Exploiting complementarity in applied general-equilibrium models

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

Abstract

Both analytical general-equilibrium models and applied applications generally make compromises to avoid endogenous zeros and discrete regime shifts. In open economy models, assumptions such as national-level product differentiation, a continuum of firms, probabilistic production, and smooth parametric distribution functions ensure interior closed-form solutions to systems of equations. But resulting complexities restrict models to sparse general-equilibrium structures. I offer an alternative by exploiting non-linear complementarity, focusing on heterogeneous firms, endogenous exporting and multinational production. There is a discrete set of firm types allowing for any differences in productivity which can be calibrated from observed market shares. The model incorporates large firms with endogenous markups and solves for the set of active firms and their supply modes (no entry, domestic, exporting, multinational). While focusing on international trade, I believe that complementarity has applicability to many fields of economics where endogenous boundary solutions are important to the narrative.

May 2026

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, Julia, R and Matlab; details 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

Keywords: Complementarity, Karush-Kuhn-Tucker, Endogenizing zeros, Applied general-equilibrium models

JEL codes: C63, C68, F12, F23

1. Introduction

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. This has occurred in both analytical models where dealing with inequalities is awkward past very small-dimension problems and in formulating numerical simulation models. In both cases, assumptions are made such that a model consists of a set of equations, preferably with close-form solutions for analytical versions. For counter-factual simulations in numerical models or comparative statics in analytical models, assumptions ensure that positive benchmark values stayed positive or initial zeros are constrained to remained zero.

An early tactic for ruling out switching between active and inactive activities in numerical models is referred to as the Armington assumption in trade applications. 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). Micro-foundations for this approach are weak, but it has remained a workhorse tool to the present. Restrictions ensuring no endogenous zeros persist today in the popular and highly cited Eaton and Kortum (2002) model. This approach relies on probabilistic production for analytical solutions, with all countries producing and exporting at least some small amount in all sectors. There can be no changes in the extensive margin of trade, and indeed under some parameterizations it is equivalent to Armington. 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 equally important contribution is that of Melitz (2003), who tackled the problem of heterogenous firms and endogenous choices of entry and the decision to export. This is also an analytical model, relying on atomistic firms, a smooth continuous distribution of firm productivities, and random productivities draws. It is a major contribution but has difficulties dealing with large firms and discrete jumps in variables such as firm numbers or asymmetries among firms in the continuum (e.g., variable versus fixed cost differences in technologies).

I adopt a deterministic formulation as a simple alternative to what I term the probabilistic-continuum approaches. Instead, I formulate general equilibrium as a complementarity problem. This has been relatively rarely exploited in general-equilibrium analytical models. It has been used in perfect-competition applied general-equilibrium models (AGE) however. This was first developed 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.

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. The models developed do not have analytical solutions, and I believe that this explains their relative scarcity of applications. At least in international economics, there is a high preference for analytical models with closed-form solutions, but their direct translation into numerical simulations for counter-factuals favors the probabilistic-continuum approaches restricted to interior solutions. Here, I develop analytical models which are then solved and simulated numerically.

My view is that analytical tools based on probabilistic production are complex and difficult to incorporate into high-dimension applied general-equilibrium models. Their resulting simple structures cannot address some current policy issues such as trade diversion (China to Vietnam) or foreign firms switching from exporting to building foreign plants (Taiwanese and Korean semi-conductor firms) . In addition to ruling out regime, location, and mode switches, a number of other restrictive assumptions are typically (but not universally) made.

First, much of the literature continues to use "large-group" monopolistic competition (MC) 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 and an inability to deal with large market-share firms. 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 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 simultaneity in general equilibrium, allowing total income to be independent of profits.

The alternative way to model heterogeneous firms offered here avoids 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., sales) into a discrete set of firm "types". Firm types differ in their marginal cost. There is free entry into firm types, but only up to a limited number for each type. This limit per types is the discrete equivalent of the continuous parametric distribution used in the literature which trivially allows only one firm at each point in the continuum. The model will solve for the set of active firm types depending on parameters such as the economy's size or trade costs. Positive aggregate profits are added to total income. The entry limit could result from scarce entrepreneurs, patents, or other types of scarce firm-specific intangible assets that prevent entrants from capturing the productivity of top firms. Entrants cannot just duplicate Nividia's or TSCM's technology. Bernard, Dhyne, Magerman, Manova, and Moxnes (2022) identify additional sources of heterogeneity.

---

[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), but it is not formulated as a complementarity problem.

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 are needed. Second, there is no need to impose any parametric distribution function on firm productivity or cost. These can be calculated directly from market-share data once those are divided into firm classes or type. Third, there is no need to assume MC. Nash Cournot (NC) or Nash Bertrand (NB) can allow for large firms and 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 GAMS. Then I specify a basic general-equilibrium model: a two-sector, one-factor closed-economy model. Paying for productivity draws is unnecessary and the aggregate profits of the active firms are added to the income of the representative consumer. The specific formulation for simulation has five firm types in one increasing-returns sector, free entry up to an upper bound on each firm type. Two simple experiments are introduced to exploit the model properties, comparing results under NC, NB and MC. These sections also show how unobserved marginal costs or productivity can be derived from observed market shares and provide an illustrative application to semi-conductor market-share data.

The second model extends the first to a two-country trade casel, 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 NB to MC (the Melitz original). 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 multinational models of Horstmann and Markusen (1992) and Markusen (2002).[2]

The second and third models yield simulation results that resemble those in Melitz 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 serve only their domestic market and the least productive firms may not enter. The simulations show how the configuration of modes across firm types changes as trade costs fall from a high level to costless trade. Several cases of country or firm asymmetries are included to argue that complementarity allows for the incorporation of many real-world features found in data and needed for policy analysis.

[2]The non-linear complementarity approach 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. 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.

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 a solution to $F(Z) = 0$ violates one (or more) of the inequality constraints. 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 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. 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 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. The inequality cannot run the other way, price greater than marginal cost leaves an unexploited profit opportunity.

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

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 unprofitable under current economic conditions ((2) refers to a specific one of those technologies). Second, the world trade matrix is full of zeros: unprofitable trade links.

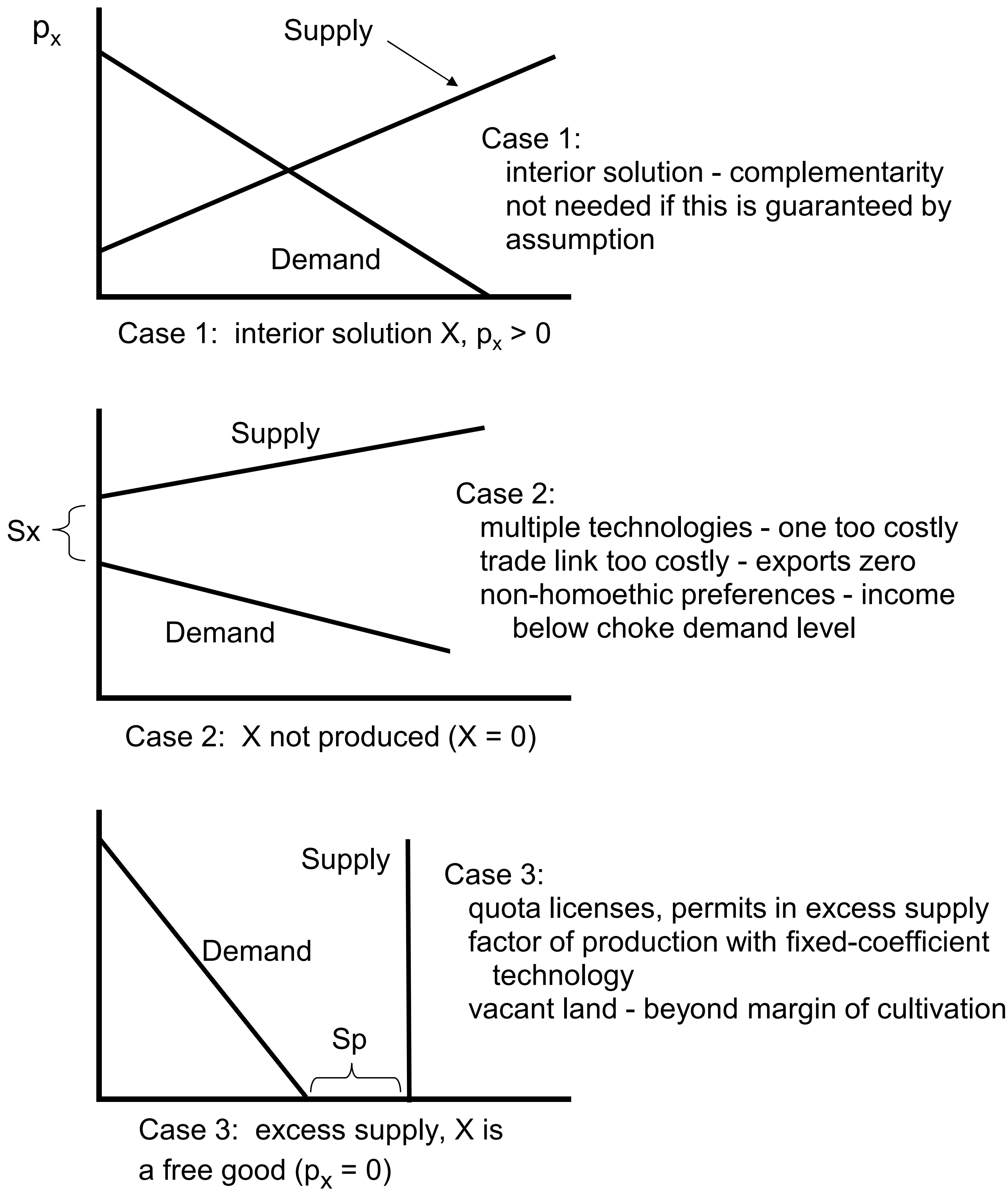
The other type of situation that requires a complementarity condition due to a non-negativity constraint is a market clearing equation, supply equal demand, $D(p_x)$. But it may be that the solution involves excess supply in equilibrium, implying that the price is zero.

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

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 supply of licenses or quota level), or a factor of production in excess supply with fixed-coefficient (Leontief) technologies.

Figure 1gives the three possible outcomes of this simple partial-equilibrium, supply-

Figure 1: Three equilibrium outcomes



demand problem. The top panel is the classic undergraduate textbook case of an interior solution. If guaranteed, complementarity is not needed: 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 introduce non-negative "slack" variables into the problem, which 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, the complementarity conditions.

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

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

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

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

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: the modeler just specifies the weak inequalities and matched variables. 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.[3]

3. Complementarity and KKT with bounded variables

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). 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, the variable associated with the weak inequality is the number of firms in equilibrium. If fixed costs are greater than (potential) markup revenues, then the number of firms is zero.

---

[3]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).

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. For most of the paper, the upper bound will be *N.up(i)* = 1. The solution will converge to a single firm type (the most productive) as *N.up(i)* becomes large, the classic MC Krugman outcome. There are three solutions in general equilibrium, ignoring knife-edged solutions.

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

$$FC(i) > MKR(i) \quad \Rightarrow \quad N(i) = N.lo(i) = 0 \quad \text{losses, no entry} \tag{9}$$

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

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.

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

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

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

Equations (12) and (13) are the complementary conditions. Both *w(i)* and *v(i)* are zero in an interior (zero profit) solution. As just noted, 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. Markups, entry and profit income

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.. *N(i)* is the number of active type *i* firms, σ is the elasticity of substitution among the *X* goods. Welfare (*W*) of the representative consumer, and the symmetry of varieties of the *X* goods gives welfare 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 $\eta(i)$, 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. These are MC, NC, and NB. 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 MC, the perceive elasticity of firm demand is just $\eta = \sigma$ with the markup as defined here $mk = 1/\sigma$. The markups for Bertrand and Cournot are

$$mk_b(i) = \frac{1}{\sigma - s(i)(\sigma - 1)} \qquad mk_c(i) = s(i) + (1 - s(i))\frac{1}{\sigma} \tag{16}$$

In Markusen (2023), I show that 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. The limiting values of NB and NC markups converge to $1/\sigma$ as a firm’s market share goes to zero.

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 firms 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 Bertrand as the example but show results for all three cases. I will provide the code for all three versions on request.

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

$$mk(i) \geq \frac{1}{\sigma - s(i)(\sigma - 1)} \qquad \perp \quad mk(i) \geq 0 \tag{17}$$

$$ENTRE(i) \geq mk(i)p_x(i)X(i) \qquad \perp \quad ENTRE(i) \geq 0 \tag{18}$$

$$ENTRE(i) \geq p_l fc \geq ENTRE(i) \qquad \perp \quad N.up(i) \geq N(i) \geq 0 \tag{19}$$

where p_l is the price of labor, the single factor of production in the model to follow, and fixed cost 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 (19) as described in the previous section. The Bertrand markups formula is the same as Gaubert and Itskhoki (2021).

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

$$CONS \geq p_l ENDOW + \sum_i [N(i)ENTRE(i) - N(i)p_l fc] \qquad \perp \quad CONS \tag{20}$$

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

The model set up computes what the optimal price $p_x(i)$ and output $X(i)$ would be for an inactive firm type but they don't affect the solution: $N(i) = 0$. This has the advantage of showing how unprofitable inactive firm types are at the solution.[4]

The final item for this section is the relationship between marginal costs $mc(i)$ across firm types, which are held constant in the counter-factuals to come. This is unobservable and it is unclear how this could be parameterized. Fortunately, CES demand allows for a mapping between observed market shares and unobserved marginal costs. The parameter $N.up(i)$ is quantitatively important to the results as well but that is postponed until section 6. The following shows that there is convenient mapping in which the observed market shares allow the recovery of marginal costs. We can use the sub-utility function W_x for the X sector because all the Y variables cancel out in deriving the equation we need. Setting $N(i) = 1$ for the moment to simplify the notation without loss of generality, the optimization problem is

$$\max W_x = (\sum_i X(i)^\alpha)^{1/\alpha} + \lambda(M - \sum_i p(i)X(i)) \tag{21}$$

where M is income and λ is the Lagrangean multiplier. Take the first-order conditions for types 1 and 2 and form the ratio, cancelling out some algebra including the Lagrangean multiplier λ.

[4] 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.

$$\frac{X_1^{\alpha-1}}{X_2^{\alpha-1}} = \frac{p_1}{p_2} \qquad \frac{X_1^{-1/\sigma}}{X_2^{-1/\sigma}} = \frac{p_1}{p_2} \qquad \frac{X_1}{X_2} = \frac{p_1^{-\sigma}}{p_2^{-\sigma}} \qquad => \tag{22}$$

$$\frac{p_1 X_1}{p_2 X_2} = \frac{s_1}{s_2} = \frac{p_1^{1-\sigma}}{p_2^{1-\sigma}} \qquad \frac{s_1^{1/(1-\sigma)}}{s_2^{1/(1-\sigma)}} = \frac{p_1}{p_2} \qquad cs(i)^{\frac{1}{1-\sigma}} = p(i) \tag{23}$$

where c is a (endogenous) constant of proportionality. Use the Bertrand conjecture for the relationship between p and mc.

$$p(i)(1 - mk(i)) = mc(i) \qquad mk(i) = \frac{1}{\sigma - s(i)(\sigma - 1)} \tag{24}$$

Substitute (24) in to (23) for $p(i)$. With five firm types, there are six equations in six unknowns, the sixth being the adding-up condition that is complementary to the proportionality variable c. σ is treated as exogenous and would be taken from literature estimates in an empirical application. Ratios of equations in (25) reduce to Melitz equation (6) for MC when $mk(i) = 1/\sigma$.

$$cs(i)^{\frac{1}{1-\sigma}}\left[1 - \frac{1}{\sigma - s(i)(\sigma - 1)}\right] = mc(i) \qquad \perp \quad s(i) \tag{25}$$

$$\sum_i s(i) = 1 \qquad \perp \quad c \tag{26}$$

Data should give market shares as parameters, so for these six equations the variables are $mc(i)$ and c ((26) already satisfied). But this could also be solved in the other direction, with marginal costs as the parameters, solving for the market shares and c that are consistent with those values (Meltiz does this assuming productivities ($1/mc(i)$). This latter is solved as an mcp to show the relationship between marginal costs and market shares in section 6.

5. 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.

Table 2 specifies the equations of the model, each with its complementary variable assignment. 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. Fixed-cost 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*. In two cases in section 6, these are chosen to produce figures that show the full range of outcomes, but I check using (25) and (26) that the implied market shares make sense. In two other cases in section 6, market shares are the data and these are used to solve for the marginal costs. 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. [5]

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
s(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
N.up(i)	maximum number of type i firms that can enter

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

[5]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 distribution, and then that calibrated distribution is in turn used to assign values to inactive firm types in the continuum. Some similar procedure could be used here.

Table 2: Simulation model formulated as a non-linear complementarity problem
All variables restricted to non-negative values

Pricing inequalities	Complementary variables are quantities	
$p_l mc(i) \geq p_x(i)(1 - mk(i))$	$\perp$ X(i)	(27)
$ENTRE(i) \geq p_l fc \geq ENTRE(i)$	$\perp$ $N.up(i) \geq N(i) \geq 0$	(28)
$p_l \geq p_y$	$\perp$ Y	(29)
$p_e^{0.5} p_y^{0.5} \geq p_w$	$\perp$ W	(30)
Market-clearing inequalities	Complementary variables are prices	
$X(i) \geq p_x(i)^{-\sigma}(p_e^{(\sigma - 1)})\, CONS/2$	$\perp$ $p_x(i)$	(31)
$Y \geq CONS/(2p_y)$	$\perp$ p_y	(32)
$W \geq CONS/p_w$	$\perp$ p_w	(33)
$ENDOW \geq Y + \sum_i N(i)(mc(i)X(i) + fc)$	$\perp$ p_l	(34)
Income balance	Complementary variables are incomes	
$CONS \geq p_l L + \sum_i (ENTRE_i - p_l fc)N_i$	$\perp$ $CONS$	(35)
$ENTRE(i) \geq p_x(i)\, mk(i) X(i)$	$\perp$ $ENTRE(i)$	(36)
Behavioral and definitional inequalities		
$s(i) \geq \dfrac{p_x(i)X(i)}{\sum_i p(i)X(i)N(i)}$	$\perp$ $s(i)$	(37)
$mk(i) \geq [\sigma - s(i)(\sigma - 1)]^{-1}$	$\perp$ $mk(i)$	(38)
$p_e \geq [\sum_i N(i)\, p_x(i)^{1-\sigma}]^{(1/(1-\sigma))}$	$\perp$ p_e	(39)

6. Calibrating different conjectures and interpreting *N.up(i)*

Before presenting simulations, a world about calibration when comparing differing formulations is in order. Three versions of the model are calibrated to the same (artificial) data. These are Cournot (NC), Bertrand (NB) and large-group monopolistic competition (MC). We look at the role of the maximum number of firms of a type which we could think of as a density measure. This formulation allows for bunches of firms at particular productivity/cost levels or alternatively a "jagged" firm size distribution which characterizes real data. As mentioned earlier, these *N.up(i)* values could be due to scarce top entrepreneurs, patents or other intangibles which prevent potential entrants from capturing the productivity of top firms.

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 (16), changing from one to the other requires changing the number of firms (the market share s) and/or changing the elasticity of substitution σ. Assume here that there are no upper bounds on firm numbers so that only the most productive firm type is active. I calibrate with the following elasticities for the three conjectures implying a different number of (identical) firms and market shares for the Cournot and Bertrand cases.

$$\begin{array}{llll} \text{NC:} & \sigma = \infty & s = 0.20 & (N = 5) \\ \text{NB:} & \sigma = 6.3333 & s = 0.25 & (N = 4) \\ \text{MC:} & \sigma = 5 & & \end{array} \tag{40}$$

With these assumptions, all three values in (40) yield a markup $1/\eta_c = 1/\eta_b = 1/\eta_{lg} = 0.2$. All three have the same output per firm *X(i)*, the same price $p_x(i)$ and the same welfare level (caveat: preferences are not the same!). In short, the benchmark results of the three cases with no *N.up(i)* constraint, we would see identical outcomes with only type 1 active. Assuming perfect substitutes in the Cournot case is conceptually useful, because then the Cournot example contrasts sharply with MC: in the former, welfare gains from a larger economy will be in the form of increase firms scale (productivity) and lower prices and markups. In MC, gains are in the form of more varietynot firm scale.

Two experiments are performed, with a focus on clarifying the role of *N.up(i)*. In the first, we increase *N.up(i)* uniformly across firm types holding the size (factor endowment) of the economy constant, although the code allows values to be set independently. Methodologically, it confirms that we converge to the single, most-productive firm type only as *N.up(i)* grows large: we are back to the classic identical-firm model. In the second, we raise *N.up(i)* but the economy's size grows in the same proportion. This is Krugman's (1979) parable for adding together identical economies. I chose to use different calibrations of *mc(i)* and *s(i)* for illustration because when I used the same values, one or the other cases were dominated by a couple of firm types and there was little regime switching. So the results don't have side-by-side continuity. In both cases, I use the NB assumption. To show it can work either way, in the constant-size case I set market shares and compute marginal costs. In the second case, I start with assumed marginal costs and solve for the market shares using the procedure of section 4.

Table 3: Relationship between market shares and marginal costs using NB

Experiment 1 Increase *N.up(i)* holding size constant Assumed *s(i)*	 Calibrated *mc(i)*	Experiment 2 Increase *N.up(i)* and size in proportion Calibrated *s(i)*	 Assumed *mc(i)*
0.400	1.000	0.447	1.000
0.167	1.263	0.175	1.300
0.153	1.286	0.148	1.350
0.143	1.305	0.125	1.400
0.137	1.319	0.106	1.450

We will look at data from the semi-conductor industry shortly, but I should note that these patterns of market share are qualitatively consistent with actual data for a number of industries I have (casually) looked at: the is generally a pattern where the market shares, starting with the biggest firm, are negatively sloped but convex, often with a long and fairly flat right-hand tail.

Figure 2 presents results with experiment 1 on the left panel and experiment 2 on the right. In both panels, each of the three cases is run in a loop on *N.up(i)*. The vertical axis indexes the five firm types, T1 being the lowest cost (most productive). As noted, the two panels are not quantitatively comparable as they use different calibrations to see the full range of outcomes. Results in panel A seem intuitive. With all three conjectures, adding firms lead to the successive elimination of less productive types as we move to the right. More firms per capita depresses the price and profits of the firms and so the less productive firms drop out.

Figure 2 panel B for the Cournot and Bertrand cases show the importance of the 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 all firms decrease, their markups fall and firm scale increases. This forces down the prices of the *X* goods and leads to losses for the least productive firm types.

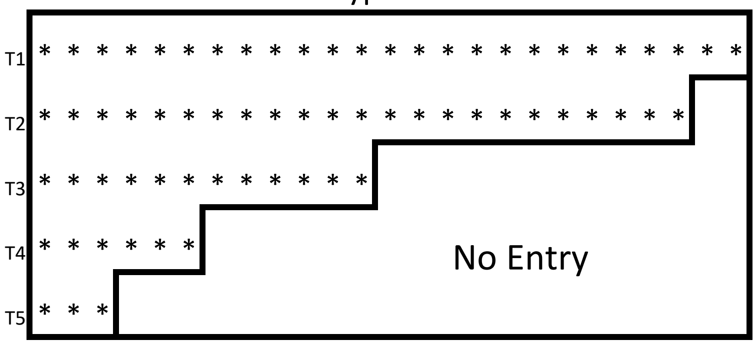
However, the MC case in Figure 2 panel B is a stark comparison and highlights the limitations of this traditional approach. Because markups are fixed at 1/σ for all firm types, growing the economy and *N.up*(i) in proportion to it just replicates quantities and prices, and active firm numbers for each firm type grow in the same proportion. Figure 2 suggests caution in making statements such as trade causes sorting and exit: this is not true in the Krugman experiment under MC.

Simulation results behind Figure 2 panel B, shown in Table 4, emphasize the differences in the composition of the effects contributing to welfare changes. The welfare gains from MC are purely from increased variety: no firm scale effects, no pro-competitive effects, no firm-type-selection effects, and constant markups. Cournot, with the goods being perfect substitutes, is the opposite 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

Figure 2: Role of N.up in determinng active firm types

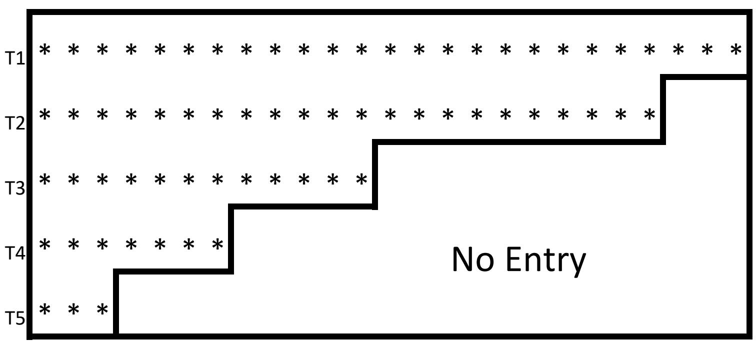
Panel A: Increase N.up(i)
holding economy size constant
converges to classic single-type
identical-firm outcome

Cournot: Active firm types *

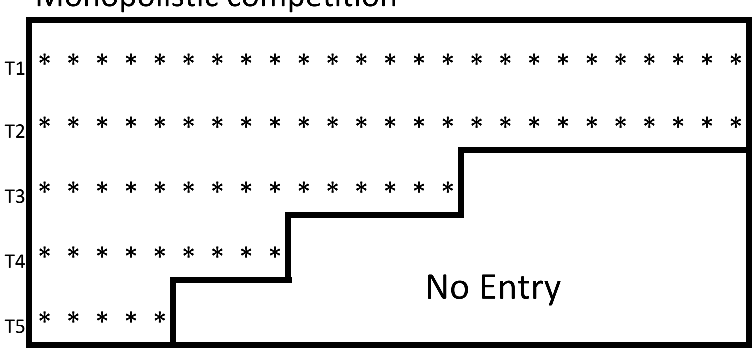


Horizontal axis: N.up from 1.0 to 3.4

Bertrand



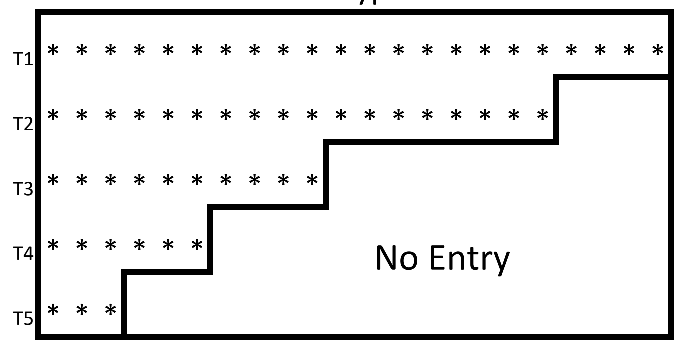
Monopolistic competition



MC will converge to just type T1, but higher
maximum value of N.up(i) needed

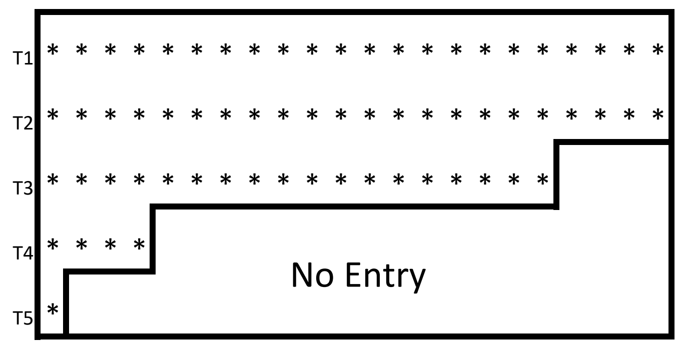
Panel B: Increase N.up(i)
Increasing economy size
in the same proportion

Cournot: Active firm types *

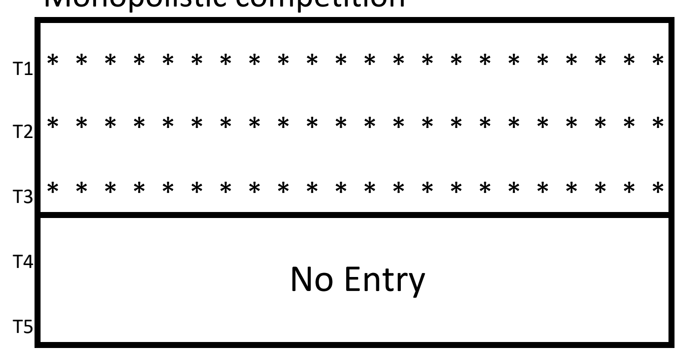


Horizontal axis: N.up from 0.5 to 4.0
Endowment from 300 to 2400

Bertrand



Monopolistic competition



cost), elimination of less productive firms, and a large fall in markups. Bertrand is in between, with added variety, higher firm scale, and lower markups all contributing to welfare.

Table 4: Left (small economy) to right (large economy) in Figure 2, panel B

	MC	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

Overall welfare changes do not show the path. Cournot welfare increases quickly when size is initially small, as firms move down the steep section of their average cost curves. But this effect diminishes when size becomes large. The implication for modeling is that initially small economies benefit greatly from trade under NC or NB, but this effect diminishes relative to MC when economies grow large. Again, the welfare changes are not really comparable insofar as each has different preferences.

The remainder of this section offers a simple example using data for the semi-conductor industry. I take the market shares of just the top five firms for comparison and normalize their shares to sum to one. I choose to use market capitalization as revenue shares are influenced a lot by the types of chips produced. In all the data I have looked at, market shares are negatively sloped but convex as noted above. There are of course many more firms in the market, but their market share are very small after firm 5 and the right-hand tail of the distribution is very flat.

Table 5: Observed market shares and marginal costs: semi-conductors

Experiment: Increase *N.up(i)* holding size constant (initially *N.up(i)* = 1)

Data *s(i)*	Calibrated *mc(i)*	(Bertrand, $\sigma = 6.333$)
0.476	1.000	
0.187	1.314	
0.186	1.315	
0.095	1.523	
0.056	1.694	

Figure 3a graphs the (actual) market sizes *s(i)* for the top five firms normalized to sum to one, showing the convex relationship alluded to earlier. The right-hand numbers in Table 5 are then the computed marginal costs *mc(i)* using the procedure outlined in equations (25) and (26) for NB. Figure 3b computes the counter-factual of increasing the density of firms holding the market size constant. The results look qualitatively very much like those in Figure 2 panel A, although I needed to have a larger range of *N.up(i)* in order to see the full range of configurations appearing. From this actual data and eye-balling other industries, I feel that the experiments above and values given in Table 3 bear a reasonable relationship to data in many industries.

Figure 3a: Market shares of top 5 semi-conductor firms
the five shares normalized to sum to 1 (top 5 is 82 percent of top 10)

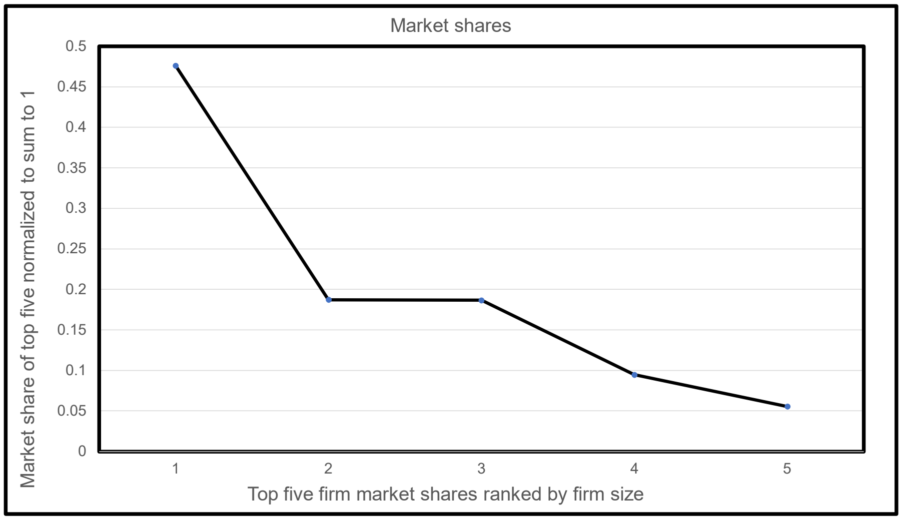
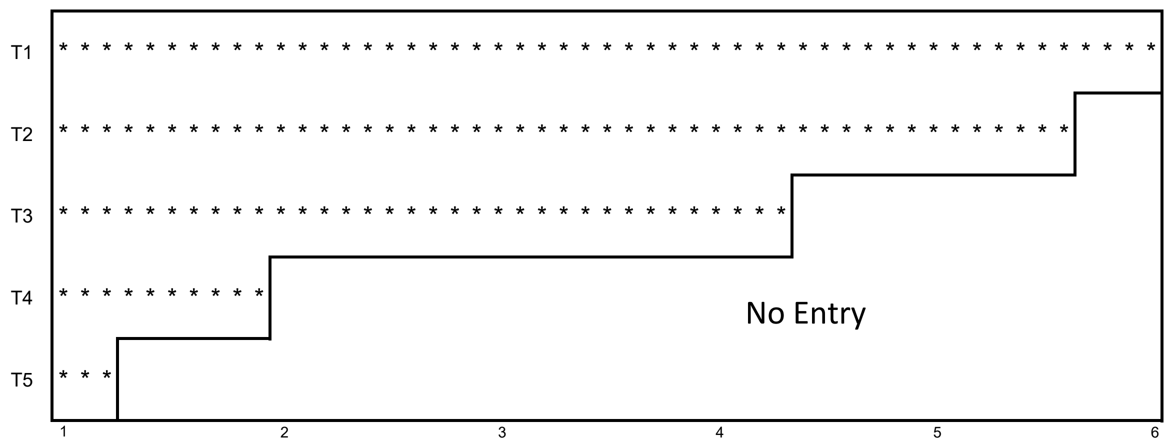


Figure 3b: Increase N.up(i) holding size constant
using market shares and calibrated marginal costs from semi-conductor data



Horizontal axis: N.up from 1.0 to 6.0   First column is the share data and calibrated mc

https://companiesmarketcap.com/semiconductors/largest-semiconductor-companies-by-market-cap/
N.B.: Top panel: these data are updated very frequently
different numbers will appear when you view the site

7. 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 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.

For the differentiated-goods Bertrand case used in this section I will assume that firms can price exports and domestic sales independently. which has two consequences. First, firms will charge different markups on domestic sales and export sales in the presence of trade costs: markets are segmented but arbitrage constraints don't bind in any case. This works through the market share variables in the markup equations. 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*. 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 equals the expenditure by the importer. I continue to use the five-firm-type formulation (set i).

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 there 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 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. *N.up(i)* denotes the maximum number of a type that can enter (ignoring country subscripts), which will be set equal to one for all firm types unless otherwise noted. 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.

$$N.up(i) \geq N_x(i) + N_d(i) \quad \perp \quad \lambda(i) \qquad (N.up(i) \text{ is a parameter, } N(i) \text{ are variables}) \tag{41}$$

This added weak inequality requires a complementary variable, denoted $\lambda(i)$ in (41), 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 (41). Let $\lambda(i)$ be the shadow tax on fixed costs for a particular country, with $\lambda(i)$ complementary to (41). 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

$$p_l fc_d + p_l \lambda(i) \geq ENTRE_d(i) \quad \perp \quad N_d(i) \tag{42}$$

$$p_l fc_x + p_l \lambda(i) \geq ENTRE_x(i) \quad \perp \quad N_x(i) \tag{43}$$

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 of 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 $\mathrm{N}.up(i)$ in equilibrium, and $\lambda(i)$ is endogenously set to make (42) and (43) equations when the number of firms hits $N.up(i)$. With tax revenue returned to (profits assigned to the consumer), the income balance constraint for country h is given as follows.

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

Table 6 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 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.[6]

Table 7 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 Tables 6-7 converts to a single firm-type model by simply setting the dimension of set *i* to a singleton. This can also be done by removing *N.up(i)* constraints. The model converts to MC by fixing the markup variables to $1/\sigma$ (GAMS then drops the complementary equations from the model).

---

[6]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*. A couple of experiments with asymmetric countries are included in a final section.

Table 6: Variables and parameters of the two-country Bertrand model

Quantity 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

Price 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

Income 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 constraints *N.up(i)*

$\lambda_h(i)$	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)
N.up(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 7: Inequalities of the two-country Bertrand model

Pricing inequalities, quantities complementary variables

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

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

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

Market clearing inequalities, prices complementary variables

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

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

$$ENDOW_h \ = \ Y_h + \sum_i \left[(N_{hd}(i) + N_{hx}(i)) X_{hh}(i) + N_{hx}(i) X_{hf}(i) \right] mc(i) + \sum_i \left[N_{hd}(i) fc_d + N_{hx}(i) fc_x \right] \qquad \perp \quad p_{lh}$$

$$ENDOW_f \ = \ Y_f + \sum_i \left[(N_{fd}(i) + N_{fx}(i)) X_{ff}(i) + N_{fx}(i) X_{fh}(i) \right] mc(i) + \sum_i \left[N_{fd}(i) fc_d + N_{fx}(i) fc_x \right] \qquad \perp \quad p_{lf}$$

Income balance inequalities, incomes complementary variables

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

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

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)

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

$$N.up(i) \geq N_{hd}(i) + N_{hx}(i) \qquad \perp \quad \lambda_h(i)$$

$$N.up(i) \geq N_{fd}(i) + N_{fx}(i) \qquad \perp \quad \lambda_f(i)$$

Figure 4 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.[7] 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. *N.up(i)* = 1 for all firm types for both countries, with type T1 in the top row of Figure 4. Results on active firm types in Figure 4 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.

Figure 5a 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 4 and 5a.

Figure 5b plots the markups for the two most productive firm types and shows both the domestic and export markups. 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 where domestic and export markups, as well as across types, are the same.

Figures 6a and 6b show some results for MC. Markups are fixed at 0.20 and the elasticity of substitution is changed from 6.333 to 5.0 as in (40) above in order to calibrate to the same benchmark data. Otherwise this model version is exactly the same. Fixed markups do not lead to results substantially different from the traditional MC approach with respect to firm exit and/or entry into exporting. Figure 6a is quantitatively a little different from Figure 4, but qualitatively similar. I have left the scale on the horizontal axis the same for comparison purposes. Types T2 and T3 begin exporting with falling trade costs, and T5 and T4 exit.

Figure 6b completes the analysis by showing the pattern of firm output for the MC case. Figure 6b looks qualitatively similar to Figure 4a. Again, the point is that the traditional results about firm entry to exporting and firm exit under MC continue to hold up in my alternative formulation.

---

[7] 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.

Figure 4: 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 5a: Output per firm by type
Small-group Bertrand competition

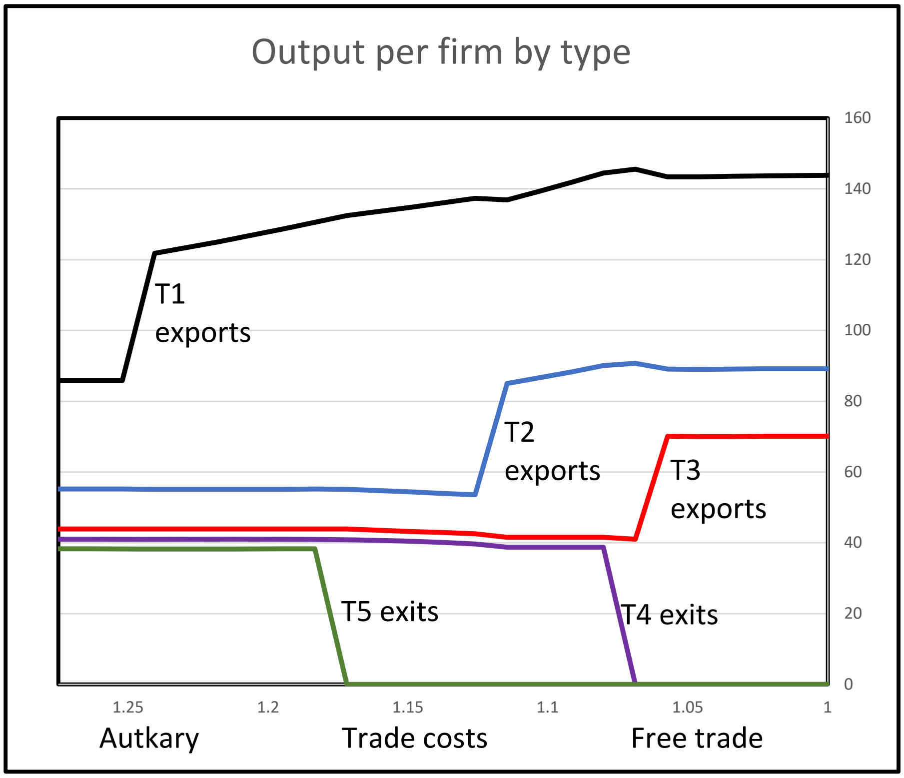


Figure 5b: Dometic and export markups
two most productive firm types: SGB

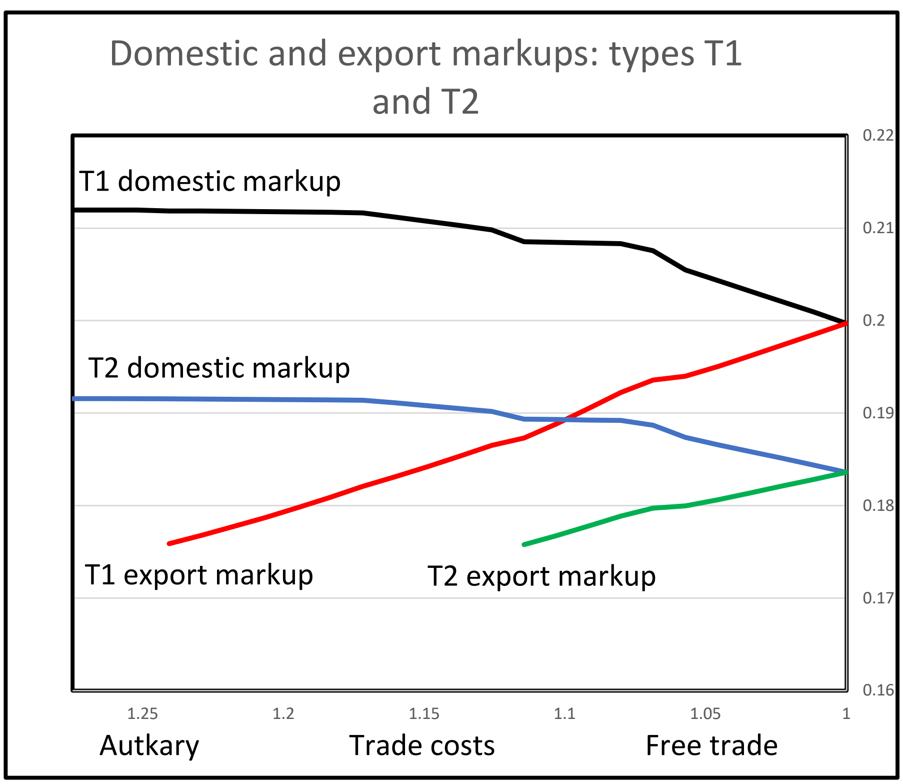


Figure 6a: Sequence of equilibria, autarky to free trade

Large group monopolistic competition

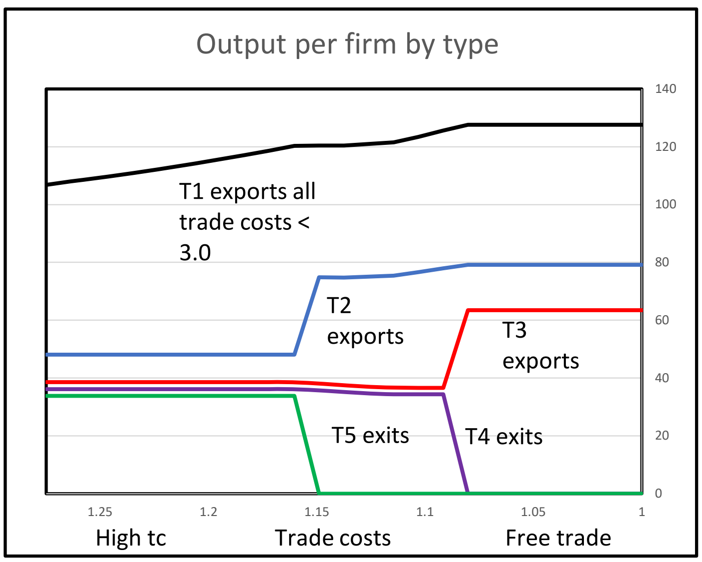
D = domestic sales only E = export sales N = no entry

	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
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

High tc Trade cost Free trade

Figure 6b: 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)

It is interesting to note that the welfare gains under MC (not shown) in Figures 6a, 6b 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 the text to the right of Figure 5b notes, 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 one issue identified and explored systematically in Matsuyama and Ushchev (2023). 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.

8. Extension 1: adding a horizontal multinational mode

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.[8]

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: trade and foreign production are substitutes.

Including three modes - 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 7 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

[8]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 firms only arise if countries are of different size or have different factor prices.

domestic and foreign sales. I will not present the equivalent of Tables 6 and 7 here, but the GAMS code is given in at the end of the paper. I use a parameter *NUP(i)* in the code, allowing users to specify levels other than one within or across firm types.

$$NUP(i) \geq N_{hd}(i) + N_{hx}(i) + N_{hm}(i) \quad \perp \quad \lambda_h(i) \tag{45}$$

$$NUP(i) \geq N_{fd}(i) + N_{fx}(i) + N_{fm}(i) \quad \perp \quad \lambda_f(i) \tag{46}$$

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

$$CONS_h = p_{lh}ENDOWH + \sum_i (N_{hd}(i) + N_{hx}(i) + N_{hm})\, p_{lh}\lambda_h(i) \tag{47}$$

Results of a simulation are shown in Figure 7. I have adjusted only one parameter (fixed cost of exporting) a small amount from Figure 4 in order to show the effect of the multinational option more clearly. So the upper Panel A of Figure 7 gives the equivalent of Figure 4 (multinational mode suppressed) for these slightly different parameters.

The introduction of the multinational option in Figure 7 Panel B leads to regime changes for high or moderate trade costs. At high trade costs, the two lowest-cost (most productive) firm types switch to foreign production from exporting and domestic-only production for types 1 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 switches to exporting. From that point on, panel B is the same as panel A.

There are two columns in Figure 7 Panel B that illustrate the focus of Helpman-Melitz-Yeaple (2004). These are two values of trade costs (1.08 and 1.07) 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.

9. Extension 2: examples of asymmetric countries and technologies

A principal motivation or rather "selling point" for this paper is that complementarity approach to general-equilibrium modeling is fully scalable to complex and high-dimension models. It expands to multiple sectors, factors of production, countries and accommodates intermediate use matrices, etc. Figure 8 present three examples of how asymmetries are easily examined, requiring only a couple of keystroke changes to parameters. This flexibility is valuable for empirically-relevant policy counter-factuals which are difficult in the probabilistic-continuum models.

The top two matrices of Figure 8 give the results corresponding to Figure 4 but with the countries of different sizes (exporting allowed but no mnes). Country h has twice the labor

Figure 7: Adding horizontal multinationals:
Sequence of equilibria, autarky to free trade

Panel A: multinational mode suppressed

	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
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

High tc Trade cost Free trade

Panel B: multinationals allowed

	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
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

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

Columns 1.08 and 1.07, Panel B, have the full set of four modes active and correspond to the analysis of Helpman et. al. (2004)

endowment of country f, and I have let country h have twice the *N.up(I)* values of country f. At very high trade costs, there are a couple of points where firm type 1 in country h has a (fractional) firm that is exporting and one that sells only domestically, cells denoted "B". Skipping to the right-hand area with low to zero trade costs, country h will have some middle-productivity firms that are only domestic due to the fixed cost of entering exporting. Country f has only exporting and no-entry firm types. However, country h is the net exporter of *X* good (net importer of *Y*), consistent with the classic presumption of home-market advantage.

It is interesting to note however that if both countries have the same value of *N.up(I)* (not shown), then country h turns out to be the net importer of *X*, net exporter of *Y*. Home-market advantage for the larger country depends on it having proportionately larger numbers of *potential* firms.

The middle two panels of Figure 8 consider a second example of asymmetry. Both countries have the same labor endowment or size, but country h has an absolute/comparative advantage in good *Y* production requiring 0.95 units of labor to produce one unit of *Y* instead of 1.0 units in country f. Both countries have the same technologies for *X* production for a firm type. Examining the right-hand area of the two panels with low trade costs, I think that the results are quite intuitive. Country f has more firms/firm types producing *X* and is the net exporter of *X*. As a application, think of the US as having a comparative advantage in services and agriculture, the *Y* sector in our example, though equally productive in *X*, manufacturing. The US will be a net importer of *X*, for a reason that is not caused by being a being a low productivity producer of manufactured goods.

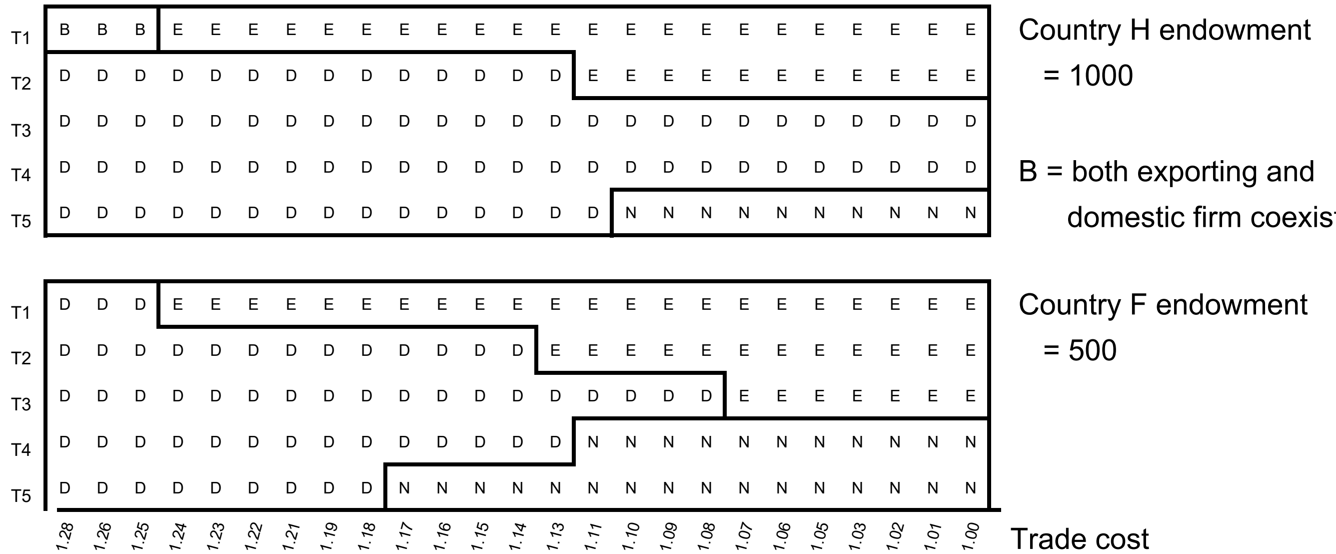
The bottom panel of Figure 8 is motivating the idea that the complementarity approach offers a richness of possibilities which I believe are not feasible in the probabilistic-continuum approaches. Specifically, in the Melitz model, small output firms are necessarily branded inefficient firms. But I imagine that we can all think of many counter examples where smaller market-share firms are highly profitable (Porsche). The last matrix in Figure 8 assumes that the ranking of the five firms by marginal cost is the same and thus so is their ranking by market shares. But firm type T4 has a lower fixed cost for domestic production and exporting: T4 is small but profitable. For the other parameters chosen, a fall in trade costs leads firm T4 to start exporting and T3 to exit rather than the other way around. There are certainly alternative reasons why some smaller firms may nevertheless be highly profitable, such as targeting niche markets, or high quality but high marginal cost products. But my point is that the complementarity framework allow for rich differences across individual firms as well as countries.

10. Summary

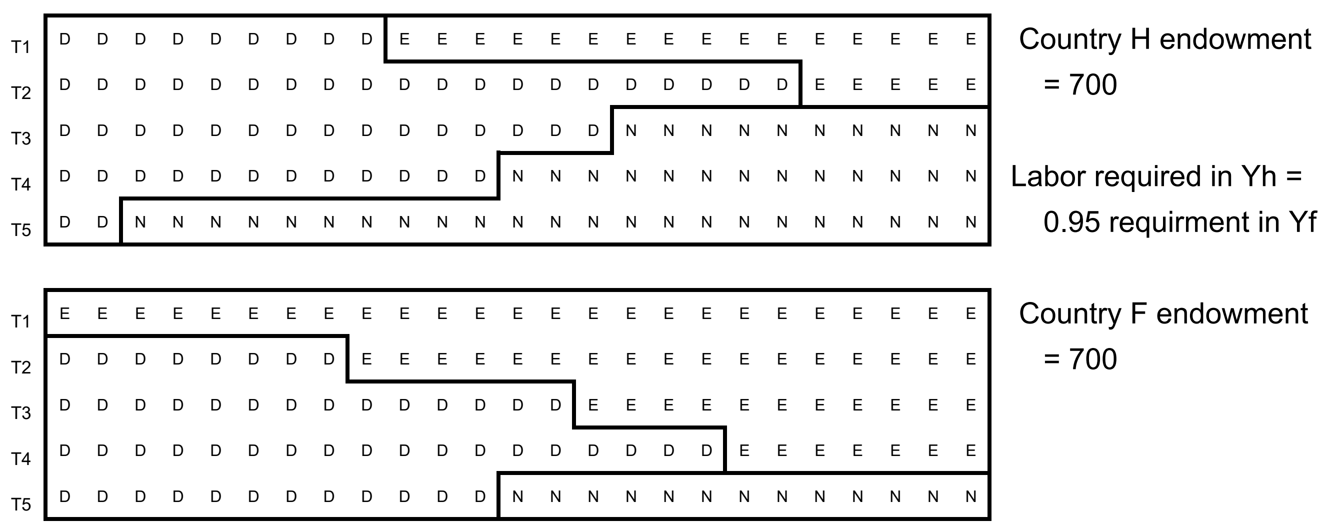
The general objective of this paper is to illustrate the advantages of formulating general-equilibrium models as non-linear complementarity problems. My view is that this allows modelers a much easier option for incorporating corner solutions, regime shifts, capacity constraints and endogenous zeros 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). They also have difficulty with large firms and discrete regime shifts.

Figure 8: Examples of asymmetries in countries or firm types

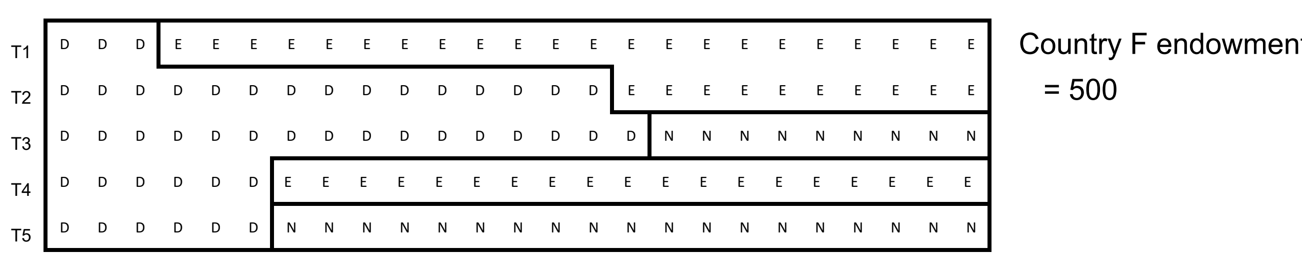
N(i).up is proportional to country size: Nh(i).up = 2*Nf(i).up, same for all i within a country



N(i).up = 1 is the same for both countries, Country h has 5 percent comparative advantage in Y



N(i).up = 1 is the same for both countries, type 4 has 35 percent lower fixed costs



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. Assuming that firms are atomistic yet produce with increasing returns is rather absurd. 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, 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.[9] Fourth, the addition of endogenous markups eliminates 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 bypasses an awkward endogeneity by firms paying for "draws" to find out their productivity is not needed. Sixth, I argue that my approach is completely scalable to high-dimension multi-country, multi-factor, multi-sector models, and many types of country and firm asymmetries. Finally, I imagine that working with real data requires aggregating firm-size distributions into discrete size classes in any case. All of these features are important for realistic policy analysis involving large firm entry/exit, mode switches and trade diversion. While I don't extend the analysis to protectionist policies here, it is easy to add tariffs for example in order to examine how policies can induce the switch from foreign firms exporting to producing via a subsidiary inside the protecting country.

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, with data on both sales market share and market capitalization, so as to not rule out small but highly productive firms. A difficulty is that data which are rich in firm detail are generally disjoint from detailed general-equilibrium data.

For generalizations to the theory, the knowledge-capital model (Markusen (2002)) incorporates vertical multinationals and could be extended in a heterogeneous firm context, allowing the analysis of the effects of policy on supply chains. More generally, there is a great need to allow for endogenous zeros in the world trade matrix in large competitive AGE models, such as allowing for minimum choke income levels and maximum choke prices.

---

[9] 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.

REFERENCES

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

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

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., Emmanuel Dhyne, Glenn Magerman, Kalina Manova, and Andreas Moxnes (2022), "The Origins of Firm Heterogeneity: A Production Network Approach", *Journal of Political Economy* 130(7), 1765-1804.

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(5), 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.

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

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.

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: A brief note on firm-level capacity constraints

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. Refineries may have maximum output levels. 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 (27) 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 $X.up(i) \geq X(i)$ 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 (27) becomes three equations in three unknowns

$$p_l mc(i) = p_x(i)(1 - mk(i)) + w(i) - v(i) \qquad (48)$$

$$w(i)X(i) = 0 \qquad (49)$$

$$v(i)(X.up(i) - X(i)) = 0 \qquad (50)$$

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

$$p_x(i) - p_l mc(i) \geq p_x(i) mk(i) \qquad \text{(e.g., } mk(i) \text{ is the Bertrand formula)} \tag{51}$$

$$ENTRE(i) = (p_x(i) - p_l mc(i))X(i) \geq p_x(i) mk(i) X(i) \tag{52}$$

Consider the Bertrand panel in Figure 2, panel B for example. The maximum *number of firms* of any type increases in proportion to the size of the economy. Assume that we constrain *output per firm* to equal the output per firm of the most productive type T1 at a small size (T for type, same index as I). 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.

However, how to implement this is not trivial. Fixing a variable creates a surplus or deficit such as positive or negative profits as we have seen above. 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 2: 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 is a simultaneous-move game.

Suppose that we constrain the most efficient firm type 1 at 0 (*N.fx(i1)* = 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(i1)* = 0. If we then free up type 1, then the next Newton iteration sets *N(i1)* 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, 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 we can add additional restrictions such that the first-mover solution blockades entry by the most efficient type 1? Yes, but added restrictions needed.

(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(i1)* = 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(i1)* = 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(i1)* = 0.6030 with all the other types making negative profits without the sunk cost assumption. Scenario S7 constrains type 1 to enter only as a full-sized firm *N(i1)* = 1. Firm type 1 now makes losses with the other types at their first-mover advantage sizes (cell with -3.09 outlined). But note that this requires restrictive assumptions, including the myopic assumption that less efficient types don't anticipate type 1 entering in the second stage and especially integer numbers of firms.

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, solulution 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
N1.up = 0.01	S3	0.08	3.48	0.71	-5.72	-6.48
N1.up = 0.25	S4	1.01	0.20	-2.18	-7.64	-8.28
N1.up = 0.50	S5	1.21	-1.15	-3.36	-8.42	-9.01
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

firm numbers
N2, N3, N4, N5
held fixed at
scentario S2 levels

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
N1.up = 0.01	S3	0.01	1	1	0	0
N1.up = 0.25	S4	0.25	1	1	0	0
N1.up = 0.50	S5	0.50	1	1	0	0
N1.up = 1.0 = >N1 = 0.60	S6	0.60	1	1	0	0
N1.fx = 1.0	S7	1	1	1	0	0

firm numbers
N2, N3, N4, N5
held fixed at
scentario S2 levels

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
N1.up = 0.01	S3	0.15	9.24	5.71	-2.49	-3.45
N1.up = 0.25	S4	2.48	5.06	2.03	-4.93	-5.74
N1.up = 0.50	S5	3.94	3.34	0.52	-5.92	-6.67
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

firm numbers
N2, N3, N4, N5
held fixed at
scentario S2 levels

```
$TITLE  SGB-add mnes.gms  adds horizontal mne option. Figures 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 - increasing returns, imperfect competition,  Y - constant returns, perfect competition
* Four entry modes: no entry, domestic, exporting, horizontal mne,  five firm types

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(I), FCX(I), FCM(I)  fixed cost for domestic exporting and mne firms
 NUP(I)                  maximum number of firms by type;

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

FCD(I) = 8;
FCX(I) = 14;
FCM(I) = 17;
NUP(I) = 1;
* the following used to calculate solution with mnes blocked:  FCM(I)(I) = 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(I)*PLH + LAMH(I)*PLH=G= ENTRHD(I);
PRNHX(I)..       FCX(I)*PLH + LAMH(I)*PLH=G= ENTRHX(I);
PRNHM(I)..       FCM(I)*PLH + LAMH(I)*PLH=G= ENTRHM(I);
PRNFD(I)..       FCD(I)*PLF + LAMF(I)*PLF=G= ENTRFD(I);
PRNFX(I)..       FCX(I)*PLF + LAMF(I)*PLF=G= ENTRFX(I);
PRNFM(I)..       FCM(I)*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(I) + NHX(I)*FCX(I) + NHM(I)*FCM(I));
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(I) + NFX(I)*FCX(I) + NFM(I)*FCM(I));

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)..        NUP(I) =G= NHD(I) + NHX(I) + NHM(I);
ELAMF(I)..        NUP(I) =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, "WELFCD(I)AP") = 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)$(NHD.L(I)+NHX.L(I)+NHM.L(I) GT 0) + EPS;
RESULTS1c(J, I) = (XHF.L(I))$(NHX.L(I) GT 0) + EPS;
RESULTS1d(J, I) = (XMHF.L(I))$(NHM.L(I) GT 0) + EPS;
RESULTS1e(J, I) = MARKHH.L(I)$(NHD.L(I)+NHX.L(I)+NHM.L(I) GT 0) + EPS;
RESULTS1f(J, I) = MARKHF.L(I)$(NHX.L(I) GT 0) + EPS;
RESULTS1g(J, I) = MARKMHF.L(I)$(NHM.L(I) GT 0) + EPS;

);

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

Execute_Unload 'RESULTS-MNE.gdx' RESULTS1
execute 'gdxxrw.exe RESULTS-MNE.gdx par=RESULTS1 rng=SHEET1!A3:F28'

Execute_Unload 'RESULTS-MNE.gdx' RESULTS1ad
execute 'gdxxrw.exe RESULTS-MNE.gdx par=RESULTS1ad rng=SHEET1!H3:M28'

Execute_Unload 'RESULTS-MNE.gdx' RESULTS1ax
execute 'gdxxrw.exe RESULTS-MNE.gdx par=RESULTS1ax rng=SHEET1!O3:T28'

Execute_Unload 'RESULTS-MNE.gdx' RESULTS1am
execute 'gdxxrw.exe RESULTS-MNE.gdx par=RESULTS1am rng=SHEET1!V3:AA28'

Execute_Unload 'RESULTS-MNE.gdx' RESULTS1b
execute 'gdxxrw.exe RESULTS-MNE.gdx par=RESULTS1b rng=SHEET2!A3:F28'

Execute_Unload 'RESULTS-MNE.gdx' RESULTS1c
execute 'gdxxrw.exe RESULTS-MNE.gdx par=RESULTS1c rng=SHEET2!H3:M28'

Execute_Unload 'RESULTS-MNE.gdx' RESULTS1d
execute 'gdxxrw.exe RESULTS-MNE.gdx par=RESULTS1d rng=SHEET2!O3:T28'

Execute_Unload 'RESULTS-MNE.gdx' RESULTS1e
execute 'gdxxrw.exe RESULTS-MNE.gdx par=RESULTS1e rng=SHEET3!A3:F28'

Execute_Unload 'RESULTS-MNE.gdx' RESULTS1f
execute 'gdxxrw.exe RESULTS-MNE.gdx par=RESULTS1f rng=SHEET3!H3:M28'

Execute_Unload 'RESULTS-MNE.gdx' RESULTS1g
execute 'gdxxrw.exe RESULTS-MNE.gdx par=RESULTS1g rng=SHEET3!O3:T28'
```