SML Estimation Based on First Order Conditions†

By

Michael P. Keane
Department of Economics
Yale University
New Haven, CT 06520-8264
michael.keane@yale.edu

November, 2003

† The statistical analysis of the confidential firm-level data on U.S. multinational corporations reported in this study was conducted at the International Investment Division, Bureau of Economic Analysis, U.S. Department of Commerce, under arrangements that maintained legal confidentiality requirements. Seminar participants at Princeton and the University of Minnesota provided useful comments. Conversations with Penny Goldberg, Lee Ohanian, Tom Holmes, Sam Kortum and Chris Sims were also very helpful. All remaining errors are of course my own.

Abstract

This paper describes a strategy for structural estimation of economic models that I will refer to as SML based on FOCs. In this approach, one uses simulated maximum likelihood (SML) to estimate the structural parameters that appear in the Euler or first order conditions (FOCs) solved by an optimizing economic agent.

I will argue that the SML based on FOCs approach has certain advantages over the generalized method of moments (GMM) approach to structural estimation based on FOCs. Most importantly, the SML based on FOCs approach can easily handle economic models that involve multiple structural sources of error. In contrast, GMM requires that all the structural sources of error enter the FOCs additively, so that a single composite additive error term may be obtained. In models with multiple sources of error, very strong assumptions on functional form or on information structure are often necessary in order to put FOCs in this form. Thus, the SML based on FOCs approach gives the econometrician much more flexibility in terms of how he/she can specify utility functions and/or production functions, particularly in terms of how these functions may be heterogeneous across agents.

Implementation of the SML based on FOCs approach requires the development of a number of new simulation algorithms that I develop here. These include two new recursive importance sampling algorithms that are the discrete/continuous and purely continuous data analogues the GHK algorithm for discrete data. These algorithms should have wide applicability to a range of econometric problems beyond the specific issues discussed here.

I illustrate the SML based on FOCs approach by using it to estimate a structural model of the behavior of U.S. MNCs with affiliates in Canada. The model is estimated on confidential BEA firm level data on the activities of U.S. MNCs over the period 1983-96. The method appears to work well in practice. That is, the computation time was manageable, and the algorithm converged steadily to stable estimates that were not very sensitive to starting values or simulation size. The estimated model fits the data reasonably well, and there appeared to be little evidence against the distributional assumptions that were required to implement the SML approach.

I. Introduction

This paper describes a strategy for structural estimation of economic models that I will refer to as SML based on FOCs. In this approach, one uses simulated maximum likelihood (SML) to estimate the structural parameters that appear in the Euler or first order conditions (FOCs) solved by an optimizing economic agent. I will argue that the SML based on FOCs approach has certain advantages over the generalized method of moments (GMM) approach to structural estimation based on FOCs.

The GMM approach of Hansen (1982) and Hansen and Singleton (1982)) has been popular (at least in part) because it is widely perceived as having one key advantage over a full information maximum likelihood (FIML) approach. Namely, in the GMM approach the econometrician does not need to completely specify the economic model in order to obtain estimates. There are two common ways in which a complete specification is avoided. First, if the FOCs involve expectation terms, it is common to substitute realized quantities for expected quantities, invoke a rational expectations assumption, and then assert that the forecast errors are orthogonal to the elements of agents' information sets at the time forecasts were made (see McCallum (1976)). Second, if the model structure is such that expectation errors and other structural sources of error (e.g., taste shocks, productivity shocks, etc.) enter the FOCs in a linearly additive fashion, then it is not necessary for the econometrician to specify their joint distribution parametrically. Rather, it is only necessary to assume that the composite error term (obtained by summing the structural errors) is mean independent of a specified set of instruments. One can then obtain moment conditions on which a GMM estimator is based.

I would argue however, that the assumptions necessary to put economic models into a form amenable to GMM estimation are often quite strong. Specifically, plausible economic models often involve multiple structural sources of error. In such cases, very strong assumptions on functional form or on information structure may be necessary in order to obtain FOCs where all the structural sources of error enter additively, so that a single composite additive error term may be obtained.

A prime example of this problem is provided in the recent paper by Krusell, Ohanian, Rios-Rull and Violante (2000). They estimate a production function with quality of skilled and unskilled labor as two latent stochastic inputs. The FOCs of their model also contain an unmeasured expectation of next period's price of capital. Thus, three stochastic terms enter nonlinearly, so the FOCs cannot be written in terms of a single additive error. Thus, Krusell et el. cannot use GMM. To deal with this problem, they developed a simulated pseudo-ML procedure for estimation based on FOCs. In this procedure, they assume that the forecast errors and the two shocks to the quality of skilled and unskilled labor are normally distributed.

The present paper is very closely related to Krusell, Ohanian, Rios-Rull and Violante (2000). However, I further extend and develop their approach in a number of ways. First, I show how to implement maximum likelihood as opposed to a pseudo-maximum likelihood procedure. Second, I show how to relax normality (via the Box-Cox transformation). Third, I show how to simulate the posterior distributions of the stochastic terms (conditional on the estimated model and the data) of the model. Given draws from the posterior distributions, one can test the distributional assumptions that underlie the SML approach. Testing distributional assumptions is particularly important here, given that the main reason for preferring GMM over a likelihood based procedure is concern over violation of distributional assumptions.

The SML based on FOCs approach developed here represents a compromise between a FIML approach and a GMM approach. As in GMM, one does not need a completely specified model. That is, one can avoid having to specify stochastic processes for all of the exogenous forcing variables that impact the environment of the agents. For instance, one can substitute realizations for expectations terms, and invoke a rational expectations assumption to deal with the resulting error terms. But, as in FIML, one must specify the joint distribution of all the structural sources of error (e.g., taste and technology shocks as well as forecast errors).

In my view, the SML based on FOCs approach has a number of potential advantages over GMM. A key advantage is that the SML approach can easily handle multiple structural sources of error entering the FOCs in a highly nonlinear way. This gives the econometrician the ability to be much more flexible in terms of how he/she specifies utility functions and/or production functions, particularly in terms of how these functions may be heterogeneous across agents. In my view it is ironic that researchers often claim to be using GMM in order to avoid making distributional assumptions on stochastic terms, yet at the same time they are often willing to make extremely restrictive modeling assumptions in other dimensions (i.e., the functional forms of utility and production functions, how heterogeneity enters the model, etc.) in order to put a model in a form amenable to a GMM approach.[1]

Another key advantage is that the researcher does obtain estimates of the distributions of the agent specific stochastic terms, and there are many contexts where these distributions are themselves of interest. This advantage is made clear by the particular illustrative application of the SML based on FOCs method that I present in this paper.

It is difficult to describe the SML based on FOCs approach to structural estimation in a

[1] Ackerberg and Caves (2003) provide an interesting discussion of the restrictive implications of models of firm behavior in which the number of structural sources of error is less than the number of factor inputs.

fully general context, because many of the details of how to implement the procedure will necessarily be specific to the particular economic model being studied. Thus, in this paper I illustrate the SML based on FOC approach by showing how to use it to estimate a particular model of the production and trade decisions of U.S. MNCs with affiliates in Canada. This model has been developed by Feinberg and Keane (2003a), and it gives a good illustration of how the approach works.

I should also stress that the SML based on FOCs approach is not a new estimation method. It is simply an application of SML, whose asymptotic properties have been developed, for example, in Lee (1992, 1995). Thus, I present no proofs of consistency and asymptotic normality for this method, since no new proofs are required. Rather, SML based on FOCs is a new strategy for structural estimation of economic models. That is, to my knowledge, SML has not previously been applied to estimate structural parameters using FOCs. It turns out that implementation of SML in this context is not at all straightforward. It requires the solution of a number of computational problems that do not arise in contexts where SML has previously been applied (such as estimation of discrete choice models), and therefore requires the development of a number of new simulation algorithms. Thus, to be clear, what is new in this paper is: (1) the suggestion of the SML based on FOCs strategy, and (2) the development of several new simulation algorithms necessary to implement that strategy. These new simulation algorithms are of independent interest, as they have broader applicability as well.

The particular empirical application to MNC behavior that I present illustrates well the potential advantages of SML based on FOCs over either FIML or GMM. The structural model of firm behavior that I consider is dynamic, because it allows for labor force adjustment costs. Hence, estimation using FIML would require one to specify how firms form expectations of future labor force size. This in turn would depend on the stochastic processes for several exogenous forcing variables: demand shocks, technology shocks, factor input prices, tariffs, exchange rates, etc.. It is easy to see how a researcher might feel reasonably comfortable specifying functional forms for the MNC production function, as well as for the product demand functions that the MNC faces, while being very reluctant to specify stochastic processes for all these forcing processes. The ability to avoid such assumptions is an advantage over FIML.

On the other hand, a researcher might well wish to specify a production function with several inputs, where the several parameters mapping these inputs into output are allowed to be heterogeneous across firms in a flexible way. As illustrated by the Krusell, Ohanian, Rios-Rull and Violante (2000) paper, allowing multiple production function parameters to be stochastic precludes putting the model in a form amenable to GMM estimation. However, this poses no problem for the SML based on FOCs approach. This is its main advantage over GMM.

A key feature of the model is that the stochastic terms are all continuous, yet they must fall in certain sub-regions of a high dimensional space in order for firm behavior to be rationalizable by the model. In other words, the FOCs cannot be satisfied if certain subsets of the vector of stochastic terms fall in certain regions of the error space. As a result, a high dimensional integral must be evaluated to construct the joint density of a firm's stochastic terms. In order to evaluate such integrals, I present a new recursive probability simulator that is the discrete/continuous analogue to the GHK method for simulation of discrete choice probabilities (see Keane (1994)).

This new simulator should have wide applicability in discrete/continuous simulation problems, just as GHK has been useful in a wide range of discrete choice problems. For instance, it could be useful in any situation where, in certain regions of the space of stochastic terms, a corner solution is induced where a firm would shut down, discontinue certain products or activities, etc., and where one wants to simulate the density of observed marginal decisions conditional on the firm being active, engaging in certain activities, etc.

Another aspect of the model, which is generic to situations where multiple stochastic terms enter the FOCs nonlinearly, is that the Jacobian of the transformation from the stochastic terms to the data is intractably complex. Thus, since the likelihood is the density of the stochastic terms times this Jacobian, one cannot even writer down the likelihood analytically. Nevertheless, I show how to construct a simulated numerical approximation to the Jacobian. This is what makes it possible to implement a likelihood based estimator (as opposed to the pseudo-ML estimator in Krusell et al.).

Finally, the illustrative model that I estimate here has 14 stochastic terms, yet it is estimated based on 12 FOCs. Models that allow for flexible patterns of heterogeneity in the agent specific parameters will typically have more stochastic terms than there are FOCs. For instance, the MNC production function I estimate here has a stochastic term associated with each input, and the number of FOCs is equal to the number of inputs. But then there are additional stochastic terms associated with expectation errors.

Thus, in order to generate the posterior distribution of the stochastic terms in the model conditional on the data, we need to simulate a $J\times1$ vector of continuous random variables subject to the constraint that it lies in a $K\times1$ dimensional space (where $K<J$). In order to this I develop a new recursive simulator for purely continuous distributions that is the continuous data analogue of the GHK method (Obviously this is a rather general problem that may arise in many contexts).

When I test the distributional assumptions of the model of MNC behavior by simulating the posterior distributions of the stochastic terms, I find that the distributions of the forecast errors appear strikingly close to normal. Statistical tests do not reject normality even at very low levels of significance. One argument for favoring GMM for the estimation of dynamic structural models based on FOCs is that we have no basis in theory for imposing distributional assumptions on forecast errors. Thus, the finding here that normality is not rejected is quite interesting.

The outline of the remainder of the paper is as follows. Sections II and III present the model of MNC behavior that I will use to illustrate the SML based on FOCs approach. Section IV shows how this model can be estimated using the SML based on FOCs approach. Section V presents the recursive simulation algorithm for simulating from the posterior distribution of the model parameters. Section VI describes the estimation results, including the tests of distributional assumptions. Section VII concludes.

II. The Illustrative Model

II.1. Overview

This section presents a model of the marginal production and trade decisions of a U.S. MNC with an affiliate in Canada, conditioning on the MNC's decision to place an affiliate in Canada. Each period, the MNC chooses the levels of factor inputs to utilize in both the U.S. and Canada. In addition, it chooses the levels of four types of trade flows: arms-length imports and exports, and intra-firm trade in intermediates from parent to affiliate and vice versa. Feinberg and Keane (2003a) develop and estimate a model of MNC behavior, and I will use that model to illustrate the SML based on FOCs approach.

The key assumptions of the Feinberg and Keane (2003a) model are as follows:

1) The parent and affiliate each produce a different good.

2) The good produced by the affiliate may serve a dual purpose: it can be sold as a final good to third parties (in Canada or the U.S.), or it may be used as an intermediate input by the parent. We make a symmetric assumption for the good produced by the parent.

3) Both the parent and affiliate have market power in final goods markets. They each produce a variety of a differentiated product. These products are non-rival (i.e., not substitutes).

4) The parent and affiliate both produce output using a CRTS Cobb-Douglas production function that takes labor, capital and materials as inputs. In addition, intermediates produced by the affiliate may be a required input in the parent's production process, and intermediates produced by the parent may be a required input in the affiliate's production process.

5) The affiliate and the parent both face iso-elastic demand functions in both the U.S. and Canadian final goods markets.

6) The parent and affiliate both face labor force adjustment costs.

7) The MNC maximizes the expected present value of profits in U.S. dollars, converting Canadian earnings to U.S. dollars using the nominal exchange rate.

8) The expected rate of profit is equalized across firms.

9) Parameters of technology and of demand are allowed to be heterogeneous both across firms and within firms over time.

Feinberg and Keane (2003a) estimate this model using the Benchmark and Annual Surveys of U.S. Direct Investment Abroad administered by the Bureau of Economic Analysis (BEA) for the 1983-1996 period. A key data problem that influences the set up of the model is that the BEA data do not contain separate information on quantities of production and prices. This problem plagues most production function estimation. It has been typical in the literature on production function estimation to simply use industry level price indices to deflate nominal sales revenue data in order to construct real output. But Griliches and Mairesse (1995) and Klette and Griliches (1996) have pointed out that this procedure is only valid in perfectly competitive industries, so that price is exogenous to the firms. This condition is obviously violated for MNCs, since they have market power. This problem has received a great deal of attention recently in the IO literature (see, e.g., Katayama, Lu and Tybout (2003) and Levinsohn and Melitz (2002)).

The only general solution to the problem of endogenous output prices is to estimate the production function jointly with an assumed demand system. But in the present case the problem is further exacerbated by the fact that, while the BEA data reports nominal values of intra-firm flows, imports and exports – the prices and quantities for these flows cannot be observed separately. Nor can we separate price and quantity for capital and materials inputs, or for intermediate inputs shipped intra-firm. Furthermore, the price of such intermediate inputs is endogenous, since it depends on the MNC's other input and trade decisions.

The only general solution to the problems created by the inability to observe prices and quantities of outputs or intermediate inputs separately is to assume: 1) constant returns to scale (CRTS) Cobb-Douglas production functions for both the parent and affiliate, and 2) that both parent and affiliate face isoelastic demand in the market for final goods. These two assumptions enable one to identify the price elasticities of demand faced by parents and affiliates using only information on revenues and costs (i.e., by exploiting Lerner type conditions). Then, given the elasticities of demand, one can pin down the Cobb-Douglas share parameters using only information on factor shares of revenues (appropriately modified to account for market power).

The solutions proposed by Katayama, Lu and Tybout (2003) and Levinsohn and Melitz

(2002) allow estimation of more general production functions, but these solutions assume that real input quantities are observed. In the present case, generalizations of Cobb-Douglas seem infeasible, because one does not observe input price variation that identifies substitution elasticities. This is what motivates the Cobb-Douglas production and isoelastic demand assumptions in the Feinberg and Keane (2003a) model.

Data limitations also motivate assumption 8. The BEA capital stock data is rather imprecise (i.e., PPE at historical cost). This is of course a very general problem not limited to the BEA data. As we discuss below, the assumption of an equalized (expected) profit rate across firms will enable us to dispense with the capital stock data entirely, and to instead construct payments to capital as a residual using the other available cost and revenue data. The theoretical justification for this assumption, as well as the issue of how the profit rate is identified, is discussed further below. We refer the reader to Feinberg and Keane (2003a) for further discussion of the modeling assumptions.

II.2. Basic Structure of the Model

In this section I present the equations of the model in the most general case in which a single MNC exhibits all four of the potential trade flows (arms-length imports and exports and bilateral intra-firm trade). Instances where an MNC has only a subset of these four flows are special cases.[2] Let Q_d and Q_f denote total output of the parent and affiliate, respectively. Let N_d denote the part of affiliate output shipped to the parent for use as intermediate. Similarly, let N_f denote intermediates transferred from the parent to the affiliate. I (imports) denotes the quantity of goods sold arms-length by the Canadian affiliate to consumers in the U.S., and E (exports) denotes arms-length exports from the U.S. parent to consumers in Canada. Thus, $S_d \equiv (Q_d - N_f - E)$ is the quantity of its output the parent sells in the U.S., and $S_f \equiv (Q_f - N_d - I)$ is the quantity of its output the affiliate sells in Canada.

Finally, let P denote prices, with the superscript $j=1,2$ denoting the good (i.e., that produced by the parent or the affiliate) and the subscript $c=d,f$ denoting the point of sale. Since we do not observe prices and quantities separately in the data, we will work with the six MNC firm-level trade and domestic sales flows, which are $P_d^2 I$, $P_f^1 E$, $P_d^1 N_f$, $P_f^2 N_d$, $P_d^1 S_d$, and $P_f^2 S_f$.

The MNC's domestic and Canadian production functions are Cobb-Douglas, given by:

[2] If MNC decisions about whether to utilize each of the 4 potential trade flows are correlated with firm specific unobservables, then treating these decisions as exogenous could create bias in estimates of the structural model. To deal with this problem, the model presented in this section was actually estimated jointly with reduced from decision rules for whether an MNC chose to engage in each of the 4 potential trade activities. Estimates of that reduced form model are discussed in Feinberg and Keane (2003b).

(1) $Q_d = H_d K_d^{\alpha^{Kd}} L_d^{\alpha^{Ld}} N_d^{\alpha^{Nd}} M_d^{\alpha^{Md}}$

(2) $Q_f = H_f K_f^{\alpha^{Kf}} L_f^{\alpha^{Lf}} N_f^{\alpha^{Nf}} M_f^{\alpha^{Mf}}$

Note that there are four factor inputs: capital (*K*), labor (*L*), intermediate goods (*N*) and materials (*M*). I assume that the share parameters α sum to one for both the parent and affiliate (CRTS). I will allow the constants H_d and H_f to follow time trends in order to capture TFP growth.

For the domestically produced good (good 1), the MNC faces the following iso-elastic demand functions in the U.S. and Canada:

(3) $P_d^1 = P_{0d}^1 S_d^{-g_1}$ $\qquad P_f^1 = P_{0f}^1 E^{-g_1}$ $\qquad 0<g_1<1$

Similarly, for the good produced in Canada (good 2), the MNC faces the demand functions:

(4) $P_f^2 = P_{0f}^2 S_f^{-g_2}$ $\qquad P_d^2 = P_{0d}^2 I^{-g_2}$ $\qquad 0<g_2<1$

Recall that S_d denotes the quantity of the U.S. produced good sold in the U.S., and S_f denotes the quantity of affiliate sales in Canada. The g_1 and g_2 are the (negative) inverses of the price elasticities of demand for the domestic and foreign produced good, respectively.

Next, we assume the MNC faces labor force adjustment costs. It is often assumed such costs are quadratic, e.g.: $AC_{dt} = \delta_d [L_{dt} - L_{d,t-1}]^2$, where $\delta_d > 0$. However, Feinberg and Keane (2003a) found that a generalization of this function led to a substantial improvement in fit and could accommodate many reasonable adjustment cost processes[3]:

(5) $AC_{dt} = \delta_d \left((L_{dt} - L_{d,t-1})^2\right)^{\mu} / L_{d,t-1}^{\Delta}$ $\qquad$ where $\delta_d > 0, \mu > 0, \Delta \geq 0.$

A similar adjustment cost function is specified for the affiliate, which will be allowed to have a different δ parameter (δ_f). The curvature parameters μ and Δ are assumed to be common.

We can write the MNC's period specific profits (suppressing the time subscripts) as:

(6)
$$\begin{aligned}\Pi = {} & P_d^1(Q_d - N_f - E) - P_d^1 N_f(T_f + C_f) + P_f^1 E(1 - T_f - C_f) \\ & + P_f^2(Q_f - N_d - I) - P_f^2 N_d(T_d + C_d) + P_d^2 I(1 - T_d - C_d) \\ & - w_d L_d - w_f L_f - \phi_d M_d - \phi_f M_f - \gamma_d K_d - \gamma_f K_f - AC_d(L_d, L_d^{(-1)}) - AC_f(L_f, L_f^{(-1)})\end{aligned}$$

Here, T_f and C_f are the *ad valorem* Canadian tariff and transportation costs the MNC faces when shipping products from the U.S. to Canada (and similarly for T_d and C_d). Note that the use of an *ad valorem* transport cost is consistent with the common "iceberg" assumption.

---

[3] For example, setting $\mu = 1$ and $\Delta = 0$ produces $\delta_d [L_{dt} - L_{d,t-1}]^2$. Similarly, $\mu = \frac{1}{2}$ and $\Delta = 1$ gives $\delta_d \left|(L_{dt} - L_{d,t-1}) / L_{d,t-1}\right|$.

The exchange rate enters (6) implicitly because Canadian affiliate costs and revenues are converted into U.S. dollars using the nominal exchange rate. Thus, the MNC cares about U.S. dollar profits (and hence U.S. dollar output and input prices). w_d and w_f are the domestic and foreign real wage rates respectively, and ϕ_d and ϕ_f are the domestic and foreign materials prices. γ is the price of capital, which we assume is equal for the parent and the affiliate ($\gamma_d = \gamma_f$).

The MNC's problem is to maximize the expected present value of profits in real U.S. dollars $E\sum_{\tau=1}^{\infty}\beta^{\tau}\Pi_{t+\tau}$ by choice of eight control variables $\{L_{dt}, M_{dt}, K_{dt}, N_{dt}, L_{ft}, M_{ft}, K_{ft}, N_{ft}\}$. The solution to this problem will generate shadow prices on intermediates shipped from the parent to affiliate and vice-versa.

Finally, recall assumption 8, that the rate of profit, defined as $R=\Pi/\gamma K$, is equalized (in expectation) across firms. This can be justified by assuming a world where capitalists (as residual claimants) decide how to allocate capital across industries (or varieties of differentiated products) based on expected profit rates. In equilibrium, the profit rate will be equalized across industries (as entry reduces R). The equilibrium R will be positive if there is a fixed cost of entry that is proportional to size of the capital stock.

The virtue of assuming a particular profit rate is that one can back out the payments to capital from data on total revenues and payments to the other factors (rather than using PPE data).[4] Thus, I treat the profit rate R as an unknown parameter to be estimated. I discuss the intuition for its identification in section IV.3.B.

II.3. Solution of the Firm's Problem and Derivation of the Estimable FOCs

We can express the FOCs more compactly if we first define:

$$A=\left(\frac{P_d^1(Q_d-N_f-E)-P_d^1N_f(T_f+C_f)}{P_d^1(Q_d-N_f-E)}\right)=\left(\frac{P_d^1S_d-P_d^1N_f(T_f+C_f)}{P_d^1S_d}\right)$$

[4] The procedure works as follows. Denote domestic revenue by RD, domestic costs by CD* and domestic costs excluding capital costs by CD1. These quantities are given by:

$$RD = P_d^1 S_d + P_d^1 N_f + (1-T_f-C_f)(P_f^1 E)$$

$$CD^* = w_d L_d + \gamma_d K_d + \phi_d M_d + P_f^2 N_d (1+T_d+C_d) + AC_d$$

$$CD1 \equiv CD - \gamma_d K_d$$

Now, let R_K denote the fraction of operating profit that is pure profit, leaving $(1-R_K)$ as the fraction that is the payment to capital. This gives $\Pi_d = R_K \cdot [RD-CD1]$ and thus:

$$\gamma_d K_d = (1-R_K)\cdot[RD-CD1]$$

Thus, the rate of profit for domestic operations is $R = \Pi_d / \gamma_d K_d = R_K/(1-R_K)$. We treat R as a common parameter across firms and countries that we estimate (we also assume it is equal for the parent and the affiliate). That is, for the affiliate we have the analogous equation:

$$\gamma_f K_f = (1-R_K)\cdot[RF-CF1]$$

$$B=\left(\frac{P_f^2(Q_f-N_d-I)-P_f^2N_d(T_d+C_d)}{P_f^2(Q_f-N_d-I)}\right)=\left(\frac{P_f^2S_f-P_f^2N_d(T_d+C_d)}{P_f^2S_f}\right)$$

and express the adjustment cost term in the FOC for domestic labor (L_d) as $E(FD)$, where:

$$FD=\partial\mu((L_{dt}-L_{d,t-1})^2)^{\mu-1}(L_{dt}-L_{d,t-1})/L_{d,t-1}^{\Delta}$$
$$-\beta\mu((L_{dt+1}-L_{dt})^2)^{\mu-1}(L_{dt+1}-L_{dt})/L_{dt}^{\Delta}-\beta\Delta((L_{dt+1}-L_{dt})^2)^{\mu}(L_{dt+1}-L_{dt})/L_{dt}^{\Delta+1}$$

The adjustment cost term in the FOC for Canadian labor (L_f) is $E(FF)$, where FF is defined similarly.

The first order conditions for parent factor inputs and parent's exports to Canada are then:

$$L_d:\alpha^{Ld}(1-g_1A)\left(\frac{P_d^1Q_d}{L_d}\right)-w_d-E(FD)=0$$

$$K_d:\alpha^{Kd}(1-g_1A)\left(\frac{P_d^1Q_d}{K_d}\right)-\gamma_d=0$$

$$M_d:\alpha^{Md}(1-g_1A)\left(\frac{P_d^1Q_d}{M_d}\right)-\phi_d=0$$

$$N_d:\alpha^{Nd}(1-g_1A)\left(\frac{P_d^1Q_d}{N_d}\right)+g_2P_f^2B-(1+T_d+C_d)P_f^2=0$$

$$E:(1-g_1)P_f^1(1-T_f-C_f)-(1-g_1A)P_d^1=0$$

For the affiliate, the first order conditions for L_f, K_f, M_f, N_f and I are similar.

Note that the FOC for N_d, intermediates shipped from the affiliate to the parent, equates the marginal revenue product from increasing the input of N_d in domestic production to the effective cost of importing N_d. The effective cost (or shadow price) can be written $(1-g_2B)P_f^2+(T_d+C_d)P_f^2$. The first term is the marginal revenue from selling N_d in Canada. The component $g_2BP_f^2$ arises because the affiliate has market power. The second term is the tariff and transport cost, which is based on the "transfer price" P_f^2. Thus, I assume that the transfer price is equal to price that the firm charges third parties in Canada.[5] Note that the shadow price is a

[5] MNCs must set transfer prices for accounting purposes. It enables them to allocate profits to different countries for tax purposes, and it enables then to calculate tariffs. I assume that the transfer price is equal to P_f^2 because both the U.S. and Canadian revenue services require a "third party standard" whereby the transfer price should be set equal to the price charged to unaffiliated buyers. Canada has lower tax rates on manufacturing than the U.S., so there is an incentive to manipulate the transfer price to shift profits to Canada. However, Eden (1998) finds no evidence that transfer price manipulation is significant in eth U.S.-Canada context. Given that the tax differential is small and enforcement is relatively strict, it may be easier to transfer profits using licensing fees on intangibles.

completely separate quantity from the transfer price, since obviously it is not optimal for the affiliate to charge the parent the same price it charges third parties.

Since prices and quantities are not separately observed, one cannot take these FOCs directly to the data. They must first be manipulated to obtain estimable equations that contain only observed quantities and unknown model parameters. First, by multiplying each first order condition by the associated control variable, we obtain:

$$
\begin{aligned}
&L_d : \alpha^{Ld}(1-g_1A)(P_d^1Q_d) - w_dL_d - \delta_d E(FD)L_d = 0 \\
&K_d : \alpha^{Kd}(1-g_1A)(P_d^1Q_d) - \gamma_d K_d = 0 \\
(7)\qquad &M_d : \alpha^{Md}(1-g_1A)(P_d^1Q_d) - \phi M_d = 0 \\
&N_d : \alpha^{Nd}(1-g_1A)(P_d^1Q_d) + g_2(P_f^2N_d)B - (1+T_d+C_d)(P_f^2N_d) = 0 \\
&E : (1-g_1)(P_f^1E)(1-T_f-C_f) - (1-g_1A)(P_d^1E) = 0
\end{aligned}
$$

In the FOC for E, the quantity P_d^1E is not observable.[6] However, one can exploit the fact that:

$$
(P_d^1E) = \left(\frac{P_d^1}{P_f^1}\right)(P_f^1E) = \left(\frac{P_{0d}^1}{P_{0f}^1}\right)\left(\frac{P_d^1S_d}{P_f^1E}\right)^{-g1} \cdot (P_f^1E)
$$

to express the FOC for E in terms of observable quantities and the demand function intercepts $\left(P_{0d}^1 / P_{0f}^1\right)$, which are treated as unknown parameters, as follows:

$$
(8)\qquad \text{E: } (1-g_1)(P_f^1E)(1-T_f-C_f) - (1-g_1A)\left(\frac{P_{0d}^1}{P_{0f}^1}\right)\left(\frac{P_d^1S_d}{P_f^1E}\right)^{-g1} \cdot (P_f^1E) = 0
$$

Similarly, in the FOCs for the factor inputs, the quantity $(P_d^1Q_d)$ is also not observed. We can rewrite this quantity as $P_d^1Q_d = P_d^1S_d + P_d^1N_f + \left(P_d^1 / P_f^1\right)(P_f^1E)$ but, again, P_d^1E is not observed. I therefore repeat the same type of substitution to obtain, for domestic labor:

$$
(9)\quad L_d : \alpha^{Ld}(1-g_1A)\left[P_d^1S_d + P_d^1N_f + \left(\frac{P_{0d}^1}{P_{0f}^1}\right)\left(\frac{P_d^1S_d}{P_f^1E}\right)^{-g_1}(P_f^1E)\right] - w_dL_d - \delta_d E(FD)L_d = 0
$$

The FOCs for K_d, M_d and N_d, and for the affiliate, are obtained similarly.

---

[6] Note: $P_d^{1}E$ is the physical quantity of exports times their domestic (not foreign) price – an object we cannot construct since we do not observe prices and quantities separately

III. Stochastic Specification

The model contains eight parameters (R, β, δ_d, δ_f, μ, Δ, H_d and H_f) that are common across firms. The model also contains eight technology parameters (α^{Kd}, α^{Ld}, α^{Nd}, α^{Md}, α^{Kf}, α^{Lf}, α^{Nf}, α^{Mf}) and six demand function parameters (g_1, P^1_{0d}, P^1_{0f}, g_2, P^2_{0d} and P^2_{0f}) that are allowed to be heterogeneous both across firms and within firms over time. Given CRTS, two of the Cobb-Douglas share parameters (α) are determined by the other six. Thus, there are 12 fundamental parameters that can vary independently. In addition, the two unobserved expectation terms $E(FD)$ and $E(FF)$ that appear in the FOCs for U.S. and Canadian labor must also be dealt with.

III.1. Production Function Parameters

Allowing the Cobb-Douglas share parameters to be stochastic, while also imposing that they are positive and sum to one (CRTS), is challenging. To impose these constraints, I use a logistic-type transformation, treating the share parameters as analogous to choice probabilities in a multinomial logit (MNL) model. For instance, for the domestic labor share parameter, we have, suppressing firm and time subscripts:

$$\text{(10)} \qquad \alpha^{Ld} = \frac{\alpha_R^{Ld}}{1+\ \alpha_R^{Ld} + \alpha_R^{Md} + \alpha_R^{Kd}} = \frac{G\{x\overline{\alpha}^{Ld} + \varepsilon^{Ld}\}}{1+G\{x\overline{\alpha}^{Ld} + \varepsilon^{Ld}\} + G\{x\overline{\alpha}^{Md} + \varepsilon^{Md}\} + G\{x\overline{\alpha}^{Kd} + \varepsilon^{Kd}\}}$$

where $G(\cdot)>0$ is a positive function, the vector x_{it} includes all firm characteristics that shift the share parameters, and $\overline{\alpha}^{Ld}$ is a corresponding vector of parameters.

The expressions for α^{Kd} and α^{Md} are similar. Note the expression for α^{Nd} is:

$$\text{(11)} \qquad \alpha^{Nd} = \frac{1}{1+G\{x\overline{\alpha}^{Ld} + \varepsilon^{Ld}\} + G\{x\overline{\alpha}^{Md} + \varepsilon^{Md}\} + G\{x\overline{\alpha}^{Kd} + \varepsilon^{Kd}\}}$$

So α^{Nd} plays the role of the "base alternative" in a multinomial logit model. This specification insures that, given any values for the x_{it} and any values for the stochastic terms ε_{it}, the Cobb-Douglas share parameters are guaranteed to be positive and sum to 1.

Note that in (10) the quantities α_R^{Ld}, α_R^{Md} and α_R^{Kd} are simply latent variables that map into the firm specific share parameters. If we specify that $G(a) = exp(a)$ and that the ε are normal we obtain a specification where α_R^{Ld}, α_R^{Md} and α_R^{Kd} are log normal. Then, for example, the stochastic term α_R^{Ld} would be given by:

$$\ln\alpha_R^{Ld} = x\overline{\alpha}^{Ld} + \varepsilon^{Ld} \qquad\qquad \varepsilon^{Ld} \sim N(0,\sigma^2_{Ld})$$

and similar equations could be specified for α_R^{Md}, α_R^{Kd}.

We can easily generalize log normality by using a Box-Cox transformation, given by

$G(a)=(a^{bc}-1)/bc$, where bc is the Box-Cox parameter.[7] Using a Box-Cox transformation with parameter $bc(1)$, we obtain:

$$(12) \qquad \frac{\left(\alpha_R^{Ld}\right)^{bc(1)}-1}{bc(1)} = x\overline{\alpha}^{Ld} + \varepsilon^{Ld} \qquad\qquad \varepsilon^{Ld} \sim N(0,\sigma^2_{Ld})$$

Expressions similar to (12) hold for the parameters α_R^{Md}, α_R^{Kd} and also for the affiliate parameters. I will denote the Box-Cox parameters in these equations as $bc(2)$ through $bc(6)$.

Next, consider the specification of x_{it}, the vector of firm characteristics that shift the share parameters. In the empirical application I allow x_{it} to include an intercept and a time trend t ($t=0$ in 1983). I also allow these intercepts and time trends to differ for parents (affiliates) that do and do not use intermediate inputs from affiliates (parents).

If the U.S. parent is not structured to use intermediate inputs from the affiliate, then $\alpha^{Nd}=0$, and we must constrain the remaining three share parameters, α^{Ld}, α^{Md} and α^{Kd}, to sum to one. This is done just as above, except that now we let α^{Kd} play the role of the base alternative. A similar construct is used for affiliates that do not use intermediates from the parent. Because the scale of the coefficients in a MNL model with three alternatives is quite different from that of a MNL model with four alternatives, I also introduce a scaling parameter, denoted SC_d, that scales down the error terms in the three alternative case.

Thus, for the α_R^{Ld} equation, we have:

$$(13) \qquad \frac{\left(\alpha_R^{Ld}\right)^{bc(1)}-1}{bc(1)} = \alpha_0^{Ld} + \alpha_{shift}^{Ld}\, I[N_d>0] + \alpha_{Time}^{Ld}\cdot t\cdot I[N_d>0] + \alpha_{Time}^{Ld}\cdot t\cdot I[N_d=0]\cdot SC_d$$
$$+\ \varepsilon^{Ld}\,\{ I[N_d>0] + SC_d\cdot I[N_d=0]\}$$

Similar equations hold for α_R^{Md} and α_R^{Kd}, except that we simply have $\alpha_R^{Kd}=1$ in the $N_d=0$ case.[8] The same scaling parameter, SC_d, applies in the α_R^{Md} equation in the $N_d=0$ case. The equations for the affiliate share parameters are similar.

---

[7] The model was originally estimated assuming log normality, but this was severely rejected for some of the stochastic terms. Thus, I turned to a Box-Cox transformation. Strictly speaking, this Box-Cox transformation does not impose positivity on the share parameters. But, given my estimates of the Box-Cox parameters and the variances of the stochastic terms, negative outcomes would be extreme outliers.

[8] Note that we only have an equation for α_R^{Kd} in the case of $N_d > 0$, because if $N_d = 0$ we normalize $\alpha_R^{Kd} = 1$ and if $N_f = 0$ we normalize $\alpha_R^{Kf} = 1$. Thus, for illustration, the equation for α_R^{Kd} is just:

$$\frac{\left(\alpha_R^{Kd}\right)^{bc(3)}-1}{bc(3)} = a_0^{Kd} + a_{time}^{Kd}\cdot t + SC_d\cdot\varepsilon^{Kd}$$

Turning to the correlations of the ε, I specify that:

$$(14) \quad \left(\varepsilon^{Ld} \quad \varepsilon^{Md} \quad \varepsilon^{Nd}\right)' \sim N\left(0, \Sigma^{d}\right),$$

where Σ^d is unrestricted. Similarly, for affiliates, Σ^f is unrestricted. But, in order to conserve on parameters, I do not allow for covariances between the parent and affiliate share parameters.[9]

Finally, consider the TFP parameters H_d and H_f in equations (1) and (2). Since we do not observe output prices and quantities separately, we cannot identify the scale of the H (either absolutely or for the affiliate relative to the parent). However, we can identify technical progress. Thus I normalize $H_d = H_f = 1$ at $t=0$ (1983) and let each have a time trend:

$$(15) \quad H_{jt} = (1 + h_j)^t \qquad for\ j=d,f.$$

A specification with equal time trends could not be rejected, so I set $h_d = h_f = h$.

III.2. Demand Function Parameters

Now I turn to the stochastic specification for the demand function parameters. For the inverse price elasticity of demand, or market power, parameter g_1 we have:

$$(16) \quad \frac{g_1^{bc(7)} - 1}{bc(7)} = g_{10} + g_{1,time} \cdot t + g_{1,shift} \cdot I[Nd > 0] + \varepsilon^{g_1}$$

A similar equation holds for g_2, and the Box–Cox parameter in that equation is $bc(8)$.

For the demand function intercepts for good 1 in the domestic market, I specify:[10]

$$(17) \quad \frac{(P_{0d}^1)^{bc(9)} - 1}{bc(9)} = P_{0d,0}^1 + P_{0d,time}^1 \cdot t + P_{0d,shift}^1 \cdot I[Nd > 0] + \varepsilon^{P_{0d}^1}$$

Similar equations hold for P_{0f}^2, P_{0f}^1 and P_{0d}^2, and the Box-Cox parameters in these equations are denoted by $bc(10)$, $bc(11)$ and $bc(12)$, respectively.

Preliminary results suggested that cross correlations between the three groups of parameters (technology, price elasticities, and demand function intercepts), were not important. Allowing for such correlations leads to a severe proliferation of parameters. Thus, I assume ε^{g_1} and ε^{g_2} are independent of other stochastic terms. I let the $(\varepsilon^{P_{0d}^1}, \varepsilon^{P_{0f}^2}, \varepsilon^{P_{0f}^1}, \varepsilon^{P_{0d}^2})$ vector be correlated within itself with covariance matrix Σ^P, but it is independent of the other stochastic terms.

[9] Interpretation of the Σ^d and Σ^f terms is rather subtle. The logistic transformation already incorporates the negative correlation among the share parameters that is generated by the CRTS assumption. If $\Sigma^d = I$, we have an "IIA" setup, where if one domestic share parameter increases, the other share parameters decrease proportionately. The correlations in Σ^d and Σ^f allow firms to depart from this IIA situation. For example, if Σ^d_{12} is very large, then we get a pattern where firms with large domestic labor shares also have large domestic materials shares.

[10] Technically, we should impose that g_1 and g_2 are positive and less than 1, and that the P_o terms are positive. Equations (16)-(17) do not impose these constraints. But, given the estimates, violations would be extreme outliers.

III.3 Labor Force Adjustment Cost Parameters

Recall that labor force adjustment costs are given by equation (5). The parameters δ_d and δ_f are allowed to vary across firms as follows:

$$(18) \qquad \delta_{dt} = exp\{\delta_{d1,0} + \delta_d w_{dt} + \delta_d \cdot t + \delta_d \cdot I[N_{dt} > 0]\}$$
$$\delta_{ft} = exp\{\delta_{f1,0} + \delta_f w_{ft} + \delta_f \cdot t + \delta_f \cdot I[N_{ft} > 0]\}$$

As with the other structural parameters, I allow for the possibility that adjustment costs vary over time, and between firms that do and do not have intra-firm flows. I also allow the δ to be functions of the wage rate, since search and severance costs for high skilled labor are higher.

III.4. Serial Correlation

I model serial correlation of the errors for each firm using a random effects structure. For example, for the stochastic part of the parent labor share parameter ε^{Ld} we have:

$$\varepsilon^{Ld}(it) = \mu^{Ld}(i) + \nu^{Ld}(it) \qquad \text{for } t=1,\ldots T_i$$

and similarly for the other eleven parameters. Let $\mu_i \sim N(0, \Sigma_\mu)$ denote the 12×1 vector of random effects for firm i, and let $\nu_{it} \sim N(0, \Sigma_\nu)$ denote the 12×1 vector of firm/time specific error components. Then, $V_t \equiv Var(\varepsilon_{it}) = \Sigma_\mu + \Sigma_\nu$ and $C_{t-j,\,t} \equiv Cov(\varepsilon_{it}, \varepsilon_{i,t-j}) = \Sigma_\mu$. Note that Σ_μ and Σ_ν each contain 78 unique elements, but these are restricted as described earlier. For instance, the non-zero elements of $\Sigma_\mu + \Sigma_\nu$ consist entirely of elements of Σ^d, Σ^f and Σ^P along with σ_{g1}^2 and σ_{g2}^2 . Other cross-correlations are set to zero. Defining $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT_i})$ we have:

$$(19) \qquad Var(\varepsilon_i) = \begin{pmatrix} V_1 & & & \\ C_{12} & V_2 & & \\ \vdots & & \ddots & \\ C_{1T_i} & \cdots & \cdots & V_{T_i} \end{pmatrix}$$

So far, I have only considered the most general case where a firm has all 4 potential trade flows. If a firm has $N_{dt}=0$ (or $N_{ft}=0$) at time t, then there is no value for $\varepsilon^{Kd}(it)$ (or $\varepsilon^{Kf}(it)$). Similarly, if $E_t=0$, there is no value for $P_{0f}^1(it)$, and if $I_t=0$ there is no value for $P_{0d}^2(it)$. Additionally, some firms are not observed for consecutive years. In such cases $Var(\varepsilon_i)$ is collapsed in the obvious way (by removing the relevant rows and columns).

III.5. Unobserved Expectation Terms

I deal with the unobserved expectation term $E(FD)L_d$ in (9) by invoking a rational expectations assumption:

$$(20) \qquad E_t(FD_{it})L_{dit} = FD_{it}L_{dit} - \eta_{it}^d$$

where η_{it}^d is a forecast error assumed orthogonal to all information available at time t. A similar expression is specified for $E(FF)L_f$. The distributions of η_{it}^d and η_{it}^f are discussed Section IV.

IV. Estimation using the SML Based on FOCs Approach

IV. 1. Motivation

If we substituting (20) into (9), we see that the FOC for domestic labor will contain five firm specific stochastic terms (α_{it}^{Ld}, g_{1it}, P_{0d}^{1}, P_{0f}^{1} and η_{it}^{d}) that enter non-linearly. Hence, it is not possible to express the equation as a moment condition with a single additive error term.[11] The same is true of all 12 FOCs of the model, since they all involve multiple stochastic terms that enter nonlinearly. Thus, GMM estimation is infeasible.

One alternative to GMM would be a full information ML approach (FIML). This would require the econometrician to specify how firms form expectations of future labor inputs. This in turn would require that one specify how firms forecast future demand and technology shocks, tariffs, exchange rates, etc. This means completely specifying the stochastic processes for all these forcing variables. Thus, a FIML approach would force a researcher to make assumptions about a wide range of issues that go well beyond the nature of the firm's technology and the structure of demand.

The SML based on FOCs approach is another alternative. The fact that multiple stochastic terms enter the FOCs in (8)-(9) in a highly nonlinear way creates no problems for this approach. SML based on FOCs can be thought of as a compromise between FIML and GMM. As in a FIML approach, the econometrician must specify parametric distributions for the demand and technology shocks. But, rather than specify stochastic processes for all the forcing variables (e.g., tariffs, wages, etc.), one simply substitutes realizations of the $t+1$ labor demand terms for their expectations, as in a typical GMM approach. However, the SML based on FOCs approach requires that we specify a distribution for the forecast errors η_{it}^{d} and η_{it}^{f}.

In terms of what one can do with the model once it is estimated, SML based on FOCs also represents a compromise between FIML and GMM approach. Since we estimate the complete distribution of technology and demand parameters for the MNCs, we can do *steady state* simulations of the responses of the whole population of firms to changes in the tariffs and other features of the environment.[12] But, since we do not model the evolution over time of all the forcing processes, we cannot simulate transition paths to a new steady state.[13]

[11] Even if we could linearize (9), finding valid instruments is difficult. Usual candidates like input prices would be correlated with firm specific technology parameters if technology changes over time in response to price changes.
[12] Note that, even if GMM estimation were feasible for our model, it would not be adequate for this purpose. The usual argument for GMM over ML is that one avoids making distributional assumptions on the stochastic terms, and thereby obtains more robust estimates of model parameters. But we must estimate the distributions of the firm specific parameters if we want to simulate the response of the population of firms to changes in the environment.
[13] It is worth emphasizing that the key difficulty in estimation arises not from dynamics, but rather because multiple stochastic terms enter the FOCs. This problem would be present in a static model without labor adjustment costs.

IV.2 The Stochastic Process for Forecast Errors

Without loss of generality I rewrite (20), the equation for forecast errors, as follows:

(21) $E(FD_{it})L_{dit} = FD_{it}L_{dit} + \sigma_{dt}\eta^*_{dt}$

where η^*_{dt} is standard normal. Of course, there is no reason one could not specify a more flexible parametric distribution. For instance, Feinberg and Keane (2003a) estimated a generalized version of this model where η^*_{dt} was assumed to be normal subject to a Box-Cox transform. But the estimated Box-Cox parameter was extremely close to one, implying that the distributions of the forecast errors are indeed well described by normality in this model.

It is plausible that that σ_{dt}, the standard deviation of the labor adjustment cost forecast error, will be increasing in L_{dit}, so I write:

(22) $\sigma_{dt} = exp\{\tau_{d0} + \tau_1 L_{dt}\}, \qquad \sigma_{ft} = exp\{\tau_{f0} + \tau_1 L_{ft}\}$

Regarding serial correlation, I assume that the forecast errors are independent over time, as implied by rational expectations. But I allow parent and affiliate forecast errors to be correlated within a period, as must be the case if their production processes are integrated, or if they face common shocks. Thus, I let $(\eta^d\ \eta^f)' \sim N(0\ ,\ \Sigma_\eta)$. I also let $\Sigma_\eta = CC'$, where C is the lower triangular Cholesky decomposition $\begin{pmatrix} C_{11} & 0 \\ C_{12} & C_{22} \end{pmatrix}$. Finally, let $\tau = (\tau_{d0},\ \tau_{f0},\ \tau_1)$.

IV.3. Construction of the Simulated Likelihood Function

IV.3.A. Overview

Let θ denote the vector of all model parameters. It includes values for the common (or non-stochastic) parameters of the model, which are R, β, δ_d, δ_f, μ, Δ, τ and h, as well as the parameters of the joint distribution of the 12 firm specific stochastic terms (see section III). Given a value of θ, simulation of a firm's likelihood contribution involves the following steps:

Step 1: Take a draw from the joint distribution of the forecast errors, η^d and η^f.

Step 2: Use the ten first FOCs for L_d, K_d, M_d, N_d, L_f, K_f, M_f, N_f, E, and I (see eqns. 8-9), and the production functions (1) and (2) as a system of 12 equations to solve for the 12 stochastic terms that rationalize the firm's behavior in each time period.

Step 3: Calculate the joint density of the stochastic terms, using the multivariate normal distribution with covariance matrix given by (19).

Step 4: Multiply by the Jacobian to obtain the data density.

Repeating this process at independent draws for η^d and η^f, and averaging the data densities so obtained, we obtain a simulation consistent estimate of the likelihood contribution for the firm.

While this process is conceptually straightforward, some aspects of the computation require new simulation methods that will be developed below. In step 2, solving the system of 12 nonlinear equations for the 12 stochastic terms in the model is cumbersome. But it is not computationally difficult. However, there are some draws for η^d and η^f in step 1 such that the system in step 2 has no solution. That is, there are regions of the space of forecast errors such that firm behavior is not rationalizable, and the boundaries of these regions depend on θ.

Thus, in order to simulate the likelihood, it is not advisable to naively take draws from the unconditional distribution of η^d and η^f in step 1. We have a model where all the stochastic terms are continuous, yet they must fall in certain sub-regions of a high dimensional space in order for firm behavior to be rationalizable by the model. As a result, a high dimensional integral must be evaluated to construct the joint density of a firm's stochastic terms. Section IV.3.E presents a recursive importance sampling algorithm that can deal effectively with this integration problem. The algorithm is the discrete/continuous analogue of the GHK algorithm.

The second problem is that the Jacobian required for step 4 is analytically intractable. But in sections IV.3.C and IV.3.E I show how it can be approximated using simulation methods and numerical derivatives.

IV.3.B. Solving for the Error Terms (and Identification of the Model Parameters)

Solving the system of 12 nonlinear equations for the 12 stochastic parameters in the model is cumbersome. I relegate the details to Appendix 1, but here I give an overview of the process. Understanding this process is important both for understanding the simulation methods that are developed later, as well as for understanding how the model parameters are identified. In this section, assume for the moment that we have already obtained a valid draw for the forecast errors, η^d and η^f, and that we seek to solve for the remaining 12 stochastic terms of the model.

The market power parameters g_1 and g_2 are identified by simple markup relationships (i.e., Lerner conditions), which we can construct using data on sales revenues and costs. These relationships are modified slightly to account for labor force adjustment costs, and also for the complication that the MNC has an incentive to hold down prices of final goods it ships intra-firm as intermediates (in order to avoid tariff costs). (See equation A1.11). The U.S./Canadian price ratios for goods 1 and 2, denoted PR_1 and PR_2, are determined by the tariff and transport cost wedge, again modified by the incentive to hold down prices of intra-firm intermediates. Since the strength of this incentive depends only on g_1 and g_2, we can solve for PR_1 and PR_2 once g_1 and g_2 are obtained (see equation A1.12). Also, given g_1 and g_2, the Cobb-Douglas share parameters are identified by cost shares of modified revenues (see equation A1.13).

Finally, given g_1 and PR_1, we can infer the *ratio* of the U.S. to Canadian demand function intercepts for good 1 (i.e., P_{0d}^1/P_{0f}^1) by observing the ratio of U.S. to Canadian sales for good 1.

Similarly, by comparing affiliate sales in Canada vs. imports to the U.S., we can infer the *ratio* of domestic to foreign demand function intercepts for good 2 (i.e., P_{0f}^2/P_{0d}^2).

Thus, without separate data on prices and quantities (except wages and employment, which we need only to identify the labor force adjustment cost function), we can identify the market power parameters, the Cobb-Douglas share parameters, and the ratios of the demand function intercepts for goods 1 and 2. To identify the *levels* of the demand intercepts, we need capital and materials price indices. Then, we can construct real capital and materials inputs, and use the production functions (1)-(2) as additional equations to determine quantities of output.

The preceding discussion assumes that we are solving for the stochastic terms at a given value of θ, which means at a given value of the profit rate R. Knowledge of R enables us to construct total costs, which enables us in turn to construct g_1 and g_2 using markups inferred from revenue and cost data. A key issue is how R is identified. As I show in Appendix 2, the model implies a relation $g/(1\text{-}g) = \alpha^K \cdot R$ between market power and capital share. Thus, if the profit rate is low (high), there is a strong (weak) tendency for firms with more market power to also have larger capital shares, so that profits accrue to a larger (smaller) stock of capital. In other words, the larger is R, the greater the extent to which firms with larger capital shares also have larger markups (i.e., face more inelastic demand).

Thus, to the extent that firms with larger capital shares act as if they face less elastic demand (in terms of how they respond to changes in the forcing variables), we will infer a higher value of R. At first this may seem like a strange argument for identification, since the capital share is not observed. However, given the assumption that R is equal for all firms, the capital share is perfectly negatively correlated with the sum of the labor, materials and intermediate shares. Thus, the greater the extent to which firms with larger labor plus materials plus intermediate shares act as if they face more elastic demand, the larger the implied R.[14]

Having solved for all the firm specific parameters, we use the equations of the stochastic specification (see section III) to construct the vector of error terms for the firm. Let ε_{it} denote the vector of (up to) 12 error terms for firm i in period t (or, as few as 8 if $N_d = N_f = E = I = 0$):

$$\varepsilon_{it} \equiv (\ \varepsilon_{it}^{Ld},\ \varepsilon_{it}^{Md},\ \varepsilon_{it}^{Nd},\ \varepsilon_{it}^{Lf},\ \varepsilon_{it}^{Mf},\ \varepsilon_{it}^{Nf},\ \varepsilon_{it}^{g1},\ \varepsilon_{it}^{g2},\ \varepsilon_{it}^{P_{0d}^1}, \varepsilon_{it}^{P_{0f}^2}, \varepsilon_{it}^{P_{0f}^1}, \varepsilon_{it}^{P_{0d}^2}\ )$$

Finally, I construct the joint multivariate normal density of the error vector for firm i, which I denote by $\varepsilon_i \equiv (\varepsilon_{i1}, \ldots, \varepsilon_{iT(i)})$, where $T(i)$ is the number of time periods that firm i is observed, using the covariance structure given by equation (19).

[14] Appendix 2 also gives an intuitive explanation of how the time trends in TFP (h) and in the demand function intercepts (the P_0) are separately identified. Briefly, to the extent that growth is more than proportionately slower for firms with more market power, it implies that growth is induced by TFP rather than growth in demand.

IV.3.C. The Jacobian of the Transformation from the Error to the Data Density

Of course, the likelihood is the joint density of the data, not of the stochastic terms. I now turn to the construction of the Jacobian. If y_i denotes the vector of data elements for firm i, then $f(y_i) = |\partial\varepsilon_i/\partial y_i|\ f(\varepsilon_i)$, where $\partial\varepsilon_i/\partial y_i$ is the Jacobian of the transformation from the data to the stochastic terms. In the present case, the 12 data items observed for the firm (or as few as 8 if $N_d = N_f = E = I = 0$) at time t are:

$$y_{it} = \left\{P_d^1 S_d\,, P_f^2 S_f\,, L_d\,, L_f\,, w_d L_d\,, w_f L_f\,, P_f^2 N_d\,, P_d^1 N_f\,, P_f^1 E, P_d^2 I, \phi_d M_d\,, \phi_f M_f\right\}$$

Observe that the Jacobian is not block diagonal by period t, because ε_{it} is affected by $L_{d,t-1}$, $L_{f,t-1}$, $L_{d,t+1}$ and $L_{f,t+1}$. It is not possible (as far as I can determine) to obtain an analytic expression for the Jacobian, because the mapping from the data to the stochastic terms is so highly nonlinear. Furthermore, the mapping depends on the values of the forecast errors (which we condition on here, but which must be integrated out). Therefore, I construct the Jacobian numerically.

To calculate the numerical Jacobian, I bump the elements of the data vector (y) one at a time. When a data element is bumped, I recalculate all the elements of ε_i, and form numerical derivatives of ε_i with respect to that element of y. I then use these numerical derivatives to fill in the column of the Jacobian that corresponds to the bumped element of y. We have:

$$J(\eta_i, y_i) = \begin{pmatrix} \partial\varepsilon_{i1,1}/\partial y_{i1,1} & \partial\varepsilon_{i1,1}/\partial y_{i1,2} & \cdots & \partial\varepsilon_{i1,1}/\partial y_{iT_i,12} \\ \partial\varepsilon_{i1,2}/\partial y_{i1,1} & . & & . \\ . & & & . \\ \partial\varepsilon_{iT_i,12}/\partial y_{i1,1} & \cdots & \cdots & \partial\varepsilon_{iT_i,12}/\partial y_{iT_i,12} \end{pmatrix}$$

where y_{itk} denotes the kth element of the data vector for firm i in year t. For instance, if we bump $y_{i11} = P_d^1 S_{d(i1)}$, and form numerical derivatives of the ε_i elements, we obtain the first column of the Jacobian. I denote the Jacobian by $J(\eta_i, y_i)$ to highlight that it depends on the forecast errors.

Since the Jacobian depends on η_i, it must be simulated. However, to form the likelihood, it is not correct to simulate the Jacobian separately and then multiply by the error density. Rather, the product of the Jacobian and data density must be averaged over draws for η_i, as I discuss below. I will not comment further on the fact that the simulated Jacobian is based on numerical derivatives. Numerical procedures are generally used to construct even "exact" likelihood functions (e.g., exponentials are calculated via series expansions), so I don't feel that the use of a numerical procedure here warrants special comment.

IV.3.D. Simulating the Likelihood Function: Naïve Approach

I form the likelihood by integrating the data density over the forecast error distribution:

$$\mathscr{L}(\theta) = \prod_{i=1}^{N} \int_{\eta_i} f(y_i | \eta_i, \theta)\ f(\eta_i)\, d\eta_i$$

$$= \prod_{i=1}^{N} \int_{\eta_i} |J(\eta_i, y_i)|\ f(\varepsilon_i(\eta_i, y_i)|\theta)\ f(\eta_i, y_i)\, d\eta_i$$

Here, $\eta_i \equiv (\eta_{i1}^d, \eta_{i1}^f \ldots\ldots, \eta_{iT_i}^d, \eta_{iT_i}^f)$, $f(\eta_i)$ denotes the joint density of forecast errors for firm i, and θ denotes the vector of model parameters. The notation $\varepsilon_i(\eta_i, y_i)$ emphasizes that the mapping from the data to the stochastic terms for a firm depends on the vector of forecast errors η_i.

Naively, one could simulate the likelihood function by taking *iid* draws from the distribution of η_i. An important complication arises here, however. Firm behavior cannot be rationalized by any arbitrary values for the forecast errors. Conditional on a particular draw for (η^d, η^f), firm behavior is only rationalizable if, when we solve the mapping from the data to the model parameters, we obtain $1>g_1>0$, $1>g_2>0$, all technology parameters positive, and all demand shift parameters positive. But this will not be the case for all possible (η^d, η^f) draws.

In Appendix 1, I derive the following equation relating g_1 and the forecasts $E(FD)L_d$:

$$(23)\quad g_1[AYE - \frac{AB(P_d^1 N_f)(P_f^2 N_d)}{BVI}] = RD - CD - \delta_d E(FD)L_d + \frac{(RF - CF - \delta_f E(FF)L_f)}{BVI}(P_f^2 N_d)B$$

In the data, the term $AYE - AB(P_d^1 N_f)(P_f^2 N_d)/BVI$ is always positive.[15] So, the right-hand side of (23) must be positive in order for g_1 to be positive. But if the forecasted adjustment costs $E(FD) = FD - \eta^d$ or $E(FF) = FF - \eta^f$ are too large, the right-hand side will be driven negative. Thus, observed firm behavior implies bounds on the forecast errors. I derive the exact bounds on (η^d, η^f) in Appendix 3.

For any draw (η^d, η^f) that does not satisfy the bounds, no values of the firm specific parameters can rationalize firm behavior. Let BD_i denote the region of the space of forecast errors such that behavior of firm i is rationalizable. Then, we can rewrite the likelihood as:

$$\mathscr{L}(\theta) = \prod_{i=1}^{N} \int_{\eta_{i1} \in BD_{i1}} \cdots\cdots\cdots \int_{\eta_{iT_i} \in BD_{iT_i}} |J(\eta_i, y_i)|\quad f(\varepsilon_i(\eta_i, y_i)|\theta)\quad f(\eta_{iT_i}) \ldots f(\eta_{i1})\, d\eta_{iT_i} \ldots d\eta_{i1}$$

[15] A and B are usually close to 1. Therefore, AYE is close to sales of good 1, and BVI is close to sales of good 2 (see Appendix 1 for definitions). Thus, the second term includes $(P_f^2 N_d / BVI)$, a number < 1, times $P_d^1 N_f$, which is just one component of sales of good 1. So the whole term in brackets is roughly sales of good 1 minus a fraction thereof.

$$= \prod_{i=1}^{N} \int_{\eta_{i1}} \ldots \int_{\eta_{iT_i}} \left|J(\eta_i, y_i)\right| \; f(\varepsilon_i(\eta_i, y_i)|\theta) \; I[\eta_{iT_i} \in BD_{iT_i}] \ldots\ldots I[\eta_{i1} \in BD_{i1}] \; f(\eta_{iT_i}) \ldots\ldots f(\eta_{i1}) \, d\eta_{iT_i} \ldots d\eta_{i1}$$

where $I[\eta_{it} \in BD_{it}]$ is an event indicator. Then, given M *iid* draws from the unconditional distribution of the forecast errors, the likelihood contribution for firm i can be approximated using a simple frequency simulator as follows:

$$(24) \qquad \hat{\mathscr{L}}_i(\theta) = M^{-1} \sum_{m=1}^{M} \left|J(\eta_i^m, y_i)\right| \; f(\varepsilon_i(\eta_i^m, y_i)|\theta) \; I[\eta_{i1}^m \in BD_{i1}] \ldots I[\eta_{iT_i}^m \in BD_{iT_i}]$$

There are two fundamental problems with this approach, however:

1) For some draws m, the likelihood contribution is zero. For example, suppose $P(\eta_{it} \in BD_{it}) = .95 \;\; \forall t$ and there are 11 periods. Then only 56% of draws would belong to BD_i on average. This creates a serious numerical inefficiency, since many draws are useless for evaluating the $|J(\eta_i, y_i)|\, f(\varepsilon_i(\eta_i, y_i)|\theta)$ term. They are only used (implicitly) to evaluate event probabilities $P(\eta_i \in BD_i)$, which could be much more accurately evaluated by other means (see below).
2) The simulated likelihood in (24) is not a smooth function of the model parameters θ. It takes discrete jumps at θ values such that one of the draws η_i^m is exactly on the boundary of BD_i. This means that gradient-based search algorithms cannot be used to maximize the likelihood function, and derivatives are not available to calculate standard errors of parameter estimates. Without derivatives, estimation is practically impossible for a model with many parameters

IV.3.E. A Recursive Importance Sampling Approach

In this section I present a more efficient smooth simulator of the likelihood using a recursive importance sampling algorithm that is a discrete/continuous data analogue of the GHK algorithm for simulating event probabilities in discrete choice models. As in GHK, the idea is to draw the η_i from the "wrong" density, chosen so that <u>all</u> η_i with positive mass under this density are consistent with firm behavior. The likelihood is then simulated using a weighted average over these draws, where the weights are ratios of the draw's likelihood under the correct density $f(\eta_i)$ to its likelihood under the incorrect density $f(\eta_i | \eta_i \in BD_i)$.

Consider first the simulation for firm i in period t. Define $\eta_t \equiv (\eta^d_t, \eta^f_t)$. From (A3.11), the constraints on η_t are (suppressing firm subscripts):

$$BL_t^d < \eta_{dt} < BU_t^d \qquad\qquad BL_t^f < \eta_{ft} < BU_t^f(\eta_{dt})$$

Let:

$$\begin{pmatrix} \eta_{dt} \\ \eta_{ft} \end{pmatrix} = \begin{pmatrix} \sigma_{dt}\eta_{dt}^* \\ \sigma_{ft}\eta_{ft}^* \end{pmatrix}$$

where σ_{dt} and σ_{ft} are the standard deviations of η_{dt} and η_{ft}, and η^*_{dt} and η^*_{ft} are standard normal. Recall that I allow the forecast errors to be correlated across the parent and affiliate within a period. This makes sense because, given the structure of the model, any shock that affects $L_{d,t+1}$ would also potentially affect $L_{f,t+1}$. Then, employing a lower triangular Cholesky decomposition, we have:

$$\begin{pmatrix} \eta^*_{dt} \\ \eta^*_{ft} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} \zeta_d \\ \zeta_f \end{pmatrix}$$

where ζ_d and ζ_f are iid N(0,1) and $a_{12}^2 + a_{22}^2 = 1$. Let $F(\cdot)$ denote the standard normal distribution function. The simulation algorithm proceeds as follows:

Step 1: Calculate $F\left(\frac{BL_t^d}{\sigma_{dt}}\right)$ and $F\left(\frac{BU_t^d}{\sigma_{dt}}\right)$

Step 2: Draw a uniform random variable on [0,1]. Denote it by u_{1t}^m.

Step 3: Construct $u_1^* = F\left(\frac{BL_t^d}{\sigma_{dt}}\right) + u_{1t}^m \cdot \left[F\left(\frac{BU_t^d}{\sigma_{dt}}\right) - F\left(\frac{BL_t^d}{\sigma_{dt}}\right)\right]$

Step 4: Construct $\zeta_d^m = F^{-1}(u_1^*)$ and $\eta_{dt}^m = \sigma_{dt}\zeta_d^m$

Step 5: Calculate the bounds on ζ_f, given ζ_d^m. These are:

$$BL_t^f < \sigma_{ft}\left[a_{12}\,\zeta_d^m + a_{22}\zeta_f\right] < BU_t^f(\eta_{dt}^m)$$

Or, rearranging terms:

$$\frac{1}{a_{22}}\left[\frac{BL_t^f}{\sigma_{ft}} - a_{12}\,\zeta_d^m\right] < \zeta_f < \frac{1}{a_{22}}\left[\frac{BU_t^f}{\sigma_{ft}} - a_{12}\,\zeta_d^m\right]$$

We write these bounds more compactly as $BL(\zeta_d^m) < \zeta_f < BU(\zeta_d^m)$.

Step 6: Draw a uniform random variable on [0,1]. Denote it by u_{2t}^m.

Step 7: Construct $u_2^* = F(BL(\zeta_d^m)) + u_{2t}^m \cdot \{F[BU(\zeta_d^m)] - F[BL(\zeta_d^m)]\}$

Step 8: Construct $\zeta_f^m = F^{-1}(u_2^*)$ and $\eta_{ft}^m = [a_{12}\,\zeta_d^m + a_{22}\zeta_f^m] \cdot \sigma_{ft}$

Repeat Steps 1 through 8 for $m=1...M$ draws and $t=1...T$ periods. Save the following:

1) The uniform draws (u_{1t}^m, u_{2t}^m) $m=1, M$; $t = 1, T_i$

2) The forecast error draws $(\eta_{dt}^m, \eta_{ft}^m)$ $m=1, M$; $t = 1, T_i$

3) The event probabilities:

$$P(\eta_{dt} \in BD_t) \equiv P\left(BL_t^d < \eta_{dt} < BU_t^d\right) = F\left[\frac{BU_t^d}{\sigma_{dt}}\right] - F\left[\frac{BL_t^d}{\sigma_{dt}}\right] \qquad \text{for } t=1,\ldots,T_i$$

$$P\left(\eta_{ft} \in BD_t \mid \eta_{dt}^m\right) \equiv F(BU(\zeta_d^m)) - F(BL(\zeta_d^m)) \qquad \text{for } m=1, M;\quad t=1,\ldots, T_i$$

where BD_t denotes the region in which (η_t^d, η_t^f) must fall in order for firm behavior to be rationalizable in year t.

Finally, defining $\eta_i^m = (\eta_{d1}^m, \eta_{f1}^m, \ldots, \eta_{dT_i}^m, \eta_{fT_i}^m)$, the smooth unbiased simulator for the likelihood contribution of firm i is:

$$\hat{\mathscr{L}}_i(\theta) = M^{-1} \sum_{m=1}^{M} \left|J(\eta_i^m, y_i)\right| \, f(\varepsilon_i(\eta_i^m, y_i)|\theta) \;\; P(\eta_{d1} \in BD_1) P(\eta_{f1} \in BD_1 \mid \eta_{d1}^m) \cdot \ldots \cdot P(\eta_{dT_i} \in BD_{T_i}) P(\eta_{fT_i} \in BD_{T_i} \mid \eta_{dT_i}^m)$$

Since many economic models have a structure where certain stochastic terms must be in particular ranges in order for a continuous outcome to be observed (e.g., a productivity shock must be in a certain range in order for a firm to operate), the basic approach used here can be useful in many contexts. Since the simulator is recursive it can be extended trivially to accommodate serial correlation in the η_t. The bounds for η_2 would then be a function of η_1^m, and so on. I rule out such correlation here because of the forecast error interpretation of η.

V. Simulation from the Posterior Distribution of Model Parameters

In this section I present a recursive importance sampling algorithm for drawing from the posterior distribution of firm specific parameters (conditional on the firm's observed history). This is a non-trivial problem, for the following reason: The model contains a 14×1 vector of stochastic terms for each firm and time period (i.e., 12 firm specific parameters, and two forecast errors). Even though this vector is multivariate normal unconditionally, its posterior distribution conditional on firm behavior is very complex. This is because the 12×1 vector of data elements places complex constraints on the 14×1 vector of stochastic terms. Thus, I need a way to draw a K×1 vector of random variables that lie in a J×1 dimensional space, where J<K due to the constraints imposed by the data and model. I do this using a recursive algorithm that is a continuous data analogue to the GHK algorithm.

I want to construct the posterior distribution of the firm-specific parameters for two reasons. First, I want to test the model's distributional assumptions. I assumed that (after Box-Cox transformations) the stochastic parts of the firm specific parameters are jointly normal. If this assumption is valid, then we should not reject normality for their posteriors conditional on the data. Second, I want to simulate the response of the population of firms to changes in the environment, and for this I need draws from the posteriors of the firm specific parameters.

The distribution of firm specific parameters conditional on the data cannot be expressed analytically, because the mapping from the data to the stochastic terms is highly nonlinear. (Contrast this with a linear random effects model, $y_{it} = \alpha_i + x_i\beta + \varepsilon_{it}$ where the distribution of the random effect α_i for a firm i conditional on $\{y_{i1}, ..., y_{iT}\}$ is simple to construct). But here I develop an importance sampling algorithm that can be used to obtain draws from $f(\varepsilon_i, \eta_i \mid D_i)$ where D_i is the data observed for firm i. The basic principle of the algorithm is quite general, and it can be used to obtain draws from the posterior distribution of stochastic terms conditional on data in the general class of continuous nonlinear models where this distribution is not degenerate (e.g., random effects models, mixture models, factor models).

To describe the algorithm, it is useful to first explain implementation for a single time period (ignoring the panel aspect of the data). First I establish some definitions. Let D denote the vector of data elements for a particular firm, and let S denote the vector of stochastic terms for this firm (in the present case case D is 12x1 and $S = (\varepsilon, \eta)$ where ε is 12x1 and η is 2x1, so S is 14x1). Let M denote the mapping from S to D: $D=M(S)$. Note that M is not invertible, since the dimension of S exceeds that of D. Let $B(D)$ denote the set of S values that are consistent with D:

$$B(D) = \{ S \mid D = M(S) \}$$

If we partition S into sub-vectors $S = (S^1, S^2)$, then we can also define sets of values for the sub-vectors that are consistent with D:

$$B^1(D) = \{ S^1 \mid D = M(S^1,S^2) \text{ for some } S^2\} \quad = \quad \{ S^1 \mid \exists S^2 \text{ such that } D=M(S^1,S^2)\}$$

We can construct these sets recursively:

$$B^2 (D \mid S^1) = \{ S^2 \mid D = M(S^1 , S^2) \}$$

Use $P(X)$ to denote the measure of a set X and note that $P(B(D))=0$ under the unconditional distribution of S. The fact that $P(B(D))=0$ means that not only is acceptance/rejection sampling impractical, it is <u>impossible</u> in this case. A recursive algorithm is essential.

Since $D = M(S)$ we have $f(S , D) = f(S)$, and therefore:

$$f(S \mid D) = \begin{cases} f(S)/f(D) \text{ if } S \in B(D) \\ 0 \text{ if } S \notin B(D) \end{cases}$$

It is worth noting that:

$$\int_{S \in B(D)} \left[f(S)/f(D)\right] dS \;=\; f(D)^{-1} \int_{S \in B(D)} f(S)\; dS = 1$$

Thus, we have that $f(S \mid D) \propto f(S)$ *if* $S \in B(D)$. But this is not immediately useful, because we cannot draw directly from $f(S)$ subject to the constraint that $S \in B(D)$, since $P(B(D)) = 0$.

Feasible methods to obtain draws from $B(D)$ will generally involve drawing from an incorrect "source" density $\phi(S)$ chosen so that $P_\phi(B(D)) > 0$, or ideally, $P_\phi(B(D)) = 1$, and $\phi(S) > 0$ for all $S \in B(D)$ subject to $f(S) > 0$. Thus, the general problem is to simulate draws from the target density $f(S \mid D)$ given that we only have access to draws from the source density $\phi(S)$.

The solution is an importance sampling algorithm in which the draws $\{S_m\}_{m=1,\,M}$ from $\phi(S)$ are weighted using weights w_m such that $w_m \propto f(S_m)/\phi(S_m)$ *for* $S_m \in B(D)$ *and* $w_m = 0$ otherwise. Since weights must sum to 1, we have:

$$w_m \;=\; \begin{cases} Kf(S_m)/\phi(S_m) \text{ if } S_m \in B(D) \\ 0 \text{ otherwise} \end{cases} \quad \text{where } K = \sum_{\substack{m=1 \\ s.t.\, S_m \in B(D)}}^{M} f(S_m)/\phi(S_m).$$

Maximal efficiency of an importance sampling algorithm is achieved by choosing a $\phi(S)$ as "close" as possible to $f(S \mid D)$, e.g., using Kullback–Liebler distance. Note that if $\phi(S)$ is very close to $f(S \mid D)$, then the w_m will be close to $1/M$.

I now describe the application of this algorithm to the model of this paper. Consider a single period of data for a firm. The data vector D is 12x1. The vector of stochastic terms S is 14x1. It consists of ε which is 12x1, and η which is 2x1. We have that $D = M(\varepsilon,\, \eta)$ and consider a partition:

$$\begin{aligned} B(D) &= \{\, (\varepsilon,\, \eta) \mid D = M(\varepsilon,\, \eta) \,\} \\ B^1(D) &= \{\, \eta \mid \exists\, \varepsilon \text{ s.t. } D = M(\varepsilon,\, \eta) \,\} \\ B^2(D \mid \eta) &= \{\varepsilon \mid D = M(\varepsilon,\, \eta) \,\} \end{aligned}$$

Given our model and distributional assumptions, we have seen that $P(B(D)) = 0$, $0 < P(B^1(D)) < 1$, and $P(B^2(D \mid \eta)) = 0$. It is not feasible to draw directly from the joint distribution $f(\varepsilon,\, \eta \mid D)$. Rather, we draw sequentially, first obtaining $\eta_m \in B^1(D)$ and then constructing the implied ε_m using the relation $D = M(\varepsilon_m,\, \eta_m)$, which is given by the nonlinear system of equations described in Appendix 1. Given this procedure, we obtain the source density:

$$\phi(\varepsilon,\, \eta) = \begin{cases} g(\eta) \text{ if } \eta \in B^1(D) \\ 0 \text{ otherwise} \end{cases}$$

where $g(\eta)$ is the density from which we draw the η_m.

I emphasize that g(η) is not the "correct" density of η, given by $f(\eta \mid \eta \in B^1(D))$. Rather, $g(\eta)$ is the density induced by the procedure of drawing $\eta = (\eta^d, \eta^f)$ sequentially, rather than drawing directly from the correct joint density $f(\eta^d, \eta^f \mid \eta \in B^1(D))$. Specifically, we sequentially partition the set $B^1(D)$ as follows:

$$
\begin{aligned}
B^{1d}(D) &= \{ \eta^d \mid \exists \eta_f \ s.t.\ (\eta^d, \eta^f) \in B^1(D) \} \\
B^{1f}(D, \eta^d) &= \{ \eta_f \mid (\eta^d, \eta^f) \in B^1(D) \}
\end{aligned}
$$

We first draw $\eta_d \in B^{1d}(D)$ and then, conditional on the draw $\eta_{d,m}$, we draw $\eta_{f,m} \in B^{1f}(D, \eta_{d,m})$.

Given this sequential procedure, we have that g(η) is given by:

$$
g(\eta_m^d, \eta_m^f) = \frac{f(\eta_m^d)}{\int\limits_{\eta^d \in B^{1d}(D)} f(\eta^d)\, d(\eta^d)} \quad \frac{f(\eta_m^f)}{\int\limits_{\eta^f \in B^{1f}(D, \eta_m^d)} f(\eta^f)\, d(\eta^f)} = f(\eta_m^d)\ f(\eta_m^f) / P(B^{1d}(D))\ P(B^{1f}(D, \eta_m^d))
$$

Thus we have that our source density is:

$$
\phi(\varepsilon_m, \eta_m) = \begin{cases} f(\eta_m^d)\ f(\eta_m^f) \big/ P(B^{1d}(D))\ \ P(B^{1f}(D, \eta_m^d)) \text{ if } \eta_m \in B^1(D) \\ 0 \text{ otherwise} \end{cases}
$$

Therefore, the importance sampling weights $w_m = f(\varepsilon_m, \eta_m)/\phi(\varepsilon_m, \eta_m)$ are:

$$
w_m = \begin{cases} K \ \ f(\varepsilon_m) P(B^{1d}(D))\ \ P(B^{1f}(D, \eta_m^d)) \ \text{ if } (\eta_m, \varepsilon_m) \in B^1(D) \\ 0 \text{ otherwise} \end{cases}
$$

$$
\text{where: } K = \left[\sum_{m=1}^{M} f(\varepsilon_m) P(B^{1d}(D))\ \ P(B^{1f}(D, \eta_m^d)) \right]^{-1}
$$

Since we have chosen the source density $\phi(\cdot)$ so that all draws $(\varepsilon_m, \eta_m) \in B(D)$, and since $P(B^{1d}(D))$ appears in the numerator and denominator of w_m, we can simplify to:

$$
w_m = f(\varepsilon_m)\ P(B^{1f}(D, \eta_m^d)) \Big/ \sum_{m=1}^{M} P(B^{1f}(D, \eta_m^d)) \ .
$$

The sequential construction of the importance sampling weight for draw sequence m is exactly analogous to the construction of sequence weights in the GHK algorithm. The sequence that begins with η_m^d is given more (less) weight to the extent that η_m^d makes a valid draw for η^f more (less) likely. And a sequence that begins with (η_m^d, η_m^f) is given more (less) weight if the implied ε_m is more (less) likely. In fact, the algorithm described here is the continuous data analogue to the algorithm for constructing and weighting draw sequences that underlies the GHK algorithm.

Given the recursive structure of the sequence weights, extension to the multi-period case is obvious – but notationally burdensome – so I omit the specific expressions.

VI. Empirical Results

VI.A. Parameter Estimates for the Structural Model

The smooth SML algorithm described in section IV.3 was used to estimate the model of MNC described in sections II-III, using the BEA data on U.S. MNCs and their Canadian affiliates for the period 1983-1996 described in Feinberg and Keane (2003a). The recursive importance sampling algorithm described in section IV.3.E was implemented using 50 draws. The algorithm was numerically very well behaved and converged without difficulties in about 300 iterations (that required several minutes each on a Pentium II processor). Results were not significantly affected by use of alternative starting values, or changing the number of draws.

Table 1 reports the estimates of the structural model of MNCs' marginal production and trade decisions.[16] Feinberg and Keane (2003a) discuss these estimates in detail, so I will only highlight a few main results. The first panel of Table 1 reports estimates of parameters related to the labor share in the parent's Cobb-Douglas production technology. Recall that these parameters map into the share parameter itself through the transformation given by equations (10) and (13). The estimates in Table 1 are for the parameters in equations like (13).

The first term is the intercept (α_0^{Ld}). The second term, α_{shift}^{Ld}, is a shift parameter that allows the labor share to differ for the subset of firms with positive intra-firm flows (i.e., it multiplies $I[N_d>0]$). The third and fourth terms are time trends, which are relevant for parents that do and do not use intermediate inputs from the affiliate, respectively. Finally, the fifth term is the Box-Cox parameter, *bc(1)*. This captures departures of the stochastic term in the labor share equation from log normality.

The second and third panels of Table 1 report parameters relevant to the parent's materials and capital shares, respectively. Note that the capital share equation has fewer parameters. When a parent does not utilize intermediates from the affiliate, it has only three inputs, so the capital share is just $1-\alpha^{Ld}-\alpha^{Md}$. Thus, the capital share equation is only relevant for parents that do use intermediates from the affiliate, and so it does not include the shift parameter or the extra time trend that are included in the labor and material share equations. The fourth through sixth panels of Table 1 contain exactly the same types of parameters, but for the affiliate.

A key result is that the time trends on the share parameters are small and insignificant for parents and affiliates that do not use intermediates that are shipped intra-firm. That is, in Table 1,

[16] Recall that the structural model was actually estimated jointly with reduced form MNC decision rules for whether or not to engage in intra-firm and arms length trade. While we condition on the MNC configuration in estimating the model of marginal decisions, we did not wish to assume that configuration was exogenous. This could lead to bias in estimates of the structural model if the MNC configuration is influenced by firm specific unobservables. Feinberg and Keane (2003b) discuss the estimates of the reduced form decision rules. Most notably, these estimates imply that tariffs have a negligible influence of MNC's decisions about whether to engage in intra-firm and arms length trade. The adjustment to tariff changes appears to be primarily on the intensive rather than the extensive margin.

the terms $\alpha_{Time}{}^{Ld}\cdot t \cdot I[N_d=0]$, $\alpha_{Time}{}^{Md}\cdot t \cdot I[N_d=0]$, $\alpha_{Time}{}^{Lf}\cdot t \cdot I[N_f=0]$ and $\alpha_{Time}{}^{Mf}\cdot t \cdot I[N_f=0]$ are all insignificant and quantitatively small. Thus, the behavior of these parents and affiliates is well described by a CRTS Cobb-Douglas technology with fixed share parameters.

In contrast, for the subset of MNC parents that do utilize intermediates from affiliates, and affiliates that use intermediates from parents, the time trends for the share parameters are all highly significant and negative. Recall that these trends feed into logistic transformations like (10). Thus, the fact that the time trends are negative for labor, capital and materials means the share of the omitted category (intermediates) must be rising. Thus, conditional on MNCs having had positive intra-firm flows initially, the estimates imply that "technical change" was driving up the share of intermediates.

Another way to state this result is that the model cannot explain the observed increase in intra-firm trade in intermediates based on the changes in the exogenous forcing variables, such as tariffs, transport costs, exchange rates, wages, materials prices, etc.. Thus, the model attributes the increase in intra-firm trade to technological change in the form of trends in the Cobb-Douglas share parameters. Feinberg and Keane (2003a) discuss this result at length and argue that it would be robust to various changes in the structural model that has been estimated here. They also argue that these "technical change" estimates may be picking up recent advances in logistics that have made it easier to ship intermediates intra-firm in a way that allows fragmented production processes to operate smoothly.

The top four panels on the second page of Table 1 contain estimates of the labor cost adjustment cost function. In the top two panels note that time trends in the δ_d and δ_f equations (see equation 18) are both significant and positive, suggesting that labor force adjustment costs increased over time. The third panel contains estimates of the μ and Δ parameters in the generalized labor force adjustment cost function (equation 5). These estimates imply that adjustment costs are not well described by the common linear-quadratic in levels specification, since they depart substantially from μ=1 and Δ=0. The fact that μ<1 and Δ>0 implies that the cost of a given absolute change in labor force size is smaller to the extent that the change represents a smaller fraction of the initial labor force.

The fourth panel contains estimates of the parameters τ that determine the variance of labor adjustment forecast errors (see equation 22). Not surprisingly, forecast error variance is increasing in labor force size. Interestingly, the domestic forecast error variance is higher (for given labor force size), and domestic and foreign forecast errors are positively correlated.

The middle two panels on the second page of Table 1 contain estimates of the parameters that determine the (negative) inverse price elasticities of demand for the U.S. and Canadian produced good (g_1 and g_2). These are the parameters of equations like (16). Note that the Box-Cox parameter for g_1, which is .6084, implies that a transform close to the square root is needed to induce normality of the residuals, while the transform for g_2, which is .8424, is quite close to linear. Thus, the stochastic terms in the g equations depart dramatically from log normality.

The last four panels on page 2 of Table 1 contain estimates of the parameters that determine the four demand function intercept parameters (i.e. U.S. and Canadian demand for the U.S. and Canadian produced goods). These are the parameters of equations like (17). Note that all four demand function intercepts exhibit significant negative time trends, implying falling demand at any given price level. At the same time, the estimate of TFP growth is 4.5% per year. As I discuss in Appendix 2, as $g \rightarrow 1$ (less elastic demand) TFP growth has no effect on employment. Thus, the fact that TFP is increasing while demand (at any given price) is falling implies that growth in employment is very concentrated among firms facing elastic demand. In fact, employment must be shrinking among high market power firms (see the discussion of Table 5 in section VI.C for more on this point).

Finally, I turn the estimates of the covariance matrix of the firm specific stochastic terms. Each of the 12 technology and demand function parameters is allowed to be heterogeneous across firms and over time. Recall from Section III.4 that I specify a stochastic structure where each of these 12 parameters has a firm/time specific component that consists of a random effect, μ_i, and a transitory error, ν_{it}. Table 2 reports estimates of the correlation matrices of these stochastic terms. Panel 1 presents the cross-sectional correlation matrix of the composite error $\mu_i + \nu_{it}$. Although all correlations are highly significant, the smallest (about 0.4) are for the country specific demand shifters across the two different goods (i.e., between P_{0d}^{1} and P_{0d}^{2}, and between P_{0d}^{1} and P_{0f}^{2}). By contrast, the correlation between the U.S. and Canadian demand shifters for the domestically produced good (i.e., between P_{0d}^{1} and P_{0f}^{1}) is 0.97! That is, demand for the same good across the two countries is extremely highly correlated, while demand for the two different goods in the same country is not so highly correlated.

The second panel of Table 2 reports the correlations of the firm specific parameters at times t and $t+1$. The random effects structure implies equal correlations at all leads and lags, all the way out to $t+11$. As we would expect, all the firm specific parameters are highly serially correlated. The technology and demand function intercept parameters show more persistence (i.e., correlations in the 0.72 to 0.80 range) than do the market power parameters (i.e., correlations of 0.46 and 0.53).

VI.B. Fit of the Structural Model

The SML based on FOCs estimation method requires that the researcher specify a joint parametric distribution for the vector of stochastic terms in the model. Thus, it is important to have a method to test the validity of the assumed parametric distribution. In the illustrative model of MNC behavior that I have estimated in this paper, I assumed that the vector of 12 firm specific technology and demand parameters are jointly normally distributed, subject to a Box-Cox transformation (as described in section III). In order to investigate the validity of this assumption, I use the recursive importance sampling algorithm developed in Section V to draw from the posterior distribution of the stochastic terms for each firm, conditional on the estimated model and the firm's observed history. I then use these draws to test the joint normality assumption.

Specifically, I use the algorithm developed in Section V to draw ten vectors of simulated residuals (and their associated importance sampling weights) for each observation in the data set. Tables 3 and 4 report two tests of the normality assumption based on these simulated residuals.

Table 3 presents χ^2 tests of normality based on the deciles of the normal distribution. These tests show that the normality assumption is substantially supported. Indeed, I reject normality at the 1% level for only one of the 12 residuals—the affiliate materials share—and at the 5% level, I reject normality only for one additional residual—the parent capital share.[17]

Table 4 presents the simulated deciles for each residual, and compares these to the deciles of the standard normal distribution. Starred entries in Table 4 denote deciles of the residuals that depart significantly from normality. Only the deciles for the parent capital and affiliate materials shares show any significant departures from the deciles of the normal distribution. I can reject normality for four of the deciles of the α^{Kd} residual and two of the deciles of the α^{Mf} residual. However, even in these worst cases, the departures from normality are not dramatic.

Recall that a novel feature of SML based on FOCs approach when applied to dynamic models is the assumption that forecast errors are normal. This is also a feature of the pseudo-SML based on FOCs approach developed in Krusell, Ohanian, Rios-Rull and Violante (2000). Traditionally, researchers have been reluctant to specify that forecast errors have any particular parametric distribution, such as normality, because there is no theoretical reason to expect them to have any particular distribution. Thus, it is important that we have a means of testing the normality assumption.

The statistics in Table 4 indicate strong support for the normality assumption on the forecast errors. According to the results in Table 4 the deciles of the simulated distributions for the domestic and foreign labor force adjustment cost forecast errors, η^d and η^f are essentially indistinguishable from normality. This finding is particularly encouraging with regard to the potential usefulness of the SML based on FOCs approach.

In summary, the model appears to fit the data quite well, in the sense that the parametric assumptions on the distributions of the stochastic terms appear (for the most part) to be strongly supported.

---

[17] The test statistic is χ^2 (9), based upon the deciles of the normal distribution. I bootstrap the critical values to deal with the fact that residuals for a firm are highly serially correlated.

VI.C. Some Simulation Experiments

A key advantage of the SML based on FOCs approach over GMM is that, in conjunction with the sampling algorithm developed in section V, it enables one to draw from the posterior distribution of the firm specific parameters (conditional on the model and the data). Given such draws, we can simulate the behavior of individual firms, and/or the whole population of firms, to changes in the economic environment. The GMM approach does not allow this, because it does not deliver estimates of the distributions of the stochastic terms.

On the other hand, in the SML based on FOCs approach the model is not completely specified. In the present case, I have not specified the stochastic processes for the forcing processes such as tariffs, exchange rates, wages, technical change, etc., nor have I specified how firms form expectations of future values of these quantities. Since I have not implemented a full-solution algorithm, I cannot simulate the short-run outcomes generated by the dynamic model. To do the latter would require a complete model, which could be estimated using FIML. However, I can use a steady-state version of the model (which assumes no labor force adjustment costs) to simulate the long-run response of the population of firms to changes in the policy environment.

In Table 5, I use simulations of the model to examine the effect of various changes in the economic environment on MNC behavior. The simulations are done using 543 vectors of technology and demand parameters drawn from the posterior distribution of the firm specific parameters of the model, using eth algorithm of Section V. Feinberg and Keane (2003a) discuss these simulations in detail, so here I just give some highlights.

The first column of Table 5 reports the actual changes in several variables of interest that occurred from 1984-1995. The second column reports the predicted change from 1984 to 1995 in the steady state level of each variable, given all changes in the environment. The third, fourth and fifth columns report the predicted changes due to changes in tariffs, technology and wages, respectively. The last column reports the combined effect of all other factors in the model.

Of course, the predicted change in steady state levels is not directly comparable to the change in actuals, since the latter include transition dynamics. However, from a face validity standpoint it is comforting that the predicted changes line up reasonably well with the actual changes. For example, the model predicts that all factors combined led to an increase of 99% in U.S. parent intra-firm sales to affiliates, and a 79% increase in U.S. parent arms-length sales to Canada. The actual changes were 73% and 75% respectively. And the model predicts that all factors combined lead to a 123% increase in affiliate intra-firm sales to parents, and a 9% increase in affiliates arms-length sales to the U.S.. The actual changes were 95% and 4% respectively. It is interesting that employment falls both in the data and according to the model. As I discussed in Section VI.A, this is consistent with the pattern we would expect for firms with substantial market power, given that TFP increased while demand (and any given price) fell over the sample period.

Now lets look at the model's decomposition of how specific exogenous factors affected MNC behavior over the sample period. The model predicts that tariff reductions over the 1984-1995 period increased U.S. parents' arms length sales to Canada by 36% and affiliates' arms-length sales to the U.S. by 32.5%. Note, however, that affiliate arms-length sales to the U.S. only increase 4% in the data, and they only increased 9% in the simulation that takes all factors into account. According to the model, other factors were hindering Canadian affiliate arms-length exports to the U.S.. One key factor was rising Canadian real wages. The real wage (in U.S. dollar terms) paid by Canadian affiliates increased by 20% from 1984-1995. The model predicts that this reduced affiliates' arms-length exports to the U.S. by 20%.

Table 5 also reveals the important role that the model assigns to technical change in increasing intra-firm trade. Of the 99% increase in parents' intra-firm sales to affiliates, the model attributes 72 percentage points to technical change. And of the 123% increase in affiliates' intra-firm sales to parents, the model attributes 84 percentage points to technical change. The changes attributed to tariff reductions are only about 10 and 5 percentage points, respectively. While these tariff effects are not trivial, they are an order of magnitude smaller than the impact of technology. Feinberg and Keane (2003a) discuss possible technological reasons for the increase in intra-firm trade.

VII. Conclusions

In this paper I have described the SML based on FOCs approach for structural estimation of economic models. The approach is a compromise between GMM and FIML because the economic model does not have to be completely specified and fully solved, a feature the method shares with GMM. But parametric distributions must be specified for all the stochastic terms that appear in the FOCS, a feature the method has in common with FIML.

The key virtue of the method relative to GMM is that it allows for multiple sources of structural error, and hence enables one to model rich patterns of heterogeneity across economic agents. In contrast, to apply GMM to models with multiple sources of structural error, one typically needs strong functional form assumptions so that the FOCS can nevertheless be written in terms of a single additive composite error term.

There are two computational problems that will typically arise in applying the SML based on FOCs method. First, in a model where the stochastic terms enter the FOCs nonlinearly, it will typically be the case that the Jacobian of the transformation from the stochastic terms to the data will be intractable. I have shown how one can deal with this problem by using numerical and simulation methods to approximate the Jacobian.

Second, economic models often have the feature that for certain ranges of the stochastic terms the FOCs cannot be satisfied (i.e., we are at a corner solution where the firm shuts down, demand is zero, etc.). I have shown how to deal with this problem by developing a new recursive importance algorithm to simulate the likelihood function. This algorithm is based on drawing the stochastic terms from the "wrong" density, chosen in such a way that all draws from the density allow the FOCs to be satisfied. The importance sampling weights adjust for the fact that the "wrong" density is used. This algorithm is the discrete/continuous analogue of the GHK algorithm.

Once one has estimated a model using the SML based on FOCs method, it is desirable to be able to draw from the posterior distribution of the firm specific parameters. This is useful both because one would like to test ones distributional assumptions, and because one would like to be able to simulate the behavior of the economic agents in the model. The GMM approach does not allow this because it does not produce estimates of the distributions of a model's stochastic terms. The SML based on FOCs approach does provide estimates of the distributions of the stochastic terms, so posterior simulation is in principle possible. However, such posterior simulation is in general a difficult problem in models where there is not a one-to-one mapping from the data to the stochastic terms. In models with rich patterns of heterogeneity across agents, there will typically be more stochastic terms than there are observed data elements.

To deal with this problem, I have presented a recursive importance algorithm that can be used to simulate from a J dimensional vector of agent specific stochastic terms that lie in a K dimensional sub-space, where $K<J$ is determined by the constraints implied by the data and the FOCs. The particular problem one faces here is that the sub-space of valid draws has measure zero. The algorithm that I develop to draw from such a sub-space is the continuous data analogue of the GHK algorithm.

The methods developed in the paper were applied to an illustrative model of the behavior of U.S. MNCs with affiliates in Canada. The methods appeared to work well in practice, and the estimates were numerically well behaved. Interestingly, there appeared to be little evidence against the distributional assumptions of the model, particularly the assumption of normally distributed forecast errors. This seems encouraging from the perspective of future application of the method.

References

Ackerberg, D. and K. Caves (2003), "Structural Estimation of Production Functions," Working paper, UCLA.

Eden, L. 1998. *Taxing Multinationals: Transfer Pricing and Corporate Income Taxation in North America*. Toronto: University of Toronto Press.

Feinberg, S. E. and M.P. Keane (2003a), "Accounting for the Growth of MNC Based Trade Using a Structural Model of U.S. MNCs," Working paper, University of Maryland.

Feinberg, S. E. and M.P. Keane (2003b), "Tariff effects on MNC organization," Working paper, University of Maryland.

Griliches, Z. and J. Mairesse, (1995), "Production Functions: The Search for Identification," *NBER Working Paper #5067*.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, Vol. 50, No. 4., pp. 1029-1054.

Hansen, L. P. and K. J. Singleton (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, Vol. 50, No. 5., pp. 1269-1286.

Katayama, H., Lu, S. and J. Tybout, (2003), "Why Plant-Level Productivity Studies are Often Misleading, and an Alternative Approach to Interference," *NBER Working Paper #9617.*

Keane, M. (1994), "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, Vol. 62, No. 1. (Jan., 1994), pp. 95-116.

Klette, T.J. and Z. Griliches, 1996. The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous, *NBER Working Paper# 4026.*

Krusell, P., Ohanian, L. E., Ríos-Rull, J.V., and G. L. Violante (2000), "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, 68:5, p.1029.

Lee, L. F. (1992), "On the Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models," *Econometric Theory*, 8, pp. 518-552.

Lee, L. F. (1995), "Asymptotic Bias in Maximum Simulated Likelihood Estimation of Discrete Choice Models," *Econometric Theory*, 11, pp. 437-483.

Levinsohn, J. and M. Melitz (2002), "Productivity in a Differentiated Products Market Equilibrium," Working paper, University of Michigan.

McCallum, B.T. (1976), "Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates," *Econometrica*, Vol. 44, No. 1., pp. 43-52.

Table 1: Parameter Estimates for Structural Model of MNC Behavior

	Parameter Name	Symbol	Estimate	Std. Error
I. U.S. Parent Technology	Intercept	α_0^{Ld}	-0.2288	(0.0925) b
Parent Labor	Time trend ([Nd>0])	$\alpha_{time}^{Ld} \cdot t \cdot I[N_d>0]$	-0.1135	(0.0089) a
	Time trend ([Nd=0])	$\alpha_{time}^{Ld} \cdot t \cdot I[N_d=0]$	-0.0149	(0.0199)
	Intercept Shift	$\alpha_{shift}^{Ld} \cdot I[N_d>0]$	5.4032	(0.1096) a
	Box Cox parameter	bc(1)	-0.0182	(0.0035) a
Parent Materials	Intercept	α_0^{Md}	0.2311	(0.0834) a
	Time trend ([Nd>0])	$\alpha_{time}^{Md} \cdot t \cdot I[N_d>0]$	-0.0922	(0.0083) a
	Time trend ([Nd=0])	$\alpha_{time}^{Md} \cdot t \cdot I[N_d=0]$	-0.0044	(0.0168)
	Intercept Shift	$\alpha_{shift}^{Md} I[N_d>0]$	4.8860	(0.1142) a
	Box Cox parameter	bc(2)	-0.0398	(0.0038) a
Parent Capital	Intercept	α_0^{Kd}	4.5094	(0.1125) a
	Time trend ([Nd>0])	$\alpha_{time}^{Kd} \cdot t \cdot I[N_d>0]$	-0.0764	(0.0058) a
	Box Cox parameter	bc(3)	-0.0635	(0.0045) a
II. Canadian Affiliate Technology	Intercept	α_0^{Lf}	-0.2373	(0.1172) b
Affiliate Labor	Time trend ([Nd>0])	$\alpha_{time}^{Lf} \cdot t \cdot I[N_f>0]$	-0.0448	(0.0058) a
	Time trend ([Nd=0])	$\alpha_{time}^{Lf} \cdot t \cdot I[N_f=0]$	0.0167	(0.0134)
	Intercept Shift	$\alpha_{shift}^{Lf} I[N_f>0]$	1.5053	(0.0964) a
	Box Cox parameter	bc(4)	-0.0771	(0.0091) a
Affiliate Materials	Intercept	α_0^{Mf}	0.4295	(0.1920) b
	Time trend ([Nd>0])	$\alpha_{time}^{Mf} \cdot t \cdot I[N_f>0]$	-0.0301	(0.0103) a
	Time trend ([Nd=0])	$\alpha_{time}^{Mf} \cdot t \cdot I[N_f=0]$	0.0259	(0.0186)
	Intercept Shift	$\alpha_{shift}^{Mf} I[N_f>0]$	1.3688	(0.1694) a
	Box Cox parameter	bc(5)	-0.0277	(0.0052) a
Affiliate Capital	Intercept	α_0^{Kf}	1.2639	(0.2013) a
	Time trend ([Nd>0])	$\alpha_{time}^{Kf} \cdot t \cdot I[N_f>0]$	-0.0808	(0.0084) a
	Box Cox parameter	bc(6)	-0.0260	(0.0062) a
Parent and Affiliate Scaling Parameters for Nd, Nf=0	US Parent Scaling Parameter	$SC_d I[N_d=0]$	0.4065	(0.0117) a
	Affiliate Scaling Parameter	$SC_f I[N_f=0]$	0.6334	(0.0141) a
Common Parameters	Profit Rate	R_K	0.1652	(0.0033) a
	TFP growth rate	h	0.0453	(0.0080) a
	Discount factor	β	0.9500	-

Notes: a = significant at 1% level; b=significant at 5% level; c=significant at 10% level.

Table 1 (Continued)

	Parameter Name	Parameter	Estimate	Std. Error
U.S. Parent Labor Adjustment Costs	Intercept	δ_{0d}	-5.1078	(0.0995) [a]
	Domestic wage	$\delta_d w_{dt}$	1.0194	(0.0267) [a]
	Time trend	$\delta_{d,\ time} \cdot t$	0.0240	(0.0038) [a]
	Intercept Shift	$\delta_d \cdot I[N_{dt}>0]$	-0.0151	(0.0287)
CA Affiliate Labor Adjustment Costs	Intercept	δ_{0f}	-5.4817	(0.0506) [a]
	Foreign wage	$\delta_f w_{ft}$	0.9667	(0.0073) [a]
	Time trend	$\delta_{f,\ time} \cdot t$	0.0884	(0.0049) [a]
	Intercept Shift	$\delta_f \cdot I[N_{ft}>0]$	0.1563	(0.0500) [a]
Common Adjustment Cost Parameters	numerator exponent	μ	0.7161	(0.0015) [a]
	denominator exponent	Δ	0.2216	(0.0047) [a]
Forecast Errors - Standard Deviation and Correlation	parent intercept	τ_{d0}	0.3186	(0.1409) [b]
	affiliate intercept	τ_{f0}	-0.2072	(0.1247) [c]
	labor foce size	τ_{1*L}	1.1368	(0.0143) [a]
	Correlation	CORR(τd, τf)	0.2974	(0.0808) [a]
Inverse Price elasticity of demand for domestically-produced good	Intercept	$g_{1,0}$	-1.3610	(0.0960) [a]
	Time trend	$g_{1,time} * t$	0.0001	(0.0005)
	Intercept Shift	$g_{1,shift} \cdot I[Nd>0]$	0.0055	(0.0061)
	Box Cox parameter	bc(7)	0.6084	(0.0674) [a]
Inverse Price elasticity of demand for foreign-produced good	Intercept	$g_{2,0}$	-1.0878	(0.0435) [a]
	Time trend	$g_{2,time} * t$	-0.0021	(0.0004) [a]
	Intercept Shift	$g_{2,shift} \cdot I[Nf>0]$	-0.0002	(0.0037)
	Box Cox parameter	bc(8)	0.8424	(0.0433) [a]
Demand function Parameters Domestic demand for domestically-produced good	Intercept	$P_{0d}{}^{1}{}_{,0}$	2.6898	(0.0908) [a]
	Time trend	$P_{0d}{}^{1}{}_{,time} \cdot t$	-0.0526	(0.0085) [a]
	Intercept Shift	$P_{0d}{}^{1}{}_{shift} \cdot I[Nd>0]$	0.0647	(0.0346) [c]
	Box Cox parameter	bc(9)	-0.0049	(0.0157)
Domestic demand for foreign-produced good	Intercept	$P_{0d}{}^{2}{}_{,0}$	2.5330	(0.1875) [a]
	Time trend	$P_{0d}{}^{2}{}_{,time} \cdot t$	-0.0740	(0.0123) [a]
	Intercept Shift	$P_{0d}{}^{2}{}_{shift} \cdot I[Nf>0]$	0.3912	(0.1018) [a]
	Box Cox parameter	bc(12)	0.1013	(0.0235) [a]
Foreign demand for domestically-produced good	Intercept	$P_{0f}{}^{1}{}_{,0}$	2.4520	(0.0816) [a]
	Time trend	$P_{0f}{}^{1}{}_{,time} \cdot t$	-0.0539	(0.0083) [a]
	Intercept Shift	$P_{0f}{}^{1}{}_{shift} \cdot I[Nd>0]$	0.0583	(0.0269) [b]
	Box Cox parameter	bc(11)	-0.0135	(0.0136)
Foreign demand for foreign-produced good	Intercept	$P_{0f}{}^{2}{}_{,0}$	2.8451	(0.2123) [a]
	Time trend	$P_{0f}{}^{2}{}_{,time} \cdot t$	-0.0888	(0.0128) [a]
	Intercept Shift	$P_{0f}{}^{2,}{}_{shift} \cdot I[Nf>0]$	0.4714	(0.1129) [a]
	Box Cox parameter	bc(10)	0.1568	(0.0195) [a]

Notes: a = significant at 1% level; b=significant at 5% level; c=significant at 10% level.

Table 2: Correlation Matrices for the Firm-Specific Stochastic Terms

1. Correlation Matrix: Composite Error: $\mu_i + \varepsilon_{it}$

Parameter:	α^{Ld}	α^{Md}	α^{Kd}	α^{Lf}	α^{Mf}	α^{Kf}	g_1	g_2	P_{0d}^1	P_{0d}^2	P_{0f}^1	P_{0f}^2
α^{Ld}	1.00											
α^{Md}	0.84	1.00										
α^{Kd}	0.95	0.93	1.00									
α^{Lf}				1.00								
α^{Mf}				0.85	1.00							
α^{Kf}				0.88	0.89	1.00						
g_1							1.00					
g_2								1.00				
P_{0d}^1									1.00			
P_{0d}^2									0.40	1.00		
P_{0f}^1									0.97	0.44	1.00	
P_{0f}^2									0.42	0.97	0.44	1.00

2. Correlation Matrix: Composite Error: t and t+1

Parameter:	α^{Ld}	α^{Md}	α^{Kd}	α^{Lf}	α^{Mf}	α^{Kf}	g_1	g_2	P_{0d}^1	P_{0d}^2	P_{0f}^1	P_{0f}^2
α^{Ld}	0.72	0.59	0.69									
α^{Md}	0.59	0.73	0.68									
α^{Kd}	0.69	0.68	0.73									
α^{Lf}				0.79	0.68	0.68						
α^{Mf}				0.68	0.78	0.68						
α^{Kf}				0.68	0.68	0.73						
g_1							0.46					
g_2								0.53				
P_{0d}^1									0.78	0.37	0.77	0.39
P_{0d}^2									0.37	0.73	0.40	0.71
P_{0f}^1									0.77	0.40	0.80	0.40
P_{0f}^2									0.39	0.71	0.40	0.75

Table 3: χ^2 Tests of Normality based on Deciles of the Normal (0,1) Distribution

Residual	χ^2	Residual	χ^2	Residual	χ^2
α^{Ld}	9.549	g_1	8.583	τ_{d0}	9.138
α^{Md}	8.559	g_2	11.195	τ_{f0}	0.512
α^{Kd}	18.454*	P_{0d}^1	4.257		
α^{Lf}	2.983	P_{0d}^2	2.759		
α^{Mf}	27.958*	P_{0f}^1	5.131		
α^{Kf}	11.889	P_{0f}^2	7.978		

*Rejected at $p<.05$, based on the bootstrapped critical value of the χ^2 distribution

Table 4: Comparison of Quantile Points of Residual Distributions versus the Normal (0,1)

Z	-2.58	-2.24	-1.96	-1.65	-1.28	-0.52	-0.25	0.00	0.25	0.52	1.28	1.65	1.96	2.24	2.58
Prob $\varepsilon < Z$	0.01	0.025	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.99
Parameter:															
α^{Ld}	0.008	0.023	0.043	0.094	0.214	0.296	0.385	0.492	0.614	0.721	0.823	0.907	0.945	0.966	0.984
	(.003)	(.006)	(.009)	(.012)	(.018)	(.020)	(.021)	(.021)	(.019)	(.018)	(.014)	(.010)	(.008)	(.006)	(.004)
α^{Md}	0.011	0.027	0.052	0.096	0.195	0.288	0.391	0.489	0.623	0.726	0.806	0.895	0.950	0.971	0.986
	(.004)	(.007)	(.009)	(.012)	(.017)	(.019)	(.021)	(.020)	(.019)	(.018)	(.015)	(.011)	(.007)	(.005)	(.003)
α^{Kd}	0.022	0.046	0.078	0.147*	0.261*	0.362*	0.453	0.539	0.621	0.691	0.778	0.867*	0.920	0.960	0.981
	(.007)	(.010)	(.013)	(.017)	(.022)	(.024)	(.025)	(.025)	(.023)	(.022)	(.019)	(.015)	(.011)	(.007)	(.005)
α^{Lf}	0.006	0.022	0.043	0.095	0.198	0.311	0.412	0.521	0.615	0.717	0.814	0.899	0.939	0.966	0.980
	(.002)	(.005)	(.008)	(.011)	(.016)	(.020)	(.022)	(.021)	(.020)	(.018)	(.016)	(.011)	(.009)	(.006)	(.005)
α^{Mf}	0.008	0.017	0.041	0.099	0.194	0.290	0.385	0.476	0.659*	0.737*	0.814	0.904	0.937	0.971	0.980
	(.003)	(.004)	(.007)	(.012)	(.016)	(.019)	(.020)	(.020)	(.020)	(.017)	(.015)	(.012)	(.010)	(.006)	(.006)
α^{Kf}	0.011	0.023	0.058	0.115	0.226	0.349	0.467	0.566	0.650	0.726	0.803	0.901	0.941	0.957	0.977
	(.004)	(.006)	(.010)	(.014)	(.018)	(.021)	(.023)	(.023)	(.023)	(.021)	(.017)	(.012)	(.010)	(.008)	(.007)
g_1	0.007	0.020	0.044	0.093	0.188	0.281	0.381	0.485	0.596	0.708	0.821	0.923	0.966	0.985	0.996
	(.001)	(.002)	(.004)	(.006)	(.009)	(.011)	(.012)	(.012)	(.012)	(.011)	(.009)	(.007)	(.004)	(.003)	(.001)
g_2	0.012	0.035	0.066	0.119	0.197	0.281	0.372	0.474	0.584	0.696	0.807	0.908	0.951	0.969	0.982
	(.003)	(.005)	(.007)	(.011)	(.014)	(.016)	(.017)	(.017)	(.016)	(.015)	(.013)	(.009)	(.007)	(.006)	(.004)
P_{0d}^{1}	0.005	0.018	0.045	0.093	0.182	0.283	0.393	0.504	0.609	0.719	0.822	0.908	0.950	0.974	0.991
	(.002)	(.004)	(.007)	(.011)	(.014)	(.017)	(.018)	(.018)	(.017)	(.016)	(.013)	(.010)	(.008)	(.005)	(.003)
P_{0d}^{2}	0.015	0.025	0.044	0.086	0.184	0.274	0.380	0.493	0.616	0.720	0.815	0.900	0.944	0.971	0.987
	(.007)	(.009)	(.011)	(.016)	(.022)	(.026)	(.028)	(.029)	(.028)	(.026)	(.022)	(.017)	(.012)	(.009)	(.005)
P_{0f}^{1}	0.006	0.020	0.049	0.100	0.187	0.276	0.390	0.496	0.603	0.719	0.822	0.904	0.948	0.975	0.991
	(.003)	(.005)	(.009)	(.013)	(.017)	(.019)	(.020)	(.021)	(.020)	(.018)	(.015)	(.011)	(.009)	(.006)	(.003)
P_{0f}^{2}	0.012	0.028	0.053	0.093	0.183	0.265	0.384	0.503	0.614	0.721	0.817	0.908	0.953	0.975	0.990
	(.005)	(.007)	(.009)	(.012)	(.016)	(.018)	(.019)	(.019)	(.018)	(.017)	(.014)	(.010)	(.007)	(.005)	(.003)
τ_{d0}	0.011	0.030	0.057	0.112	0.215	0.311	0.406	0.502	0.596	0.690	0.789	0.888	0.940	0.968	0.986
	(.001)	(.003)	(.004)	(.006)	(.009)	(.010)	(.011)	(.012)	(.011)	(.010)	(.009)	(.006)	(.005)	(.003)	(.002)
τ_{f0}	0.010	0.025	0.049	0.100	0.202	0.301	0.398	0.498	0.600	0.698	0.798	0.900	0.950	0.974	0.989
	(.001)	(.001)	(.001)	(.002)	(.003)	(.004)	(.004)	(.004)	(.004)	(.003)	(.003)	(.002)	(.001)	(.001)	(.001)

Note: *Rejected at p<.05, based on bootstrapped distribution of the quantile points

Table 5: Percentage Changes 1984-1995: Data and Model Predictions

	Data	Model	Model Decomposition: Tariffs	Technology	Wages	Other
USP Sales						
Domestic ($P_d^1 S_d$)	-19.6	-9.0	0.8	4.0	-4.3	-9.5
Intra-firm ($P_d^1 N_f$)	73.3	99.1	9.7	71.8	14.4	3.2
Arms-Length ($P_f^1 E$)	75.3	79.1	36.0	9.0	-1.3	35.4
Total	-17.1	-6.0	1.1	7.3	-3.9	-10.5
Total Net	-20.2	-9.8	1.0	3.8	-3.9	-10.7
CA Sales						
Domestic ($P_f^2 S_f$)	-17.9	-6.9	6.4	-71.9	18.3	40.3
Intra-firm ($P_f^2 N_d$)	94.5	123	5.2	84.1	-2.9	36.6
Arms-Length ($P_d^2 I$)	4.0	8.9	32.5	37.9	-20.5	-41.0
Total	18.7	38.3	7.3	7.8	11.2	12.0
Total Net	-0.5	14.2	6.3	-23.3	9.9	21.3
MNC Sales	-19.0	-8.4	1.3	2.4	-3.0	-9.1
USP Employment	-24.4	-12.6	1.4	12.4	-1.6	-24.8
CA Employment	-14.1	-3.0	9.0	15.5	-7.0	-20.5

Notes: "Data" shows percentage change in our analysis data set. "Model" shows percentage change from 1984-95 in the model simulation. Under "Model Decomposition," we compare the % change predicted by the model under the baseline simulation, with the % change the model predicts under the counterfactual that the indicated forcing variable (tariffs, technology or wages) stayed fixed at the 1984 level. The difference is the percentage change attributable to changes in that particular forcing variable.

Appendix 1: The Mapping from the Data to the Stochastic Terms of the Model

The mapping from the data to the stochastic terms of the model is obtained by solving the system of nonlinear equations given by the FOCs and the production functions. If the MNC utilizes all four trade flows (bilateral intra-firm and arms length flows) this is a system of 12 equations in 12 unknowns. We describe the solution in this general case.

Step 1: By summing the FOCs for L_d, K_d, M_d and N_d (see eqn. (19) in Section IV.2) we obtain:

(A1.1) $$(1-g_1A)(P_d^1Q_d)-CD+g_2(P_f^2N_d)B-\delta_dE(FD)L_d=0$$

Similarly, for the affiliate, we obtain:

(A1.2) $$(1-g_2B)(P_f^2Q_f)-CF+g_1(P_d^1N_f)A-\delta_fE(FF)L_f=0$$

where CD and CF are U.S. and Canadian costs, respectively (including capital, labor, materials and intermediates, but not adjustment costs). We have utilized the CRTS assumption to eliminate the share parameters, leaving g_1 and g_2 as the only unknown parameters in (A1.1) and (A1.2). However, the variables $P_d^1Q_d$ and $P_f^2Q_f$ are not observed, so we cannot yet solve for g_1 and g_2.

Step 2: We can express $P_d^1Q_d$ and $P_f^2Q_f$ in terms of observed data as follows:

(A1.3) $$P_d^1Q_d=P_d^1S_d+P_d^1N_f+(P_d^1/P_f^1)(P_f^1E)$$

(A1.4) $$P_f^2Q_f=P_f^2S_f+P_f^2N_d+(P_f^2/P_d^2)(P_d^1I)$$

Here, only the price ratios P_d^1/P_f^1 and P_f^2/P_d^2 are not observed. At this point, it is convenient to denote these price ratios as PR_1 and PR_2, and treat them as unknown parameters.

Before proceeding, it is useful to give an intuition for identification of g_1 and g_2. First, consider what (A1.1) and (A1.2) imply in the special case of no intra-firm intermediates. In that case, the $g_2(P_f^2N_d)B$ and $g_1(P_d^1N_f)A$ terms drop out, so we can consider each equation separately. Furthermore, $A=B=1$, and the domestic to foreign price ratios depend only on the wedge created by tariff and transport costs. Therefore, $P_d^1Q_d$ and $P_f^2Q_f$ are observable, and they are exactly equal to total revenues (net of tariff and transport costs) of the parent and affiliate. Denote these by RD and RF.[18] Then we get $g_1=(RD-CD-\delta_dE(FD)L_d)/RD$, which is the usual Lerner markup equation in the CRTS case, modified to account for expected labor force adjustment costs. Thus, g_1 and g_2 are basically identified from price/cost markups.

With intra-firm trade in intermediates, the price ratios PR_1 and PR_2 are not observed, we have $0<A<1$ and $0<B<1$, and the $g_2(P_f^2N_d)B$ and $g_1(P_d^1N_f)A$ terms matter, so (A1.1) and (A1.2) must be solved jointly. All three of these changes reflect the fact that with intra-firm trade in intermediates, the firm's incentive to hold down output to increase prices is mitigated because higher output prices also imply higher tariff and transport costs.

Fortunately, the price ratios PR_1 and PR_2 depend only on tariff and transport costs and on g_1 and g_2. For instance, the FOC for E shows how the parent's incentive to hold down P_d^1 to avoid tariff and transport costs on shipping N_f to the Canadian affiliate depends on g_1, tariffs and

[18] Note that $RD=P_d^1S_d+P_d^1N_f+(1-T_f-C_f)(P_f^1E)$. In the absence of intra-firm trade in intermediates, $P_f^1/P_d^1=1/(1-T_f-C_f)$, so the last term reduces to P_d^1E, the value of exports at domestic prices. Hence, $P_d^1Q_d=RD$.

transport costs.[19] Thus, in steps 3-4, we use the FOCs for E and I to substitute for the unobserved price ratios in (A1.3) and (A1.4), obtaining two equations we can solve for g_1 and g_2.

It is worth noting that if we introduced an elasticity of substitution between goods 1 and 2, of if we let the price elasticities differ by country, we would have more unknowns that equations, and we could not identify all the elasticities using only data on sales and revenues.

Step 3: Using (A1.3)-(A1.4) to substitute for $P_d^1 Q_d$ and $P_f^2 Q_f$ in (A1.1)-(A1.2), we obtain:

(A1.5) $$[Y+(PR_1)(P_f^1 E)](1-g_1 A)+g_2(P_f^2 N_d)B = CD+\delta_d E(FD)L_d$$

(A1.6) $$[V+(PR_2)(P_d^1 I)](1-g_2 B)+g_1(P_d^1 N_f)A = CF+\delta_f E(FF)L_f$$

where we have defined $Y = P_d^1 S_d + P_d^1 N_f$ and $V = P_f^2 S_f + P_f^2 N_d$. Note that we now have two equations in the 4 unknowns: g_1, g_2, PR_1 and PR_2.

Step 4: We now use the FOCs for E and I to add two more equations in g_1, g_2, PR_1 and PR_2:

(A1.7) $$E: PR_1(1-g_1 A) = (1-T_f-C_f)(1-g_1)$$

(A1.8) $$I: PR_2(1-g_2 B) = (1-T_d-C_d)(1-g_2)$$

Next, we use (A1.7)-(A1.8) to substitute for the price ratios in (A1.5)-(A1.6), to obtain:

(A1.9) $$g_1(AYE) = RD-CD-\delta_d E(FD)L_d + g_2(P_f^2 N_d)B$$

(A1.10) $$g_2(BVI) = RF-CF-\delta_f E(FF)L_f + g_1(P_d^1 N_f)A$$

where we have defined:

$$AYE \equiv A(P_d^1 S_d + P_d^1 N_f)+(1-T_f-C_f)(P_f^1 E)>0$$

$$BVI \equiv B(P_f^2 S_f + P_f^2 N_d)+(1-T_d-C_d)(P_d^2 I)>0$$

Solving for g_1, we have:

(A1.11) $$g_1\left[AYE-\frac{AB(P_d^1 N_f)(P_f^2 N_d)}{BVI}\right] = RD-CD-\delta_d E(FD)L_d + \frac{(RF-CF-\delta_f E(FF)L_f)}{BVI}(P_f^2 N_d)B$$

and a similar equation gives g_2. [This is referred to as Equation (23) in the text.]

Step 6: Given g_1 and g_2, we can use (A1.7)-(A1.8) to solve for the price ratios:

(A1.12) $$PR_1 = \frac{1-g_1}{1-g_1 A}(1-T_f-C_f) \qquad PR_2 = \frac{1-g_2}{1-g_2 B}(1-T_d-C_d)$$

Step 7: Given PR_1 and PR_2, we construct $P_d^1 Q_d$ and $P_f^2 Q_f$ and solve for the share parameters:

$$\alpha^{Ld} = [(w_d L_d)+\delta_d E(FD)L_d] / (1-g_1 A)(P_d^1 Q_d) \qquad \alpha^{Md} = (\phi_d M_d) / (1-g_1 A)(P_d^1 Q_d)$$

(A1.13) $$\alpha^{Kd} = (\gamma_d K_d) / (1-g_1 A)(P_d^1 Q_d) \qquad \alpha^{Nd} = 1-\alpha^{Ld}-\alpha^{Kd}-\alpha^{Md}$$

and similarly for the affiliate technology parameters.

[19] With intra firm trade, the FOC for E gives: $PR_1 = P_d^1/P_f^1 = (1-T_f-C_f)(1-g_1)/(1-g_1 A)$, where $A<1$. This price ratio is less than that in the no intra-firm trade case, because $(1-g_1)/(1-g_1 A)$ is less than 1. Thus, P_d^1 is relatively lower.

Step 8: Since $PR_1 \equiv \left(P^1_d / P^1_f\right) = \left(P^1_{0d} / P^1_{of}\right)\left(P^1_d S_d / P^1_f E\right)^{-g_1} \cdot (P^1_f E)$ we can use g_1 and PR_1 to construct the ratio of the U.S. to Canadian demand function intercepts for the domestically produced good, $\left(P^1_{0d} / P^1_{0f}\right)$. Similarly, we use g_2 and PR_2 to construct $\left(P^2_{0f} / P^2_{0d}\right)$. Intuitively, given knowledge of the price elasticity of demand and the U.S./Canadian price ratio, the ratios of the demand function intercepts are determined by the ratio of U.S. to Canadian sales.

Step 9: To determine the levels of the demand function intercepts, we need to construct real quantities of output. Since we know the production function parameters, the extra information we need is data on prices of capital equipment and materials, so we can construct quantities of capital and materials inputs.[20] Note that for the parent we can write:

$$\text{(A1.14)} \qquad \frac{P^1_d Q_d}{P^1_d N_f} = \frac{P^1_d H_d L_d^{\alpha^{Ld}} K_d^{\alpha^{Kd}} M_d^{\alpha^{Md}} N_d^{\alpha^{Nd}}}{P^1_d N_f} \equiv \frac{P^1_d LKM_d N_d^{\alpha^{Nd}}}{P^1_d N_f} = \frac{LKM_d N_d^{\alpha^{Nd}}}{N_f}$$

where we have defined $LKM_d = H_d L_d^{\alpha^{Ld}} K_d^{\alpha^{Md}} M_d^{\alpha^{Kd}}$. Similarly, for the affiliate, we have:

$$\text{(A1.15)} \qquad P^2_f Q_f \big/ P^2_f N_f = LKM_f N_f^{\alpha^{Nf}} \big/ N_d$$

where $LKM_f = H_f L_f^{\alpha^{Lf}} K_f^{\alpha^{Mf}} M_f^{\alpha^{Kf}}$. Solving (A1.14)-(A1.15) for N_d, we obtain:

$$\text{(A1.16)} \qquad N_d = \left\{ \left[\left(\frac{P^2_f N_d}{P^2_f Q_f} \right) LKM_f \right] * \left[\left(\frac{P^1_d N_f}{P^1_d Q_d} \right) LKM_d \right]^{\alpha^{Nf}} \right\}^{\frac{1}{1-\alpha^{Nd}\alpha^{Nf}}}$$

Next, we substitute (A1.16) into (A1.14) to obtain a similar equation for N_f.[21] Given N_d and N_f, we can calculate Q_d and Q_f. Given real output quantities, we can form:

$$P^1_d = \frac{P^1_d Q_d}{Q_d} \qquad\qquad P^2_f = \frac{P^2_f Q_f}{Q_f}$$

Then, the prices of exports and imports are given by:

$$P^1_f = P_d^{\,1}/PR_1 \qquad\qquad P_d^{\,2} = P^2_f/PR_2$$

Now, the quantities of exports, imports, and domestic and foreign sales are given by:

$$E = P^1_f E / P^1_f \qquad\qquad I = P_d^{\,2} I / P_d^{\,2}$$

$$S_d = Q_d - N_f - E \qquad\qquad S_f = Q_f - N_d - I$$

[20] Interestingly, we do not need price data for the intermediates used by the parent or by the affiliate, because we do not need to know quantities of intermediate inputs to determine output quantities. This is because the production function can be written solely in terms of the primary inputs. Not needing to deflate intermediates is an advantage of our framework, because firms have market power in these goods. Thus, as Griliches and Mairesse (1995) discuss, industry price indices cannot be used as deflators.

[21] If $N_d = 0$, we obtain N_f immediately from (A1.14), since $\alpha^{Nd}=0$. If $N_f = 0$ we obtain N_d immediately from (A1.15).

Finally, we can write the demand function intercepts as:

$$P_{0f}^{2} = P_f^2 S_f^{g_2} \qquad\qquad P_{0d}^{1} = P_d^1 S_d^{g_1}$$

$$P_{0d}^{2} = P_d^2 I^{g_2} \qquad\qquad P_{0f}^{1} = P_f^1 E^{g_1}$$

In the case that $I=0$, P_d^2 and P_{0d}^2 will be undefined, but P_{0f}^2 can still be calculated since $S_f = Q_f - N_d$. In the case that $E=0$, P_f^1 and P_{0f}^1 will be undefined, but P_{0d}^1 can still be calculated since $S_d = Q_d - N_f$.

Step 10: Having obtained the 12 firm specific technology and demand parameters, one can easily map these into the ε vector using the equations of the stochastic specification in section III.

Appendix 2: Identification of H, P_0 and R

As discussed in Appendix 1, the sources of identification for the Cobb-Douglas share parameters, price elasticities of demand, and demand function intercept *ratios* are rather obvious in the context of our model. This is also true of the labor force adjustment cost parameters, which are pinned down by the persistence of employment over time. Much more subtle, however, is the source of identification of the H, P_0 and R parameters, which are TFP, the demand intercept *levels*, and the profit rate. Here we give the intuition for how these parameters are identified.

Regarding TFP, recall that we normalize $H_t=1$ and $t=0$ in 1983, and specify that TFP growth is determined by $H_t = (1+h)^t$, where h is an estimated parameter. It is not immediately obvious is how growth in H would be separately identified from changes in the demand function intercepts P_0 if output prices and quantities are not separately observed. One might think that, since any observed revenue level can be explained by a locus of price and quantity values, it would be impossible to disentangle H and the P_0. However, this is not the case.

Consider a simple autarky case (a single firm with no affiliate), and with K and L as the only inputs, and no adjustment costs. Then we have:

$$Q = H \cdot K^{\alpha^K} L^{\alpha^L} \qquad P = P_0 Q^{-g} \qquad \Pi = PQ - \gamma K - wL$$

In this case, labor demand is:

$$L = \left[\left(\frac{\alpha^L}{w} \right) (1-g) P_0 H^{(1-g)} \left[\left(\frac{\alpha^K}{\gamma} \frac{w}{\alpha^L} \right)^{\alpha^K} \right]^{1-g} \right]^{1/g}$$

Calculating elasticities of labor demand with respect to H and P_0, we obtain:

(A2.1) $$\frac{P_0}{L}\frac{\partial L}{\partial P_0} = \frac{1}{g} > 0 \qquad \frac{H}{L}\frac{\partial L}{\partial H} = \frac{1}{g}(1-g) > 0 \qquad 0<g<1$$

Thus, both TFP growth and growth in demand (P_0) cause growth in the firm's employment. If all growth is due to growth in demand, then a firm's employment growth rate will be inversely proportional to its g. But if all growth is due to TFP growth, then firm growth rates will decline even more quickly with g. In fact, as $g \rightarrow 1$, the effect of TFP on employment approaches zero. Thus, to the extent that firm employment growth is less than proportional to 1/g (i.e., it is much slower among high g firms), we will conclude that TFP is a bigger factor in growth than P_0.

This argument requires that we can pin down the market power parameter g without separate information on prices and quantities. This is in fact the case. As we showed in Appendix 1, g can be determined using only data on revenues and costs (i.e., price/cost markups).

Next, we turn to the issue of how the profit rate is identified. The same simple autarky model (with CRTS and a capital share of α^K) gives:

$$\text{(A2.2)} \qquad L = \left[\frac{\gamma}{w} \frac{1-\alpha^K}{\alpha^K} \right] K \qquad K = \left[\frac{\alpha^K}{\gamma} (1-g) P_0 H^{(1-g)} \left[\frac{\gamma}{w} \frac{1-\alpha^K}{\alpha^K} \right]^{(1-\alpha^K)(1-g)} \right]^{1/g}$$

and therefore profit is:

$$\text{(A2.3)} \qquad \Pi = P_0 H^{(1-g)} \left[\frac{\gamma}{w} \frac{1-\alpha^K}{\alpha^K} \right]^{(1-\alpha^K)(1-g)} K^{(1-g)} - \frac{\gamma}{\alpha^K} K$$

From the definition of the profit rate, $R \equiv \Pi / \gamma K$, we have that $\Pi = \gamma K \cdot R$. Substituting this expression for Π in (A2.3) and then solving for K we obtain:

$$\text{(A2.4)} \qquad K = \left[\left(\frac{\gamma}{\alpha^K} + \gamma R \right)^{-1} P_0 H^{(1-g)} \left[\frac{\gamma}{w} \frac{1-\alpha^K}{\alpha^K} \right]^{(1-\alpha^K)(1-g)} \right]^{1/g}$$

Comparing the expressions for K in (A2.2) and (A2.4) we see that we must have:

$$\left(\left(\gamma / \alpha^K \right) + \gamma R \right)^{-1} = \left(\alpha^K / \gamma \right)(1-g)$$

which implies that:

$$\text{(A2.5)} \qquad g/(1-g) = \alpha^K \cdot R$$

Note that a larger g, and hence a larger $g/(1-g)$, implies greater market power. From (A2.5) we see that R (which is determined in equilibrium) governs the relationship between market power and the capital share, α^K. If the equilibrium profit rate is low (high), then there is a strong (weak) tendency for firms with more market power to also have larger capital shares, so that the profits accrue to a larger (smaller) stock of capital. Also note that g, and hence $g/(1-g)$, is negatively related to firm size. So R also governs the strength of the relationship between firm size and α^K.

Appendix 3: Derivation of the bounds on (η^d , η^f)

In this appendix we derive the bounds on (η^d, η^f) such that it is possible to rationalize firm behavior. We return to equations (A1.9) – (A1.10) and write them as:

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} AYE & -(P_f^2 N_d)B \\ -(P_d^1 N_f)A & BVI \end{pmatrix}^{-1} \begin{pmatrix} RD - CD - \delta_d\, E(FD) L_d \\ RF - CF - \delta_f\, E(FF)\, L_F \end{pmatrix}$$

This gives us:

$$\text{(A3.1)} \quad \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \frac{1}{AYE \cdot BVI - A \cdot B \cdot (P_d^1 N_f)(P_f^2 N_d)} \begin{pmatrix} BVI & -(P_f^2 N_d)B \\ -(P_d^1 N_f)A & AYE \end{pmatrix} \cdot \begin{pmatrix} RD - CD - \delta_d\, E(FD) L_d \\ RF - CF - \delta_f\, E(FF)\, L_F \end{pmatrix}$$

The term $AYE \cdot BVI - A \cdot B \cdot (P_d^1 N_f)(P_f^2 N_d)$ is > 0. Thus, for $g_1>0$ and $g_2>0$, we must have:

$$BVI \cdot [RD - CD - \delta_d E(FD)L_d] + (P_f^2 N_d) B [RF - CF - \delta_f E(FF)L_f] > 0$$
$$(P_d^1 N_f) A \cdot [RD - CD - \delta_d E(FD)L_d] + AYE \quad [RF - CF - \delta_f E(FF)L_f] > 0$$

Rearranging these expressions, we obtain:

$$(A3.2) \quad BVI \cdot [RD - CD] + (P_f^2 N_d) B [RF - CF] > BVI \cdot \delta_d E(FD)L_d + (P_f^2 N_d) B \delta_f E(FF)L_f$$
$$(A3.3) \quad (P_d^1 N_f) A \cdot [RD - CD] + AYE [RF - CF] > (P_d^1 N_f) A \cdot \delta_d E(FD)L_d + AYE \delta_f E(FF)L_f$$

Substituting into (A3.2) the definitions of the expectation terms $E(FD)L_d$ and $E(FF)L_f$ (see equation 20) we obtain an expression where the forecast errors η_d and η_f appear explicitly:

$$BVI \cdot [RD - CD] + (P_f^2 N_d) B [RF - CF] > BVI \cdot \delta_d \cdot [FD \cdot L_d + \eta_d] + (P_f^2 N_d) B \delta_f [FF \cdot L_f + \eta_f]$$

If we divide through by BVI > 0 and define $PDB \equiv (P_f^2 N_d) B \big/ BVI$ we obtain:

$$[RD - CD] + PDB \cdot [RF - CF] > \delta_d FDL_d + \delta_d \eta_d + PDB \cdot \delta_f FF \cdot L_f + PDB \cdot \delta_f \eta_f$$

We can rewrite (A3.3) in a similar way. Then, we obtain the following upper bounds on (η^d, η^f):

$$(A3.4) \quad \delta_d \eta_d + PDB \cdot \delta_f \eta_f < [RD - CD] + PDB \cdot [RF - CF] - \delta_d FDL_d - PDB \cdot \delta_f FF \cdot L_f \equiv BUD$$
$$\delta_f \eta_f + PFA \cdot \delta_d \eta_d < [RF - CF] + PFA \cdot [RD - CD] - \delta_g FFL_f - PFA \cdot \delta_f FD \cdot L_d \equiv BUF$$

where we have defined $PFA \equiv \left(P_d^1 N_f\right) A \big/ AYE$. Denoting the upper bounds given by the right hand sides of the inequalities in (A3.4) by BUD and BUF, we obtain the more compact notation:

(A3.5)
$$\delta_d \eta_d + PDB \cdot \delta_f \eta_f < BUD$$
$$\delta_f \eta_f + PFA \cdot \delta_d \eta_d < BUF$$

The (η^d, η^f) vector also has an lower bound, which we can determine from the labor share equations in A(1.13). The requirement $\alpha^{Ld}>0$ gives:

$$\delta_d E(FD) L_d + w_d L_d > 0$$

Substituting the definition of $E(FD)L_d$ we obtain $\delta_d (FD \cdot L_d + \eta_d) + w_d L_d > 0$, which implies the lower bound on η_d:

$$\delta_d \eta_d > -w_d L_d - \delta_d FDL_d \equiv BLD$$

Similarly, for η_f we obtain:

$$\delta_f \eta_f > -w_f L_f - \delta_f FFL_f \equiv BLF$$

Thus, $\delta_d \eta_d$ and $\delta_f \eta_f$ must satisfy the following four constraints in order for firm behavior to be rationalizable within the model:

(A3.6)
$$\delta_d \eta_d + PDB \cdot \delta_f \eta_f < BUD \qquad \delta_d \eta_d > BLD$$
$$\delta_f \eta_f + PFA \cdot \delta_d \eta_d < BUF \qquad \delta_f \eta_f > BLF$$

Note that the upper bound for $\delta_d \eta_d$ depends on $\delta_f \eta_f$, and vice-versa. We can write an unconditional upper bound for $\delta_d \eta_d$ in terms of BLF, the lower bound for $\delta_f \eta_f$, as follows:

$$\delta_d \eta_d + (PDB)\ (BLF) < BUD$$

This equation defines the largest possible $\delta_d \eta_d$ such that firm behavior can be rationalized by some value of η_f. If $\delta_d \eta_d$ is larger than $[BUD - (PDB) \cdot (BLF)]$ then it is impossible to find a $\delta_f \eta_f$ small enough so that the first equation in (A3.6) is satisfied. So the feasible range for $\delta_d \eta_d$ is:

$$BLD < \delta_d \eta_d < BUD - (PDB)(BLF).$$

This translates into an (unconditional) feasible range for η_d of:

$$\frac{BLD}{\delta_d} < \eta_d < \frac{BUD - (PDB)(BLF)}{\delta_d}$$

or, substituting in the definitions of BLD, BUD, BLF and PDB we have:

$$\text{(A3.7)} \quad -\frac{w_d L_d}{\delta_d} - FD \cdot L_d < \eta_d < \frac{[RD - CD] + PDB \cdot [RF - CF] + PDB \cdot w_f L_f}{\delta_d} - FD \cdot L_d$$

We rewrite (A3.7) more compactly as:

$$\text{(A3.9)} \qquad BL^d < \eta_d < BU^d$$

Now, suppose we have drawn a value of η_d that satisfies (A3.9), and denote this value by $\eta_{d,m}$ where m indicates the draw. Then the conditional bounds on η_f are determined as follows: Substituting $\eta_{d,m}$ into (A3.3) we have:

$$(P_d^1 N_f)A \cdot [RD - CD] + AYE \cdot [RF - CF] > (P_d^1 N_f)A \cdot \delta_d [FD \cdot L_d + \eta_{d,m}] + AYE \cdot \delta_f [FF \cdot L_f + \eta_f]$$

which implies that:

$$\eta_f < \frac{RF - CF}{\delta_f} - FF \cdot L_f + \frac{(P_f^1 N_f)A}{AYE} \cdot [RD - CD - \delta_d \cdot (FD \cdot L_d + \eta_{d,m})]$$

Thus, draw m for η_f must satisfy this upper bound. We write this more compactly as:

$$\text{(A3.10)} \qquad \eta_f^m < BU^f(m)$$

We have already derived $\delta_f \eta_f > BLF$, which we can write as $\eta_f > BLF/\delta_f \equiv BL^f$. Combining this with (A3.9) and (A3.10) we have the following bounds for sequential draws of η_d and η_f:

$$\text{(A3.11)} \qquad \left.\begin{Bmatrix} BL^d < \eta_d < BU^d \\ BL^f < \eta_f^m < BU^f(m) \end{Bmatrix}\right. \Rightarrow \quad \eta \in BD$$

The recursive simulation algorithm described in Section IV.3.E works by first drawing a value for η_d from its truncated distribution given the first set of unconditional bounds in (A3.11), and then drawing a value for η_f from the second set of conditional bounds in (A3.11). Draws obtained in this sequential fashion will not have the correct joint distribution, so they must be weighted using an importance sampling algorithm to obtain unbiased estimates of likelihood contributions.