T2	**DM**	**DM**	**DM**	**DM**	**DM**	**DM**	**DM**	**EM**	**EM**	E	E	E	E	E	E	E	E	E	E
T3	D	D	D	D	D	D	D	D	D	D	D	E	E	E	E	E	E	E	E
T4	D	D	D	D	D	D	D	D	D	D	D	N	N	N	N	N	N	N	N
T5	**DN**	**DN**	**DN**	**DN**	**DN**	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	1.15	1.14	1.13	1.13	1.12	1.11	1.10	1.09	1.08	1.08	1.07	1.06	1.05	1.04	1.03	1.03	1.02	1.01	1.00

High tc — Trade cost — Free trade

10. Summary

The more general objective of this paper is to illustrate the the advantages of formulating general-equilibrium models as non-linear complementarity problems. My view is that this allows modelers a much easier option for incorporating corner solutions, regime shifts, capacity constraints and endogenous zeros than other analytic approaches. Common assumptions in the latter are designed to force interior solutions such as the Armington assumption or Eaton and Kortum (2002) (the latter relax this in Eaton, Kortum and Sotelo (2013)).

The specific focus is to provide an alternative way of formulating models where firms in a sector are heterogeneous in terms of their variable costs or inversely productivity. Theoretically and also in empirical applications, this involves dividing firms in an industry into a discrete number of "types". There is free entry and exit into a firm type, but only up to an upper limit of firms as determined by the modeler. There is also free entry/exit across firm types. The upper limits per type along with pattern of cost differences across types replace the continuous parametric distributions imposed in the traditional approach pioneered by Melitz (2003).

There are several advantages but also disadvantages to my offered alternative. The latter first. It is difficult in my approach to find analytical solutions to the model, and analytical methods clearly remain a desirable property of economic theory. My first defense is that, with a high level of complexity and dimensionality, analytically solvable models comes with costs, requiring multiple simplifications that often eliminate many of the most interesting parts of a problem and/or clearly employ counter-empirical assumptions. Second, when it comes to performing counter-factual experiments on models calibrated to empirical estimates, a numerical version of the underlying theory is used in any case. There seems to be growing recognition that numerical models are a valid theory tool as well as an empirical one.

The advantages of my alternative were laid out in the introduction, so just a quick recap here. First, no integrals/integration is required. Second, and closely related, there is no need to impose some parametric continuous distribution of firm productivities/costs. Third, endogenous markup rules such as Nash Bertrand or Nash Cournot are easily incorporated.[18] Fourth, the addition of endogenous markups clearly eliminate one of the nagging counter-empirical implications of the traditional approach, equal price to marginal costs for all firms, and replaces it with endogenous markups increasing in firm size / productivity. Fifth, the need to eliminate aggregate profit income in order too bypass an awkward endogeneity by having firms paying for "draws" to find out their productivity is not needed. Finally, I imagine that working with real data in empirical analysis requires, perhaps, aggregating firm-size distributions into discrete size classes in any case.

The first version of the model is a simple single-economy model, with the experiment being growth in the economy, a parable for combining identical economies. Assuming growth in

[18]Again, I acknowledge that there are a number of other papers incorporating endogenous markups, and a number of these are reviewed in detail in my pedagogic paper (Markusen (2023)). A practical advantage of my formulation, using traditional CES preferences, is that it slots directly into traditional AGE models.

the maximum number of firms of each type in proportion to the country size, the Bertrand and Cournot cases, but not the large-group monopolistic-competition case, produce a reduction in the set of active firm types with growth. Endogenous firm scale and markups are important sources of welfare gains with growth.

I then develop a two-country trade model, the experiment being reduction in trade cost from autarky to free trade. I present results only for two identical countries, but the model itself can simulate most any kind of asymmetries. Bertrand competition contrasted with large-group monopolistic competition are the cases presented. The results look qualitatively very similar to the basic Melitz model. Falling trade costs lead to entry into exporting by the more productive firms and to exit by the less productive firms.[20] Both firm scale increases and markups fall significantly for surviving firms under Bertrand competition as trade costs fall resulting in non-comparative-advantage, non-variety gains from trade.

The third version of the model adds the additional option of servicing the foreign market by a foreign plant, the horizontal multinational mode. Simulation results here look closely consistent with the well-known paper by Helpman, Melitz and Yeaple (2004) with the most productive firm type choosing the multinational mode, upper-middle choosing exporting, lower middle choosing purely domestic sales and the lowest productivity not entering.

An appendix notes that the advantages of non-linear complementarity and KKT can also be applied to the introduction of firm-level capacity constraints instead of or in addition to industry level constraint on firm numbers. I'm sure that this is fundamental in operations research, and logistics and network modeling. Capacity constraints could be easily added to shipping nodes for short-run equilibria. A final appendix argues against multiple equilibria in my model, but shows how they could arise under more restrictive assumptions.

Future work should involve applying this approach to real data, with available sources (as I understand it) allowing the distribution of firm sales in an industry to be divided into discrete sets such as quintals. First attempts could be on smaller models which focus on particular key industries such as autos or semi-conductors. For generalizations to the theory, my knowledge-capital model (Markusen (2002)), focusing on differences between countries and technology stages, incorporates vertical multinationals and this could be extended in a heterogeneous firm context.[21] More generally, there is a great need to allow for endogenous zeros in the world trade matrix in large AGE models.

---

[20]I imagine it is well understood that a fixed cost to exporting is a necessary condition for the existence of domestic firms that don't export in this CES framework and that fits well with empirical evidence. The demand price for a very small quantity of a new variety goes off toward infinity at zero supply, so if a firm can enter domestically it will always export something even at very high unit trade costs in the absence of a fixed cost. But the empirical result that small firms don't export is somewhat called into question by the evidence in Bernard et. al. (2019), which suggest small firm often export through the larger firms.

[21]The knowledge-capital model allows countries to differ in size and relative endowments, and production-cost stages to have different factor intensities (firm-level fixed costs, plant-level fixed costs, unit production costs). The model is estimated in Carr, Markusen and Maskus (2001).

REFERENCES

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

Arkolakis, Costas, Fabian Eckert and Rowan Shi (2023), "Combinatorial Discrete Choice: A Quantitative Model of Multinational Location Decision", NBER working paper 31877.

Armenter, Roc and Miklós Koren (2014), "A Balls-and-Bins Model of Trade", *American Economic Review* 104(7), 2127-2151.

Atkeson, Andrew and Ariel Burstein (2008), "Pricing-to-Market, Trade Costs, and International Relative Prices", *American Economic Review* 98(5), 1998-2013.

Balistreri, Edward J. and David G. Tarr (2022), "Mathematics of Generalized Versions of the Melitz, Krugman, and Armington Models with Detailed Derivations", *Journal of Global Economic Analysis* 7(2), 66-139.

Bernard, Andrew B. and J. Bradford Jensen (1999), "Exceptional Exporter Performance: Cause, Effect, or Both?', *Journal of International Economics* 47(1), 1-25.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen and Samuel Kortum (2003), Plants and Productivity in International Trade", *American Economic Review* 93(4), 1268-1290.

Bernard, Andrew B., Emily J. Blanchard, Ilke Van Beveren, and Hylke Vandenbussche (2019), "Carry-Along Trade", *Review of Economic Studies* 86(2), 526-563.

Carr, David L., James R. Markusen and Keith E. Maskus (2001), "Estimating the Knowledge Capital Model of the Multinational Enterprise", *American Economic Review* 91, 693-708.

Eaton, Jonathan and Samuel Kortum (2002), "Technology, Geography and Trade", *Econometrica* 70(), 1741-1779.

Eaton, Jonathan, Samuel Kortum and Sebastian Sotelo (2013), "International Trade: Linking Micro and Macro", in Daron Acemoglu, Manuel Arellano, and Eddie Dekel, *Advances in Economics and Econometrics, Tenth World Congress, Volume II, Applied Economics*, Cambridge University Press, 329-370.

Gaubert, Cecile and Oleg Itskhoki (2021), 'Granular Comparative Advantage", *Journal of Political Economy* 129(3), 871-939.

Harris, Richard (1984), "Applied General Equilibrium Analysis of Small Open Economies with Scale Economies and Imperfect Competition", *American Economic Review* 74(5), 1016-1032.

Helpman, Elhanan, Marc J. Melitz and Stephen R. Yeaple (2004), "Exports versus FDI with Heterogeneous Firms", *American Economic Review* 94(1), 300-316.

Horstmann, Ignatius J. and James R. Markusen (1992), "Endogenous Market Structures in International Trade," *Journal of International Economics* 32 (1992), 109-129.

Karush, William (1939), *Minima of Functions of Several Variables with Inequalities as Side Constraints*, M.Sc. thesis, Department of Mathematics, University of Chicago.

Krugman, Paul (1979), “Increasing Returns, Monopolistic Competition and International Trade”, *Journal of International Economics* 9, 469-80.

Kuhn, Harold W. and Albert W. Tucker (1951), “Nonlinear Programming”, *Proceedings of 2nd Berkeley Symposium*. Berkeley: University of California Press, 481-492.

Levinsohn, James (1993), “Testing the Imports-as-Market-Discipline Hypothesis”, *Journal of International Economics* 35 (1993), 1-22.

Markusen, Ann R. (1996), “Sticky Places in Slippery Space: A Typology of Industrial Districts”, *Economic Geography* 72(3), 293-313.

Markusen, James R. (2002), *Multinational Firms and the Theory of International Trade*, MIT Press: Cambridge.

Markusen, James R. (2021), “Global Comparative Statics in General Equilibrium: model building from theoretical foundations”, *Journal of Global Economic Analysis* 6(2), 86-123. https://www.jgea.org/ojs/index.php/jgea/article/view/130/149

Markusen, James R. (2023), “Incorporating Theory-Consistent Endogenous Markups into Applied General Equilibrium Models”, *Journal of Global Economic Analysis* 8(2), 60-99. https://www.jgea.org/ojs/index.php/jgea/article/view/206/237

Mathiesen, Lars (1985), “Computation of Economic Equilibria by a Sequence of Linear Complementarity Problems”, *Mathematical Programming* 23, 144-162.

Matsuyama, Kiminori and Philip Ushchev (2023), “Selection and Sorting of Heterogeneous Firms through Competitive Pressures”, CEPR discussion paper DP17092,

Melitz, Marc. J. (2003), “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity”, *Econometrica* 71(6), 1695-1725.

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

Rutherford, Thomas F. (1985), *MPS/GE user's guide*, Department of Operations Research, Stanford University.

Rutherford, Thomas F. (1995), “Extensions of GAMS for complementarity problems arising in applied economic analysis”, *Journal of Economic Dynamics and Control* 19(8), 1299-1324.

Appendix 1: Economy grows but maximum number of firms stays constant

There is another simple experiment which I think may not only have implications in trade models, but also have wide applications in other types of problems. Assume that the economy grows in the same way as in Figure 1, but assume that the maximum number of firms of each type remains constant. In a trade context, this could be a case where one country is the only one which can produce in the *X* industry, and other countries are added which can only produce good *Y*. In spatial models, it captures the idea that increased supply can only be drawn from more remote sources. Closely related are applications in logistics models (e.g., airports) where there are capacity constraints on supply nodes in the network (e.g., airports) as noted in the second appendix following.

Perhaps the effect of this is pretty obvious. As the economy grows, less efficient firm types are drawn into production rather than exit. For SGC and SGB, envision the result as simply flipping Figure 2 over horizontally but preserving the labeling on the horizontal axis, so that less efficient types are added rather than subtracted as the economy grows. In contrast to Figure 2, this effect also occurs with LGMC. These results seem sufficiently intuitive that I don't think further comment is warranted.

Appendix 2: A brief note on firm-level capacity constraints[22]

It may also or instead be the case that the capacity constraint is not on the number of firms of a given productivity class, but rather on individual firms themselves. This might be important, for example, in modeling transportation networks where firms, airports or ports have capacity constraints. Spatially, I am sure that there are many examples where firms cannot physically expands beyond their current property boundaries in built up urban areas. Oil fields may have maximum extraction rates. Electricity generating facilities have capacity constraints.

The structure of PATH and KKT permits upper bounds on output variables as noted throughout. In our case here, the weak inequalities in (24) above (marginal cost greater than or equal to marginal revenue) have the outputs *X(i)* as their complementary variables. The lower bound on *X(i)* is zero, but an upper bound can *X.up(i)* ≥*X(i)* can be added. Denoting the non-negative slack variables for the lower (zero) bound and the upper bound as *w(i)* and *v(i)* respectively, the weak inequality in (24) becomes three equations in three unknowns

$$p_l mc(i) = p_x(i)(1 - mk(i)) + w(i) - v(i) \tag{45}$$

$$w(i)X(i) = 0 \tag{46}$$

$$v(i)(X.up(i) - X(i)) = 0 \tag{47}$$

---

[22] I imagine that individuals working in operations research and business logistics modeling may find little that is novel in this section. My target audience here is primarily for economists working with applied general-equilibrium and spatial models.

When the firm's output hits the upper bound, *v(i)* > 0, marginal revenue exceeds marginal cost. This in turn requires a slight reformulation of profit income for a valid solution. The (variable) profit per unit is price minus marginal cost, which is great than price times the markup.

$$p_x(i) - p_l mc(i) \geq p_x(i) mk(i) \quad \text{implying} \tag{48}$$

$$ENTRE(i) = (p_x(i) - p_l mc(i))X(i) \geq p_x(i) mk(i) X(i) \tag{49}$$

The middle term in (49) replaces the right-hand term to determine entrepreneur's income. (When the number of firms is fixed, the middle term could also be used instead of the right-hand term since the two are equal in that case.) With entrepreneur's income correctly calculated, the formula for profit income in (32) remains unchanged.

Consider the Bertrand (middle) panel in Figure 2 for example. The maximum *number of firms* of any type increases in proportion to the size of the economy. Assume that we constraint *output per firm* to equal the output per firm of the most productive type T1 at a small size. With reference to Figure 2, the result is that types T1-T4 all remain active throughout the entire range of costs. For the particular parameter values used in this experiment (identical to Figure 2), the capacity constraint is only binding on type T1 over this range of economy sizes. Output per firm of types T2-T4 increase steadily, but they do not hit the firm output constraint over this range. The same option is available to modelers for setting an upper bound on a trade link (activity) motivated by a port or airport constraint. This option and the one in the previous appendix 1 can be useful in computing short-run effects of parameter changes in AGE models.

I don't want to push any specific conclusions from this discussion since it is not the focus of the paper and I have not looked into firm-level capacity constraints further. My point in this short appendix is similar to the point of Appendix 1: the methodology of using complementarity to introduce upper bounds on variables or fix variables may have a lot of application in general-equilibrium models. However, how to implement this is not trivial. Fixing a variable creates a surplus or deficit such as positive or negative profits. This requires that this imbalance is incorporated elsewhere in the model, such as in the representative consumer's budget constraint (or in a government budget balance equation if modeling the public sector) for it to be a valid short-run equilibrium. If, for example, losses or positive profits are ignored, the welfare number produced by the (disequilibrium) solution is meaningless.

Appendix 3: Multiple equilibria?

To date, I have presented this paper a number of times and I have usually asked about multiple equilibria. While this could refer to several different things, I believe the question refers to whether or not the solution could be "stuck" with an outcome in which inefficient firms types are active and block entry by the most efficient types. I am quite familiar with this question and examine it in detail in Horstmann and Markusen (1992) and Markusen (2002), with the latter in particular having a chapter on issues such as first-move advantage and blockaded entry. These issues have reappeared more recently in Atkeson and Burstein (2008), Eaton, Kortum and Sotelo (2013), and Gaubert and Itskhoki (2021) who rule this out by assuming sequential entry, with the most efficient firm type allowed to enter first. This type of multiple equilibria doesn't occur in my model because a "fractional firm" (size between zero and one) is permitted: *N(i)* is continuous between bounds and is not constrained to integer values as it is in the four works just referenced. Note the model formulation could be call a simultaneous-move game.

Suppose that we constrain the most efficient firm type 1 at 0 (*N.fx(1)* = 0 in GAMS notation). If we solve the model, no matter how many other less-efficient types can enter, the solution will show that (potential) markup revenues will exceed fixed costs for a small fractional firm of type 1 evaluated at *N(1)* = 0. If we then free up type 1, then the next Newton iteration sets *N(1)* at a positive value and the outputs and profits of the other firm types shrink. The iteration process will converge to the unrestricted equilibrium in which type 1 is a "full sized" firm and likely the least efficient firm(s) that was active in the constrained equilibrium exits. In other words, I am arguing that my model will arrive at the same solution independently of starting values given that the number of firms of a type is a continuous variable between zero and one.

But could we add some more restrictive assumptions that could produce multiple equilibria? This is possible as explained in Tables A3.1 and A3.2. Here are a set of assumptions that can do that, with no claim that these are necessary conditions.

(1) Depart from the simultaneous moves assumption and assume a two-stage entry procedure with less efficient types 2-5 having a first-mover advantage. Constrain the most efficient type to zero and solve.

(2) Hold the solution values *N(i)* for types 2-5 fixed, allow type 1 to enter and resolve. As I just noted, the solution must give positive profits for a small fractional type 1 firm to enter, which sets off an adjustment process. But can we can add additional restrictions such that the first-mover solution blockades entry by the most efficient type 1? The answer seem is yes.

(3) The economy must be relatively small.

(4) Firm number are constrained to integer values, 0 or 1, no fractional firms.

(5) Fixed costs for the firms entering initially at the first-mover solution are sunk, so they remain in the market if type 1 enters in the second stage assuming that their prices exceed marginal costs. Perhaps this also requires that types 2-5 have myopic views as to the fact that type 1 can enter in the second phase.

Two cases for two different sized economies are shown in Table A3.2. Scenario S1 solves the model with no first-mover advantage, and then scenario S2 solves the model with type 1 block (N(1) = 0). In S1, only types 1 and 2 can enter, while in S2 types 2 and 3 enter. Now hold the solution values *N* for types 2-5 fixed as the scenario 2 solution. Scenarios S3-S5 successively allow for a fractional type 1 firm with upper size limit *N(1)* = 0.01, 0.25, 0.50. In all these cases, type 1 make positive profits and so will not be blockaded as a fractional firm.

Scenario S6 frees up type 1 with its upper bound equal to 1, but holds the number of the less efficient firms at their S2 values. For the small economy case, the solution value for type 1 is *N(1)* = 0.6030 with all the other types making negative profits with out the sunk cost assumption (type 2 makes loses since types 4 and 5 are still in the market). Scenario S7 constrains type 1 to enter only as a full-sized firm *N(1)* = 1. Firm type 1 now makes losses with the other types at their first-mover advantage sizes (cell with -3.09 outlined). This is, I think what several readers / listeners have in mind when they think of multiple equilibria. But note that this requires quite restrictive assumptions, including the myopic assumption that less efficient types don't anticipate type 1 entering in the second stage.

With respect to the assumption requiring the economy to be small for this to occur, the second set of numbers increases the economy size by only 25 percent. In this case, firm 1 can/will enter as a full-sized firm even holding the numbers of the other firms fixed at their S2 levels. Entry cannot be blockaded.

Table A3.1 Can there be multiple equilibria?

(1) the way that the model is constructed and solved I believe not

(2) key: N is not an integer variable, it is continuous variable between 0 and N.up
frational (non-integer firm numbers) permitted as started with Dixit-Stiglitz, Krugman

(3) solve the model with type 1 fixed as zero: N1.fx = 0

(4) results always show positive profits from type 1 entering as a small fractional firm

(5) free up N1 to N1.up = 1 at the initial solution

(5) next Newton iteration sets N1 > 0, less efficient types reduce output, least efficient may exit
solution algorithm will continue until N1 hits its upper bound N1.up
model is not sensitive to starting values, but no algorithm is guarenteed to solve

Can we come up with a scenario/assumptions that produce multiple equilibria?

Experiment - quasi two-stage entry procedure: less efficient types 2-5 have first mover advantage
compute solution with type 1 blocked, N1.fx = 0
hold the N2-N5 solution values fixed and then allow type 1 to enter
BUT then we are no longer in a simultaneous-move Nash game

Conditions which may produce multiple equilibria - Firm type 1 is blockaded by moving second

(1) economy must be relatively "small"

(2) firm numbers N(i) are constrained to integer values, 0, 1, no fractional firms

(3) fixed costs for the firms entering initially at N1 = 0 are sunk
so remain in the market even if type 1 enters (markups exceed marginal costs)

=> in a small market, a "full sized" N1 = 1 firm may make losses by entering
and so is "blockaded" which we could term a (second) equilibria
but all other firm types would also make losses if fixed costs are not sunk

=> if permitted, a small fractional firm type N1 can enter and make positive profits
will enter even if conjecturing other firm numbers as fixed
least efficient firm type will likely make losses

Table A3.2: Multiple equilibria by blockaded entry or not?

Scenarios

S1	No restrictions on entry up to a type's upper bound N.up	
S2	Type 1 blocked N1.up = 0	
S3	Type 1 constrained to N1.up = 0.01	Types 2-5 fixed at scenario S2 values
S4	Type 1 constrained to N1.up = 0. 25	Types 2-5 fixed at scenario S2 values
S5	Type 1 constrained to N1.up = 0. 50	Types 2-5 fixed at scenario S2 values
S6	Type 1 N.up = 1, solultion value = 0.603	Types 2-5 fixed at scenario S2 values
S7	Type 1 fixed at N1.fx = 1	Types 2-5 fixed at scenario S2 values

Small Economy — Profits of firm types in columns

		T1	T2	T3	T4	T5	
N.up all types = 1	S1	4.19	0.34	-2.06	-7.56	-8.20	
N1.up = 0	S2		3.64	0.85	-5.63	-6.39	firm numbers
N1.up = 0.01	S3	0.08	3.48	0.71	-5.72	-6.48	N2, N3, N4, N5
N1.up = 0.25	S4	1.01	0.20	-2.18	-7.64	-8.28	held fixed at
N1.up = 0.50	S5	1.21	-1.15	-3.36	-8.42	-9.01	scentario S2 levels
N1.up = 1.0=>N1 = 0.603	S6	0.00	-3.17	-5.12	-9.57	-10.09	
N1.fx = 1.0	S7	$\boxed{-3.09}$	-5.73	-7.35	-11.01	-11.44	

Small Economy — Numbers of active firms of each type in columns

		T1	T2	T3	T4	T5	
N.up all types = 1	S1	1	1	0	0	0	
N1.up = 0	S2	0	1	1	0	0	firm numbers
N1.up = 0.01	S3	0.01	1	1	0	0	N2, N3, N4, N5
N1.up = 0.25	S4	0.25	1	1	0	0	held fixed at
N1.up = 0.50	S5	0.50	1	1	0	0	scentario S2 levels
N1.up = 1.0 = >N1 = 0.60	S6	0.60	1	1	0	0	
N1.fx = 1.0	S7	1	1	1	0	0	

Economy 25% larger — Profits of firm types in columns

		T1	T2	T3	T4	T5	
N.up all types = 1	S1	7.17	2.74		-6.26	-6.99	
N1.up = 0	S2		9.45	5.89	-2.37	-3.34	firm numbers
N1.up = 0.01	S3	0.15	9.24	5.71	-2.49	-3.45	N2, N3, N4, N5
N1.up = 0.25	S4	2.48	5.06	2.03	-4.93	-5.74	held fixed at
N1.up = 0.50	S5	3.94	3.34	0.52	-5.92	-6.67	scentario S2 levels
N1.up = 1.0=>N1 = 1.0	S6	0.91	-2.47	-4.54	-9.22	-9.76	
N1.fx = 1.0	S7	$\boxed{0.91}$	-2.47	-4.54	-9.22	-9.76	

```
$TITLE  SGB-add mnes.gms  adds horizontal mne option. Figures 6-7  James R. Markusen
* two country (h and f) trade model, small group Bertrand competition
* no comparative advanatage, one factor labor, iceberg trade costs
* X industry - increasing returns, imperfect competition
* Y industry - constant returns, perfect competition, no trade cost.
* Four entry modes: no entry, domestic, exporting horizontal mne

SETS I firm types differ by marginal costs /I1*I5/;
ALIAS (I,II);

PARAMETERS
 ENDOWH, ENDOWF        Endowment scale multiplier
 MC(I)                 marginal cost for firm types same across countries
 TC                    trade cost gross basis (1 + trade cost rate)
 SIG                   elasticity of substitution among X goods
 FCD, FCX, FCM         fixed cost for domestic exporting and mne firms;

ENDOWH = 1000; ENDOWF = 1000;
TC = 1.0001;
SIG = 6 + 1/3;

FCD = 8;
*FCX = 13;
FCX = 14;
FCM = 17;
*FCM = 30;

MC("I1") = 1;
MC("I2") = 1.1;
MC("I3") = 1.13;
MC("I4") = 1.135;
MC("I5") = 1.14;


NONNEGATIVE VARIABLES

  XHH(I)      Production by an h firm of type i for sale in h
  XHF(I)      Production by an h firm of type i for export to f
  XMHF(I)     Production by an h firm of type i in a plant in f
  XFF(I)      Production by an f firm of type i for sale in f
  XFH(I)      Production by an f firm of type i for export to h
  XMFH(I)     Production by an f firm of type i in a plant in h

  NHD(I)      Number of X sector firms of type i in h
  NHX(I)      Number of X sector firms of type i in h exporting
  NHM(I)      Number of X sector firms of type i in h + plant in f
  NFD(I)      Number of X sector firms of type i in f
  NFX(I)      Number of X sector firms of type i in f exporting
  NFM(I)      Number of X sector firms of type i in f + plant in f

  YH          Level of Y output in country h
  YF          Level of Y output in country f
  WH          Welfare of h
  WF          Welfare of f

  PXHH(I)     Price of good Xh sold in h
  PXHF(I)     Price of good Xh sold in f
  PXMHF(I)    Price of good of an h firm produced and sold in f
  PXFF(I)     Price of good Xf sold in f
  PXFH(I)     Price of good Xf sold in f
  PXMFH(I)    Price of good of an f firm produced and sold in h

  PY          World price of Y
  PWH         Price index of utility in country h
  PWF         Price index of utility in country f
  PLH         Price of labor in country h
  PLF         Price of labor in country f
```

```
  CONSH         Income of the representative consumer in country h
  CONSF         Income of the representative consumer in country f
  ENTRHD(I)     Income of the agent ENTRE for firm type i in h
  ENTRHX(I)     Income of the agent ENTRE for firm type i in f from exporting
  ENTRHM(I)     Income of the agent ENTRE for firm type i from foreign production
  ENTRFD(I)     Income of the agent ENTRE for firm type i in h
  ENTRFX(I)     Income of the agent ENTRE for firm type i in f from exporting
  ENTRFM(I)     Income of the agent ENTRE for firm type i from foreign production

  PEH           Price index for X composite in h
  PEF           Price index for X composite in f

  MARKHH(I)     Markup of a type i h firm for sale in h
  MARKHF(I)     Markup of a type i h firm for sale in f
  MARKMHF(I)    Markup of a type i h firm producing in f
  MARKFF(I)     Markup of a type i f firm for sale in f
  MARKFH(I)     Markup of a type i f firm for sale in h
  MARKMFH(I)    Markup of a type i f firm producing in h

  LAMH(I)       Shadow tax to implement entry constraint on home firms
  LAMF(I)       Shadow tax to implement entry constraint on foreign firms;

EQUATIONS

  PRXHH(I)      Pricing inequality for XHH
  PRXHF(I)      Pricing inequality for XHF
  PRXMHF(I)     Pricing inequality for XMHF
  PRXFF(I)      Pricing inequality for XFF
  PRXFH(I)      Pricing inequality for XFH
  PRXMFH(I)     Pricing inequality of XMFH

  PRNHD(I)      Pricing inequality for NHD
  PRNHX(I)      Pricing inequality for NHX
  PRNHM(I)      Pricing inequality for NHM
  PRNFD(I)      Pricing inequality for NFD
  PRNFX(I)      Pricing inequality for NFX
  PRNFM(I)      Pricing inequality for NFM

  PRICEYH       Pricing inequality for YH (PY = MC)
  PRICEYF       Pricing inequality for YF
  PRICEWH       Consumer price index for country h
  PRICEWF       Consumer price index for country f

  MKTXHH(I)     Supply >= demand for XHH
  MKTXHF(I)     Supply >= demand for XHF
  MKTXMHF(I)    Supply >= demand for XMHF
  MKTXFF(I)     Supply >= demand for XFF
  MKTXFH(I)     Supply >= demand for XFH
  MKTXMFH(I)    Supply >= demand for XMFH

  MKTFY         Export supply = import demand for Y
  MKTWH         Supply-demand for WH
  MKTWF         Supply-demand for WF
  MKTLH         Supply-demand balance for labor LH
  MKTLF         Supply-demand balance for labor LF

  ICONSH        Consumer income in h including profits of Xh firms
  ICONSF        Consumer income in f including profits of Xf firms
  IENTRHD(I)    Entrepreneur's profits (markup revenues) in h
  IENTRHX(I)    Entrepreneur's profits (markup revenues) in h on exports
  IENTRHM(I)    Entrepreneur's profits in h on foreign production in f
  IENTRFD(I)    Entrepreneur's profits (markup revenues) in h
  IENTRFX(I)    Entrepreneur's profits (markup revenues) in h on exports
  IENTRFM(I)    Entrepreneur's profits i nf on foreign production in h

  PINDEXH       Price index for X goods in h
  PINDEXF       Price index for X goods in f
```

```
  MKHH(I)     Markup inequality for XHH(I)
  MKHF(I)     Markup inequality for XHF(I)
  MKMHF(I)    Markup inequality for XMHF(I)
  MKFF(I)     Markup inequality for XFF(I)
  MKFH(I)     Markup inequality for XFH(I)
  MKMFH(I)    Markup inequality for XMFH(I)

  ELAMH(I)    Constraint to limit NHD + NHX number
  ELAMF(I)    Constraint to limit NFD + NFX number;


PRXHH(I)..       PLH*MC(I)    =G= PXHH(I)*(1 - MARKHH(I));
PRXHF(I)..       PLH*MC(I)*TC =G= PXHF(I)*(1 - MARKHF(I));
PRXMHF(I)..      PLF*MC(I)    =G= PXMHF(I)*(1 - MARKMHF(I));
PRXFF(I)..       PLF*MC(I)    =G= PXFF(I)*(1 - MARKFF(I));
PRXFH(I)..       PLF*MC(I)*TC =G= PXFH(I)*(1 - MARKFH(I));
PRXMFH(I)..      PLH*MC(I)    =G= PXMFH(I)*(1 - MARKMFH(I));

PRNHD(I)..       FCD*PLH + LAMH(I)*PLH=G= ENTRHD(I);
PRNHX(I)..       FCX*PLH + LAMH(I)*PLH=G= ENTRHX(I);
PRNHM(I)..       FCM*PLH + LAMH(I)*PLH=G= ENTRHM(I);
PRNFD(I)..       FCD*PLF + LAMF(I)*PLF=G= ENTRFD(I);
PRNFX(I)..       FCX*PLF + LAMF(I)*PLF=G= ENTRFX(I);
PRNFM(I)..       FCM*PLF + LAMF(I)*PLF=G= ENTRFM(I);

PRICEYH..        PLH =G= PY;
PRICEYF..        PLF =G= PY;
PRICEWH..        ((PEH)**0.5)*(PY**0.5) =G= PWH;
PRICEWF..        ((PEF)**0.5)*(PY**0.5) =G= PWF;

MKTXHH(I)..      XHH(I) =G=  PXHH(I)**(-SIG)*(PEH**(SIG-1))*0.5*CONSH;
MKTXHF(I)..      XHF(I)/TC =G=  PXHF(I)**(-SIG)*(PEF**(SIG-1))*0.5*CONSF;
MKTXMHF(I)..     XMHF(I) =G=  PXMHF(I)**(-SIG)*(PEF**(SIG-1))*0.5*CONSF;
MKTXFF(I)..      XFF(I) =G= PXFF(I)**(-SIG)*(PEF**(SIG-1))*0.5*CONSF;
MKTXFH(I)..      XFH(I)/TC =G=  PXFH(I)**(-SIG)*(PEH**(SIG-1))*0.5*CONSH;
MKTXMFH(I)..     XMFH(I) =G=  PXMFH(I)**(-SIG)*(PEH**(SIG-1))*0.5*CONSH;

MKTFY..          YH + YF =G= 0.5*CONSH/PY + 0.5*CONSF/PY;
MKTWH..          PWH*WH =G= CONSH;
MKTWF..          PWF*WF =G= CONSF;
MKTLH..          ENDOWH =G= YH + SUM(I,  ((NHD(I)+NHX(I)+NHM(I))*XHH(I)+NHX(I)*XHF(I)+NHM(I)*XMHF(I))*MC(I))
                                 + SUM(I, NHD(I)*FCD + NHX(I)*FCX + NHM(I)*FCM);
MKTLF..          ENDOWF =G= YF + SUM(I,  ((NFD(I)+NFX(I)+NFM(I))*XFF(I)+NFX(I)*XFH(I)+NFM(I)*XMFH(I))*MC(I))
                                 + SUM(I, NFD(I)*FCD + NFX(I)*FCX + NFM(I)*FCM);

ICONSH..         CONSH  =G= PLH*ENDOWH + SUM(I,  (NHD(I)+NHX(I)+NHM(I))*(PLH*LAMH(I)));

ICONSF..         CONSF  =G= PLF*ENDOWF + SUM(I,  (NFD(I)+NFX(I)+NFM(I))*(PLF*LAMF(I)));


IENTRHD(I)..     ENTRHD(I) =G= MARKHH(I)*PXHH(I)*XHH(I);
IENTRHX(I)..     ENTRHX(I) =G= MARKHH(I)*PXHH(I)*XHH(I)  + MARKHF(I)*PXHF(I)*XHF(I)/TC;
IENTRHM(I)..     ENTRHM(I) =G= MARKHH(I)*PXHH(I)*XHH(I)  + MARKMHF(I)*PXMHF(I)*XMHF(I);
IENTRFD(I)..     ENTRFD(I) =G= MARKFF(I)*PXFF(I)*XFF(I);
IENTRFX(I)..     ENTRFX(I) =G= MARKFF(I)*PXFF(I)*XFF(I)  + MARKFH(I)*PXFH(I)*XFH(I)/TC;
IENTRFM(I)..     ENTRFM(I) =G= MARKFF(I)*PXFF(I)*XFF(I)  + MARKMFH(I)*PXMFH(I)*XMFH(I);

PINDEXH..        PEH =E= (SUM(I, (NHD(I)+NHX(I)+NHM(I))*PXHH(I)**(1-SIG)
                             + NFX(I)*PXFH(I)**(1-SIG)  + NFM(I)*PXMFH(I)**(1-SIG)))**(1/(1-SIG));
PINDEXF..        PEF =E= (SUM(I, (NFD(I)+NFX(I)+NFM(I))*PXFF(I)**(1-SIG)
                             + NHX(I)*PXHF(I)**(1-SIG)  + NHM(I)*PXMHF(I)**(1-SIG)))**(1/(1-SIG));
```

```
MKHH(I)..        MARKHH(I) =G= 1 / (SIG - (SIG-1)*PXHH(I)*XHH(I)/
                     (SUM(II, (NHD(II)+NHX(II)+NHM(II))*PXHH(II)*XHH(II) + NFX(II)*PXFH(II)*XFH(II)/TC
                      + NFM(II)*PXMFH(II)*XMFH(II))));

MKHF(I)..        MARKHF(I) =G= 1 / (SIG - (SIG-1)*PXHF(I)*XHF(I)/TC/
                     (SUM(II, (NFD(II)+NFX(II)+NFM(II))*PXFF(II)*XFF(II) + NHX(II)*PXHF(II)*XHF(II)/TC
                      + NHM(II)*PXMHF(II)*XMHF(II))));

MKMHF(I)..       MARKMHF(I) =G= 1 / (SIG - (SIG-1)*PXMHF(I)*XMHF(I)/
                     (SUM(II, (NFD(II)+NFX(II)+NFM(II))*PXFF(II)*XFF(II) + NHX(II)*PXHF(II)*XHF(II)/TC
                      + NHM(II)*PXMHF(II)*XMHF(II))));

MKFF(I)..        MARKFF(I) =G= 1 / (SIG - (SIG-1)*PXFF(I)*XFF(I)/
                     (SUM(II, (NFD(II)+NFX(II)+NFM(II))*PXFF(II)*XFF(II) + NHX(II)*PXHF(II)*XHF(II)/TC
                      + NHM(II)*PXMHF(II)*XMHF(II))));

MKFH(I)..        MARKFH(I) =G= 1 / (SIG - (SIG-1)*PXFH(I)*XFH(I)/TC/
                     (SUM(II, (NHD(II)+NHX(II)+NHM(II))*PXHH(II)*XHH(II) + NFX(II)*PXFH(II)*XFH(II)/TC
                      + NFM(II)*PXMFH(II)*XMFH(II))));

MKMFH(I)..       MARKMFH(I) =G= 1 / (SIG - (SIG-1)*PXMFH(I)*XMFH(I)/
                     (SUM(II, (NHD(II)+NHX(II)+NHM(II))*PXHH(II)*XHH(II) + NFX(II)*PXFH(II)*XFH(II)/TC
                      + NFM(II)*PXMFH(II)*XMFH(II))));

ELAMH(I)..        1 =G= NHD(I) + NHX(I) + NHM(I);
ELAMF(I)..        1 =G= NFD(I) + NFX(I) + NFM(I);

MODEL M52 /PRXHH.XHH, PRXHF.XHF, PRXMHF.XMHF, PRXFF.XFF, PRXFH.XFH, PRXMFH.XMFH,
        PRICEYH.YH, PRICEYF.YF, PRICEWH.WH, PRICEWF.WF,
        MKTXHH.PXHH, MKTXHF.PXHF, MKTXMHF.PXMHF, MKTXFF.PXFF, MKTXFH.PXFH, MKTXMFH.PXMFH,
        PINDEXH.PEH, PINDEXF.PEF,
        MKTWH.PWH, MKTWF.PWF, MKTFY.PY,
        MKTLH.PLH, MKTLF.PLF,
        ICONSH.CONSH, ICONSF.CONSF,
        PRNHD.NHD, PRNHX.NHX, PRNHM.NHM, PRNFD.NFD, PRNFX.NFX, PRNFM.NFM,
        IENTRHD.ENTRHD, IENTRHX.ENTRHX, IENTRHM.ENTRHM, IENTRFD.ENTRFD, IENTRFX.ENTRFX, IENTRFM.ENTRFM,
        MKHH.MARKHH, MKHF.MARKHF, MKMHF.MARKMHF, MKFF.MARKFF, MKFH.MARKFH, MKMFH.MARKMFH,
        ELAMH.LAMH, ELAMF.LAMF/;

* set initial values of variables for solver

CONSH.L = 800; CONSF.L = 800;
XHH.L(I) = 20; XFF.L(I) = 20; XHF.L(I) = 20; XFH.L(I) = 20; XMHF.L(I) = 20; XMFH.L(I) = 20;
YH.L = 100; YF.L = 100; WH.L = 200; WF.L = 200;
NHD.L(I) = 0; NHX.L(I) = 0; NFD.L(I) = 0; NFX.L(I) = 0;
NHD.L("I1") = 4; NHX.L("I1") = 4; NFD.L("I1") = 4; NFX.L("I1") = 4;

PXHH.L(I) = 1.25; PXHF.L(I) = 1.25; PXFF.L(I) = 1.25; PXFH.L(I) = 1.25;
PXMHF.L(I) = 1; PXMFH.L(I) = 1;
PEH.L = 0.8464; PEF.L = 0.8464;
PLH.L = 1; PLF.L = 1; PWH.L = 1; PWF.L = 1; PY.L = 1;
ENTRHD.L(I) = 10; ENTRHX.L(I) = 10; ENTRFD.L(I) = 10; ENTRFX.L(I) = 10;
MARKHH.L(I) = 0.20; MARKHF.L(I) = 0.20; MARKFF.L(I) = 0.20; MARKFH.L(I) = 0.20;

* choose PYH as numeraire, and check calibration
PY.FX = 1;
TC = 1.001;

M52.ITERLIM = 0;
SOLVE M52 USING MCP;

M52.ITERLIM = 1000;
SOLVE M52 USING MCP;

TC = 1.3;
M52.ITERLIM = 1000;
SOLVE M52 USING MCP;
```

```
SETS j indexes 25 different trade cost levels  /J1*J25/;

PARAMETERS
 TCOST(J)
 FIRMNUMB(J,I)
 MARKUPO(J,I)
 RESULTS1(J, *), RESULTS1ad(J,I), RESULTS1ax(J,I), RESULTS1am(J,I),
 RESULTS1b(J,I), RESULTS1c(J,I), RESULTS1d(J,I),
 RESULTS1e(J, *), RESULTS1f(J,I), RESULTS1g(J,I), RESULTS1h(J,I);

LOOP(J,

* loop over trade costs from autarky 1.275 to free trade 1.0001

TCOST(J) =  1.2 - 0.008333*(ORD(J) - 1);
TCOST("J25") = 1.0001;
TC = TCOST(J);

ENDOWH = 600; ENDOWF = 600;

SOLVE M52 USING MCP;

RESULTS1(J, "TCOST") = TCOST(J);
RESULTS1(J, "WELFCDAP") = WH.L;
RESULTS1ad(J, I) = NHD.L(I) + EPS;
RESULTS1ax(J, I) = NHX.L(I) + EPS;
RESULTS1am(J, I) = NHM.L(I) + EPS;
RESULTS1b(J, I) = XHH.L(I)$(NHD.L(I)+NHX.L(I) GT 0);
RESULTS1c(J, I) = (XHF.L(I))$(NHX.L(I) GT 0);
RESULTS1d(J, I) = MARKHH.L(I)$(NHD.L(I)+NHX.L(I) GT 0);
RESULTS1e(J, I) = MARKHF.L(I)$(NHX.L(I) GT 0);
RESULTS1f(J, I) = (-ENTRHD.L(I) + PLH.L*FCD + PLH.L*LAMH.L(I))*NHD.L(I);
RESULTS1g(J, I) = (-ENTRHX.L(I) + PLH.L*FCX + PLH.L*LAMH.L(I))*NHX.L(I);

RESULTS1H(J, I) = (ENTRHX.L(I)*NHX.L(I) + ENTRHD.L(I)*NHD.L(I)
                   - PLH.L*FCX*NHX.L(I) - PLH.L*FCD*NHD.L(I))/CONSH.L;
);

DISPLAY RESULTS1, RESULTS1ad, RESULTS1ax, RESULTS1am, RESULTS1b, RESULTS1c, RESULTS1d, RESULTS1e,
RESULTS1f, RESULTS1g, RESULTS1h;

Execute_Unload 'RESULTS-MNE-2b.gdx' RESULTS1
execute 'gdxxrw.exe RESULTS-MNE-2b.gdx par=RESULTS1 rng=SHEET1!A3:C28'

Execute_Unload 'RESULTS-MNE-2b.gdx' RESULTS1ad
execute 'gdxxrw.exe RESULTS-MNE-2b.gdx par=RESULTS1ad rng=SHEET1!E3:J28'

Execute_Unload 'RESULTS-MNE-2b.gdx' RESULTS1ax
execute 'gdxxrw.exe RESULTS-MNE-2b.gdx par=RESULTS1ax rng=SHEET1!L3:Q28'

Execute_Unload 'RESULTS-MNE-2b.gdx' RESULTS1am
execute 'gdxxrw.exe RESULTS-MNE-2b.gdx par=RESULTS1am rng=SHEET1!S3:X28'

Execute_Unload 'RESULTS MNE 2b.gdx' RESULTS1b
execute 'gdxxrw.exe RESULTS MNE 2b.gdx par=RESULTS1b rng=SHEET3!E3:J28'

Execute_Unload 'RESULTS MNE 2b.gdx' RESULTS1c
execute 'gdxxrw.exe RESULTS MNE 2b.gdx par=RESULTS1c rng=SHEET3!L3:Q28'

Execute_Unload 'RESULTS MNE 2b.gdx' RESULTS1d
execute 'gdxxrw.exe RESULTS MNE 2b.gdx par=RESULTS1d rng=SHEET3!S3:X28'

Execute_Unload 'RESULTS MNE 2b.gdx' RESULTS1e
execute 'gdxxrw.exe RESULTS MNE 2b.gdx par=RESULTS1e rng=SHEET3!Z3:AE28'

Execute_Unload 'RESULTS MNE 2b.gdx' RESULTS1h
execute 'gdxxrw.exe RESULTS MNE 2b.gdx par=RESULTS1h rng=SHEET3!AG3:AL28'
```