Estibration: An Illustration of Structural Estimation As Calibration*

Edward J. Balistreri†
Colorado School of Mines

and

Russell H. Hillberry
University of Melbourne

October 2005

Abstract

Dawkins et al. (2001) propose that estimation is calibration. We illustrate their point by examining a leading econometric application in the study of international and inter-regional trade by Anderson and van Wincoop (2003). We replicate the econometric process, and show it to be a calibration of a general equilibrium model. Our approach offers unique insights into structural estimation, and we highlight the importance of traditional calibration considerations when one uses econometric techniques to calibrate a model for comparative policy analysis. (JEL F10, C13, C60)

*We thank Eric van Wincoop for providing the data used in this analysis. This paper was largely completed while the authors were employed with the U.S. International Trade Commission. The opinions and conclusions are solely those of the authors.

†Corresponding author: Engineering Hall 311, Division of Economics and Business, Colorado School of Mines, Golden, CO 80401-1887; email: `ebalistr@mines.edu`.

Estibration: An Illustration of
Structural Estimation As Calibration

1 Introduction

Contemporary economic analysis includes two broad traditions of fitting models to data. Many estimate stochastic, theory-based, reduced forms with few parameters, while others calibrate models by an extensive collection, and computation, of consistent fitted values. Although the first technique is called estimation and the second is called calibration, these exercises are identical under consistent identifying assumptions. Both calibration and estimation fit a model to data.

Dawkins et al. (2001) make this point succinctly: "Calibration is estimation, estimation is calibration." The point is widely recognized in the macroeconomic real-business-cycle literature [Hoover (1995)]. Our purpose is to demonstrate that it is equally relevant to micro-based general equilibrium models. In our view, there is too little communication between *calibrators* and *estimators* of such models, and the lack of communication impedes research.

It is important to clearly delineate the processes of data fitting and the subsequent model analyses. In analysis of fitted models the degree of concentration on counterfactual simulation versus hypothesis testing around specific parameters often cleaves with our respective notions of calibration and estimation, but counter examples are easily found. Sensitivity analysis found in computable general equilibrium studies can be specifically designed to generate higher-order moments to facilitate hypothesis testing.[1] Many econometric studies are specifically focused on model identification for the purpose of counterfactual simulation.[2]

A more informative distinction might be drawn between testing models and calibrating models. Hoover (1995) places this issue at the heart of the macroeconomic debate sur-

[1]See Harrison et al. (1992), Harrison and Vinod (1992), and Hertel et al. (2004).
[2]See Trefler (1993) or Rose and van Wincoop (2001).

rounding real-business-cycle models. We see a direct extension of the macroeconomic debate into any empirical methodology that involves general equilibrium systems. In the testing paradigm, stochastic measures of fit provide a critically important benchmark for evaluating alternative structural assumptions or analytical results derived from a particular set of assumptions. When the objective is to provide a quantitative context for counterfactual analysis, traditional measures of fit are little more than indicators of parsimony. One example of the macroeconomic debate spilling over to the empirical literature is forwarded by Leamer and Levinsohn (1995) who advise us to "Estimate, don't test." The point being that falsifiable theories are informative—conceptually and empirically.

In the context of estimating, as opposed to testing, some researchers find it preferable to exhaust the degrees of freedom available in the data by expanding the set of structural parameters. This results in a model that perfectly fits the benchmark observation, but requires external information on how it will respond to shocks. Model responses are usually summarized in a set of elasticity assumptions. Econometric estimation does not exhaust the degrees of freedom and may focus on internal measurements of response parameters.

There also seems to be convergence in the literatures; some calibrators attempt to identify response parameters [e.g., Liu et al. (2004), or Francois (2001)] and some estimators include numerous dummy variables to isolate cross-sectional fixed effects [e.g., Hummels (2001), or Feenstra (2004) Chapter 5]. Hillberry et al. (2005) contend that idiosyncratic calibration parameters in traditional CGE applications operate in a way that is similar to econometric residuals; calibration parameters allow the model to fit the data exactly. Another way to think about it is that idiosyncratic calibration parameters in CGE applications are analogous to fixed effects in an econometric model that includes a fixed effect for each observation.

The model proposed and applied by Anderson and van Wincoop (2003) (henceforth A-vW) is exceptionally well suited for communicating with both estimators and calibrators. In their two-country application, A-vW fit a structural gravity model to observed trade patterns among Canadian provinces and U.S. states. It is our contention that *estimating* and *calibrating* the A-vW model are equivalent methods for fitting the model to data. The model is, at its core, a general equilibrium that can be fitted with standard calibration techniques. One of these techniques is the econometric procedure proposed by A-vW. Our illustration that the econometric procedure is a calibration entails a structural interpretation of the parametric estimates and an itemization of those remaining structural parameters that are implied by the estimable model's identifying assumptions.

We show that the A-vW calibration is consistent with a broad class of first-order approximations of the general equilibrium system that are equally efficient statistically (as measured by the sum of squared residuals on observed trade flows). Within this class of models there is ambiguity regarding the size of cross-border trade frictions and thus ambiguity regarding the welfare effects of trade frictions. Our approach offers unique insights into structural estimation, and we highlight the importance of traditional calibration considerations when one uses econometric techniques to calibrate a model for comparative policy analysis.

We propose alternative estimations of the core general equilibrium. One that eliminates a bias in the structurally consistent fitted trade flows, and one that yields an R-squared of one. While more efficient than the A-vW estimation, our procedures are equally incomplete—they cannot identify the trade costs or price-response parameters without ad hoc identifying assumptions (or additional data). We conclude that the theoretic gravity model in question is useful as a descriptive tool, but its ability to match trade flows closely in calibration is not novel. Richer theories of trade (dependent on traditional comparative advantage or scale and variety effects) are routinely fit—exactly—to observed trade flows.[3]

It is not our intent to critique the theory behind the gravity model. We also do not advocate any particular approach to calibration. Our contribution is a bridge between the

[3]Myriad constant-returns general equilibrium models appear in the trade and tax policy literature [the early work is surveyed by de Melo (1988), and Shoven and Whalley (1984)]. Studies that incorporate new theories of scale or variety effects include Lòpez-de-Silanes et al. (1994) and Brown and Stern (1989).

methods and language used by modelers who calibrate systems regularly and those who use econometric methods.

In the following section we outline the general theoretic framework for the illustrative example, and we outline a general calibration strategy in terms of orders-of-approximation. In section three we recast the A-vW econometric method as a calibration, showing that it generates a complete set of structural parameters under particular identifying assumptions. A comparison of A-vW's restrictive case and the general set of equally efficient, or superior, approximations are examined in section four. Concluding remarks are offered in section five.

2 Theoretic and Conceptual Foundations for the Illustrative Example

2.1 The general equilibrium model

Following A-vW, consider a model in which representative consumers in each region have preferences over goods differentiated by region of origin. Goods are aggregated to a single region-specific variety. Each region is given an endowment of its variety, which it trades for foreign varieties.[4] Ad valorem trade frictions induce substitution into varieties that are from home and nearby regions.

We formulate the multi-region competitive exchange equilibrium as a Mixed Complementarity Problem (MCP). The MCP is essentially the Mathiesen (1985) formulation of an Arrow-Debreu general equilibrium [also, see Rutherford (1995b) for an introduction to MCPs]. The key advantage of using the Mathiesen formulation here is that it represents the entire multi-region general equilibrium compactly, as a set of $4n$ conditions in $4n$ unknowns (where n is the number of regions).[5]

[4] Non-traded goods are not considered by A-vW: the aggregate country specific output is fully traded. Anderson (1979), however, explores the implications of traded versus non-traded goods on the structure of the gravity equation.

[5] More generally, an MCP is a powerful numeric tool that directly accommodates complementary-slack

Income for a region, Y_i, is given by the product of its exogenous endowment, q_i, and the net (of trade-cost, or f.o.b.) price of output, P_i:

$$Y_i = P_i q_i. \tag{1}$$

The second set of conditions require market clearing for each region-specific product such that the quantity endowed equals the sum of demands by each region j:

$$q_i = \sum_j X_{ij} \left(P, Y_j, t_j, \gamma_j\right). \tag{2}$$

Demand for commodity i by region j (X_{ij}) is a function of P, the vector of f.o.b. prices across the regions; Y_j, the income level in region j; t_j, the vector of trade cost wedges faced by region j; and γ_j, a vector of structural parameters that identify region-j's preferences.

Consistent with expenditure minimization in each region, the third set of equilibrium conditions set the true-cost-of-living index, E_i, equal to the unit expenditure function:

$$E_i = e_i \left(P, t_i, \gamma_i\right), \tag{3}$$

where the utility function is assumed to be linearly homogeneous. The final set of conditions require balance between the nominal value of utility and income:

$$U_i E_i = Y_i. \tag{4}$$

Together conditions (1) through (4) are a complete multi-region general equilibrium that can be solved numerically for relative prices (P_i and E_i), regional welfare levels (U_i), and income levels (Y_i).[6] As an artifact of the equilibrium we can recover the individual bilateral trade

conditions that arise in economics [Rutherford (1995b)].

[6]The system (1) through (4) represents the equilibrium as it might be solved numerically. The $4n$

flows using the demand functions evaluated at the equilibrium:

$$X_{ij}^* = \left(P^*, Y_j^*, t_j, \gamma_j\right). \tag{5}$$

A convenient property of the equilibrium defined by (1) through (4) is that it is fully identified by endowments, trade frictions, and an explicit representation of the unit expenditure functions for each region.[7] Rarely are equilibrium systems dependent on such a modest set of unknown parameters.

2.2 Approximating the Utility Function

2.2.1 Functional-form Approach to Identification

There are many approaches to identifying utility (or production) functions. Most are familiar with the econometric approach—specific functional forms are chosen carefully because they include embedded identifying assumptions that limit the scope of structural parameters. Even with very simple functions, however, the number of identifying parameters often exceeds the degrees of freedom offered by the data. Data requirements are controlled through parsimonious assumptions about structural similarities across a section, and/or across time. The subsequent analysis of what happens *on average* is conditional upon certainty in the functional form, the identifying structural similarities across the sample, and the assumed error structure.

Calibrators also use specific functional forms to control the number of arbitrary identifying assumptions. The focus, however, on reporting idiosyncratic results leads calibrators away from assuming structural similarities across a section, and/or time. As an alternative,

unknowns are the Y_i, the U_i, the E_i, and the P_i. Only relative prices are determined, however, so one of the market clearance conditions is removed (by Walras's law) and we assign the associated price as the numeraire.

[7] The demand functions embedded in (2) also need to be identified, but it is a routine calculus exercise to recover these from the unit expenditure function.

calibrators make direct assumptions about key parameters, which are under identified in the system. The subsequent analysis is conditional upon the functional form, and direct parametric assumptions.

2.2.2 Orders-of-approximation Approach to Identification

In contrast to the functional form approach, we prefer to characterize the calibration or estimation procedure as a method of approximating a general unknown homogeneous function. From this perspective identifying assumptions cannot be hidden, and all functions are under identified from the perspective of the next higher order of approximation. This method is consistent with the *perturbation* method discussed by Judd (1996). In particular, we follow Perroni and Rutherford (1998), and Perroni and Rutherford (1995) in their approach to characterizing and analyzing flexible functional forms.

Consider identification of the primal utility function that is consistent with the expenditure function in the general equilibrium system (1)–(4)[8];

$$U_j = U_j\left(\frac{X_j}{t_j}, \gamma_j\right). \tag{6}$$

Utility in a region is a function of the vector of f.o.b. imports to that region scaled by the trade cost factors (X_j/t_j); this operationalizes the formulation of *iceberg* transportation costs.[9] The other argument in the utility function is the vector of structural parameters, γ_j.

Now consider that the demands for each commodity are chosen to maximize $U_j(\cdot)$ subject to income, prices, and distortions. The empirical challenge is to approximate $U_j(\cdot)$ by finding

[8]We go to the primal function here because it is more familiar. Our goal, however, is to characterize the expenditure function in condition (3). Of course, identification of the primal preference/technology directly implies the dual expenditure/cost function.

[9]Anderson and van Wincoop (2003) define the nominal trade flow from i to j as the product of the f.o.b. price, the trade cost factor, and the net quantity consumed $(P_i t_{ij} c_{ij})$. This also equals the product of the f.o.b. price and the export quantity $(P_i X_{ij})$. The net (of iceberg melt) quantity consumed is therefore X_{ij}/t_{ij}. We simply make this substitution and directly define utility as a function of the bilateral export quantities scaled by their respective trade costs.

Table 1: Orders of Approximation and Parametric Identification

Order of Approximation:	**Required Benchmark Estimates:**	**Identified Functional Parameters:**
Zero	Reference Quantities	Scale Parameters
First	Reference Prices	Share Parameters
Second	Reaction of Shares to Price Changes	Substitution Elasticities
Third	Reaction of Substitution Elasticities to Price Changes	Third-order Curvature Parameters

$\hat{U}_j\left(X_j/t_j, \hat{\gamma}_j\right)$, where the vector of estimated parameters, $\hat{\gamma}_j$, is determined by the data and the form of $\hat{U}_j(\cdot)$ is determined by the accuracy of the functional approximation.

A useful anchor for building an approximation of the function is the benchmark or reference point.[10] The benchmark point might be any observed solution to the consumer problem. At a solution to the consumer's programming problem, we can determine the position of a point on the *level set* (indifference curve) in goods space and its local slope (marginal rate of substitution). In Table 1 we describe this as a first-order approximation to the true function.

In the zero-order approximation the reference quantities establish the estimated scale of $\hat{U}_j(\cdot)$. If the function approximates a production process, the scale parameter establishes the units of output, but for utility functions scale is largely irrelevant (because we are only concerned with the ordinal evaluations). We note that in the general equilibrium above, condition (4) adopts a convenient cardinalization of each region's utility. This is arbitrary, however, and the analysis generalizes.[11]

[10]Also, the benchmark offers an observable point of departure for analysis or counterfactual simulations. We describe the benchmark using *reference values*. The econometric analog to these could be fitted values for dependent variables and observed data for independent variables.

[11]We define identical tastes by the resulting demand system not the particular scale of the utility function. This allows us to adopt a convenient region-specific cardinalization such that the product of the price index

The first-order approximation is completed by using the information on reference prices to establish the marginal rates of substitution local to the benchmark. Usually the first order approximation is best summarized by the identification of the estimated benchmark value shares ($\hat{\theta}_{ij}$ in this case). Note that we designate that a symbol is an estimate with a *hat* (i.e., $\hat{(\cdot)}$). For the model we analyze here the estimated value share of good i in country j's demand system local to the benchmark is given by

$$\hat{\theta}_{ij} = \frac{\hat{t}^0_{ij}\hat{P}^0_i\left(\hat{X}^0_{ij}/\hat{t}^0_{ij}\right)}{\sum_k \hat{t}^0_{kj}\hat{P}^0_k\left(\hat{X}^0_{kj}/\hat{t}^0_{kj}\right)} = \frac{\hat{P}^0_i\hat{X}^0_{ij}}{\sum_k \hat{P}^0_k\hat{X}^0_{kj}}. \tag{7}$$

Note that the terms on the right-hand side of (7) are designated as estimates, with a *hat*, and as benchmark levels with a superscript 0 (i.e., $(\cdot)^0$). Thus, $\hat{P}^0_i$ is the estimated f.o.b. price of output from country i at the benchmark. $\hat{X}^0_{ij}$ is the estimated reference export quantities at the benchmark. The benchmark reference trade distortions, $\hat{t}^0_{ij}$, are one plus the estimated ad valorem costs of moving goods from i to j. The distortions fall out of the benchmark value share, because, although the estimated reference price (gross of distortion) is $\hat{t}^0_{ij}\hat{P}^0_i$, the reference quantity that arrives for consumption is $\hat{X}^0_{ij}/\hat{t}^0_{ij}$. In this simple exchange economy the denominator of (7) is benchmark income in region j.

It is important to make a distinction between the estimated *parameters* ($\hat{P}^0_i$, $\hat{X}^0_{ij}$, and $\hat{t}^0_{ij}$), the model *variables* (P_i, and X_{ij}), and the exogenous *instruments* (t_{ij}). Notice also that the benchmark value share, $\hat{\theta}_{ij}$, is invariant because it only depends on estimated parameters. This is distinguished from the endogenous value shares, which may change away from the local benchmark.

Second-order approximations of $\hat{U}_j(\cdot)$ require estimates of how the input value shares change given price changes. This information is best summarized by a matrix of estimated

and the welfare index equals nominal income at the benchmark (this eliminates a scale parameter in either the income balance condition or the definition of the expenditure function). Identical tastes across regions, therefore, indicate identical ordinal evaluations.

substitution elasticities.[12] Standard functions used in economic and econometric models usually place restrictions on the form of the second-order approximation. The separable Constant-Elasticity-of-Substitution (CES) form, for example, restricts all cross-substitution elasticities to be equal (and invariant away from the benchmark).

The final row of Table 1 follows Perroni and Rutherford (1995), and Perroni and Rutherford (1998) in introducing the idea of third-order approximations to the function. Perroni and Rutherford highlight the need to consider third-order curvature properties in simulation analysis. They warn that traditional flexible functional forms used in econometrics, which accommodate a general matrix of cross-substitution elasticities (flexible second-order approximations), may not be appropriate in simulation analysis because of their out of benchmark properties. Perroni and Rutherford explain that choosing a functional form involves adopting an implicit conjecture about third-order properties. They propose a superior functional form that is both flexible (to any second-order curvature estimates) and globally tractable in out-of-benchmark experiments—the Nonseparable N-stage CES (NNCES) function.

2.3 Applying the Approximated Function

In their econometric study Anderson and van Wincoop (2003) adopt the CES functional form. We utilize this form to illustrate its place in our discussion of approximated functions and for our illustration of the econometric calibration. Again, adopting the CES form is consistent with a restricted second-order approximation, because all second-order curvatures must be identical across variety pairs. Furthermore, the third-order curvatures are consistent with preservation of constant second-order curvatures.

[12] These are usually represented as Allen-Uzawa cross-elasticities of substitution, but other second-order measures of curvature have been proposed. Perroni and Rutherford (1998) provide a summary of the important measures of second-order curvature.

The CES approximation to $\hat{U}_j(\cdot)$ is usually represented by

$$\hat{U}_j\left(\frac{X_j}{t_j},\hat{\gamma}_j\right)=\left(\sum_i \hat{\alpha}_{ij}\left(\frac{X_{ij}}{t_{ij}}\right)^{\frac{\hat{\sigma}_j-1}{\hat{\sigma}_j}}\right)^{\frac{\hat{\sigma}_j}{\hat{\sigma}_j-1}} \tag{8}$$

We drop the hat on subsequent representations of approximated functions to clean up the notation, but we maintain the hat on the estimated structural parameters (e.g., $\hat{\gamma}$). It is understood that the functions, because they are CES, are empirical approximations.

Rutherford (1995a) introduces a more convenient form of the CES function, the *calibrated share form*:

$$U_j\left(\frac{X_j}{t_j},\hat{\gamma}_j\right)=\hat{\phi}_j\left(\sum_i \hat{\theta}_{ij}\left(\frac{\hat{t}^0_{ij}X_{ij}}{t_{ij}\hat{X}^0_{ij}}\right)^{\frac{\hat{\sigma}_j-1}{\hat{\sigma}_j}}\right)^{\frac{\hat{\sigma}_j}{\hat{\sigma}_j-1}}, \tag{9}$$

where $\hat{\theta}_{ij}$ is the benchmark value share defined above. This form is a simple algebraic transformation that decomposes $\hat{\alpha}_{ij}$. It is more convenient for illustrating our method of approximation because it separately tracks the benchmark scale, benchmark share, and benchmark consumption components. The scale parameter, $\hat{\phi}_j$, is set equal to the benchmark utility level, which is benchmark income divided by the benchmark consumer price index, $\hat{\phi}_j=\hat{Y}^0_j/\hat{E}^0_j$.

The unit expenditure functions, under the CES approximation, are given by

$$e_j\left(P,t_j,\hat{\gamma}_j\right)=\hat{E}^0_j\left(\sum_i \hat{\theta}_{ij}\left(\frac{t_{ij}P_i}{\hat{t}^0_{ij}\hat{P}^0_i}\right)^{(1-\hat{\sigma}_j)}\right)^{1/(1-\hat{\sigma}_j)}. \tag{10}$$

The corresponding demand functions are

$$X_{ij}\left(P,Y_j,t_j,\hat{\gamma}_j\right)=\hat{X}^0_{ij}\left(\frac{t_{ij}}{\hat{t}^0_{ij}}\right)\left(\frac{Y_j}{\hat{\phi}_j e_j\left(P,t_j,\hat{\gamma}_j\right)}\right)\left(\frac{\hat{t}^0_{ij}\hat{P}^0_i e_j\left(P,t_j,\hat{\gamma}_j\right)}{t_{ij}P_i\hat{E}^0_j}\right)^{\hat{\sigma}_j}. \tag{11}$$

In computational applications the calibrated share form is preferable because it provides a simple parameter and functional check that is independent of second-order curvature.

With the expenditure function specified in (10), and provided estimates of each parameter, the general equilibrium is operational as a comparative static or simulation tool. We simply substitute (10) and (11) into (3) and (2) respectively. Table 2 lists all of the symbolic elements of the general equilibrium. Again, the equilibrium consists of $4n$ variables. There are $n + n^2$ exogenous instruments that characterize endowments and trade frictions. Parameter requirements include $3n + 2n^2$ primitives that represent the fitted benchmark and an additional n substitution elasticities.[13] Many of these parameters are routinely eliminated or identified by assumption. In the following section we show how adopting A-vW's simplifying assumptions and least-squares procedure identifies the elements of $\hat{\gamma}$.

3 Econometric Calibration and General Equilibrium Simulation

This section illustrates how the estimation made by Anderson and van Wincoop (2003) can be interpreted as a calibration of the general equilibrium. First, we outline the assumptions and least-squares procedure necessary to arrive at a full set of benchmark calibration parameters from the data. The data include trade flows among U.S. states and Canadian provinces, state and province incomes, and bilateral distances.[14] Second, once the parameters are established we bring them to the original model to complete the calibration procedure. We

[13]We do not include the familiar summary parameters in our count of primitive parameters. In general, theoretic and econometric exercises focus on these summary parameters, where as calibration exercises focus on their primitives (the estimated benchmark fitted values). There are $2n + n^2$ summary parameters: the zero-order $\hat{\phi}_i$, the first-order $\hat{\theta}_{ij}$, and the second-order $\hat{\sigma}_i$. This obscures the fact that many more primitive estimates are needed to identify the summary parameters, and our count assumes that each of the substitution elasticities has only a single underlying primitive. Realistically, the count of primitives should be expanded to include multiple first-order fitted observations (a minimum of two different observations of the first-order information are needed to identify the $\hat{\sigma}_i$).

[14]Our illustrative example follows A-vW's two-country application, which assumes that the U.S. and Canada are the only countries and that states and provinces are the relevant geographic divisions for region specific varieties.

Table 2: Scope of the A-vW General Equilibrium with CES Preferences

Variables:			Instruments:			Structural Parameters ($\hat{\gamma}$):		
Y_i	=	Incomes	q_i	=	Endowments	$\hat{Y}_i^0$	=	Benchmark incomes
E_i	=	Unit expenditure index	t_{ij}	=	Trade frictions	$\hat{E}_i^0$	=	Benchmark unit expenditure index
P_i	=	Prices (f.o.b.)				$\hat{\phi}_i^0$	=	$\hat{Y}_i^0/\hat{E}_i^0$ (zero-order summary parameters)
U_i	=	Utility levels				$\hat{X}_{ij}^0$	=	Benchmark trade flows
						$\hat{P}_i^0$	=	Benchmark prices (f.o.b.)
						$\hat{t}_{ij}^0$	=	Benchmark trade frictions
						$\hat{\theta}_{ij}$	=	$\left(\hat{P}_i^0\hat{X}_{ij}^0\right)/\hat{Y}_j^0$ (first-order summary parameters)
						$\hat{\sigma}_i$	=	Elasticities of substitution (second-order summary parameters)

purposefully return to the extensive-form represented in conditions (1)-(4). The final step is to perform the counterfactual of interest: the integration of the U.S. and Canadian economies by removing the effect of the border.

3.1 Parametric Identification

Conditions (1)-(4) are useful as an empirical tool only after the elements of $\hat{\gamma}$ are estimated using data. What follows is an illustration of how one might arrive at a set of parameters consistent with the set derived by A-vW, but from the perspective of a calibration exercise. The data are limited so some of the elements must be identified directly through parametric or structural assumptions.

As a first step in reducing data requirements, we identify some parameters by simply making the benchmark price normalization explicit. Estimates of income at the benchmark, $\hat{Y}_i^0$, and exchange rates are widely published so it seems logical to adopt the convention that common nominal units equal real units at the benchmark. This allows us to set $q_i = \hat{Y}_i^0$ for each region. Income and, therefore, endowment quantities are in units of US dollars (at the benchmark).[15] Given our choice of units and by the first equilibrium condition $\hat{P}_i^0 = 1$ for all i, which is a convenient normalization. Any other convention for measuring real units can be adopted without loss of generality.

More subtly, however, even in this seemingly innocent assumption we have implicitly asserted that there is no error associated with the measurement of income or its conversion into common nominal units. Our failure to account for uncertainty in these underlying procedures may not bias our subsequent analysis but seriously undermine the validity of statistical inference. After all, many degrees of freedom have been consumed in the generation of what we accept as primary data. Following the tradition in most of empirical economics we acknowledge this problem and ignore it.[16]

One simplification that does have clear general implications involves a further restriction and ad hoc identification of second-order curvature. Following A-vW we simply adopt a single assumed value for $\hat{\sigma}_i$ for all of the regions ($\hat{\sigma}_i = \sigma$). The results from any analysis using the model are thus conditional on the assumed value of σ (and its constancy across regions). The specific value of σ adopted is based on a casual survey of the econometric literature that specifically focuses on measuring this parameter.[17]

Another convenient simplification is to assume symmetric trade costs, that is, $\hat{t}_{ij}^0 = \hat{t}_{ji}^0$. This assumption decreases the number of parameters that need to be identified. More importantly for A-vW it enables them to arrive at a relatively simple benchmark reduced

[15] It is important to emphasize that the benchmark price normalization is only valid locally. Away from the benchmark we normalize on the price of a single region's endowment commodity, this commodity serves as numeraire.

[16] Subsequently we remove the *hat* on income because these are assumed to be primary data rather than estimates (that is $\hat{Y}_i^0 = Y_i^0$). We still refer to Y_i^0 as a structural parameter, however; we simply assume its accuracy to be perfect.

[17] Our arguments in this paper do not depend qualitatively on the value of σ, but the reader is warned that subsequent quantitative illustrations depend on $\sigma = 5$. This is the value Anderson and van Wincoop (2003) prefer based on their reading of the literature.

form.[18]

Additionally, the trade cost is assumed to follow a log-linear form such that,

$$\hat{t}^0_{ij} = \hat{b}^0_{ij} d^{\hat{\rho}}_{ij}. \tag{12}$$

The estimated border cost factor, represented by $\hat{b}^0_{ij}$, equals one plus the tariff equivalent. The observed bilateral distance is given by d_{ij}, and $\hat{\rho}$ is the elasticity of the total cost factor with respect to distance. Although convenient for A-vW's theory, the form of (12) is another ad hoc approximation.[19] Notice that (12) indicates that absolute trade costs will be affected by the units in which distance is measured, which is potentially problematic. The indeterminacy is resolved by choosing an ad hoc normalization.

A key identifying, and restrictive, assumption of the econometric calibration is one of identical tastes across regions.[20] This significantly reduces the number of parameters needed.

[18]A-vW explain that although symmetry is assumed, the econometric model cannot distinguish between this equilibrium and one in which there are asymmetries that produce the same average trade resistance [see Anderson and van Wincoop (2003), footnote 11].

[19]For example, Hummels (2001) argues that an additive form is more sensible. As with many theories that support the gravity literature, the origin of (12) is more likely the log-linear regression, not the most plausible micro foundations.

[20]In deriving their reduced form A-vW assume homogeneity in both relative and absolute tastes across regions (the cardinalization of utility is maintained across regions). Subsequently, A-vW suggest structural taste bias as an alternative to homogeneity: we explore this suggestion in Section 4. For the equations here to be consistent with absolute taste homogeneity, one could normalize the utility functions by a positive monotonic transformation. For example, multiplying (9) by

$$\hat{z}_j = \sum_i \frac{\hat{\theta}_{ij}}{\hat{\phi}_j} \left(\frac{\hat{t}^0_{ij}}{\hat{X}^0_{ij}} \right)^{(\sigma-1)/\sigma}$$

normalizes the scale of utility across regions. With this modification, the equilibrium condition (4) becomes $U_i E_i / \hat{z}_i = Y_i$. Relative homogeneity in utility is achieved by holding the distribution parameters constant across the regions. For example,

$$\hat{\alpha}_i = \hat{\alpha}_{ij} = \frac{\hat{\theta}_{ij}}{\hat{z}_j} \left(\frac{\hat{t}^0_{ij}}{\hat{X}^0_{ij}} \right)^{(\sigma-1)/\sigma}, \forall j.$$

In subsequent analysis we adopt A-vW's reduced form for calibration; noting that their cardinalization is different than ours, but also noting that the resulting demand systems and, therefore, the extensive forms are isomorphic.

A-vW utilize this assumption to derive a reduced form that is free of share parameters. Essentially, under identical tastes one can use the information that in a frictionless world each region j would consume its income share of each regional commodity i. The estimates of the distribution parameters in the CES functions are thus replaced by information on local (benchmark) income shares.

With the simplifications established A-vW derive the reduced form for the local unit expenditure functions,

$$\hat{E}_j^0 = \left[\sum_i \frac{Y_i^0}{\sum_k Y_k^0} \left(\frac{\hat{b}_{ij}^0 d_{ij}^{\hat{\rho}}}{\hat{E}_i^0} \right)^{(1-\sigma)} \right]^{1/(1-\sigma)} ; \tag{13}$$

and the local nominal demand functions

$$\hat{X}_{ij}^0 = \frac{Y_i^0 Y_j^0}{\sum_k Y_k^0} \left(\frac{\hat{b}_{ij}^0 d_{ij}^{\hat{\rho}}}{\hat{E}_i^0 \hat{E}_j^0} \right)^{(1-\sigma)} . \tag{14}$$

Note, given our benchmark price normalization, the nominal flow $(\hat{P}_i^0 \hat{X}_{ij}^0)$ equals the real flow $(\hat{X}_{ij}^0)$ at the benchmark. Thus we have simplified (14) by removing $\hat{P}_i^0$ from the left-hand side. With some slight rearranging and in log form (14) becomes the structural gravity equation presented by A-vW:

$$\ln \left(\frac{\hat{X}_{ij}^0}{Y_i^0 Y_j^0} \right) = k + (1-\sigma)\hat{\rho} \ln d_{ij} + (1-\sigma)(1-a_{ij}) \ln \hat{b}^0 - \ln \hat{E}_i^{0(1-\sigma)} - \ln \hat{E}_j^{0(1-\sigma)} \tag{15}$$

where k is a regression constant that should approximate $-\ln \sum_k Y_k^0$. The dummy variable a_{ij} is one if the shipment from i to j does not involve a border. The estimate $\hat{b}^0 - 1$ is interpreted as the ad valorem tariff equivalent of the border cost [where $\hat{b}_{ij}^0 = (\hat{b}^0)^{(1-a_{ij})}$].

Direct measures of some of the trade flows (X_{st}^0) are observed. The regional pair (s,t)

is a member of the set of potential bilateral pairs; $\{(s,t)|s \in I, t \in J\}$.[21] The X^0_{st} are assumed to be primary data, but rampant with unbiased measurement error. To reconcile (15) with these observations A-vW define an objective function that is the sum of squared errors between the direct observations and the model prediction. Calibration involves the minimization of this objective subject to the reduced form of the unit expenditure functions, (13). The non-linear programming problem is

$$\begin{array}{lll} \min_{\{k,\beta_1,\beta_2\}} & \sum_s \sum_t & \left[\ln\left(\frac{\hat{X}^0_{st}}{Y^0_s Y^0_t}\right) - k - \beta_1 \ln d_{st} - \beta_2(1-a_{st})\right. \\ & & \left. + \ln \hat{E}^{0^{(1-\sigma)}}_s + \ln \hat{E}^{0^{(1-\sigma)}}_t \right]^2 \\ \text{subject to:} & & \hat{E}^{0^{(1-\sigma)}}_j = \sum_i \left(\frac{Y^0_i}{\sum_k Y^0_k} \hat{E}^{0^{(\sigma-1)}}_i e^{\beta_1 \ln d_{ij} + \beta_2(1-a_{ij})}\right) \forall j. \quad \text{(NLP1)} \end{array}$$

The coefficient β_1 equals $(1-\sigma)\hat{\rho}$, and the coefficient β_2 equals $(1-\sigma)\ln \hat{b}^0$. Estimation requires observations on distance for each (i,j) pair and the construction of the border dummy for each pair. Given the set of assumptions necessary to arrive at the reduced form and the solution values for β_1, β_2, and the $\hat{E}^{0^{(1-\sigma)}}_j$ from the programming problem we can recover all of the elements of $\hat{\gamma}$. Comparative static experiments can then be simulated via the equilibrium conditions.

A key insight our calibration perspective offers is the absence of variables in the reduced-form model, characterized in equations (13) and (14). The estimation (calibration) procedure brings the data to the model and is, therefore, devoid of elements endogenous to the model. The reduced-form equations are not general because they rely on identifying assumptions that may only be valid local to benchmark. For example, A-vW's empirical model depends on a particular local normalization [Anderson and van Wincoop (2003), footnote 12]. This normalization is integral to the reduced-form equations, but not the economic model. Applying the reduced-form equations in a counterfactual implicitly renormalizes the problem,

[21] Anderson and van Wincoop (2003) reduce the full set of bilateral pairs by: flows that do not have observations, flows that are observed to be zero, and any measured internal flows.

contaminating inferences [Balistreri (2005)].

3.2 Completing the Calibration

Although all of the information has been processed and each identifying assumption itemized, calibration is only complete once the data has been brought back to the extensive-form theory. This removes any superfluous simplifications that may only apply to the benchmark reduced-form model. It must be noted that, although the least-squares programming problem is consistent with the general equilibrium at a local point of estimation, the econometric model does not represent an operational economic model.[22] The data should be brought back to the more general theory for proper analysis.

Given σ, the benchmark price indices are recovered from the direct estimates of $\hat{E}_j^{0^{(1-\sigma)}}$. The $\hat{t}_{ij}^0$ are recovered by using (12) and the estimated distance and border coefficients. Following Balistreri and Hillberry (2006), distance is normalized to minimize the resources devoted to melt, subject to no negative trade costs (the minimum estimated $\hat{t}_{ij}^0$ is normalized to one). The scale on distance has no relative implications, but it does change the level of the estimated benchmark price indexes and measures of resources devoted to trade costs.[23]

The estimated fitted benchmark trade flows are calculated using equation (14). Notice that we cannot use (15) to recover the fitted values if k is inconsistent with observed incomes.[24] The share and scale parameters are calculated as they are defined in Table 2. All

[22]Balistreri (2005) shows the divergence, away from the local point of estimation, of A-vW's gravity model and their proposed general equilibrium.

[23]Under an agnostic stance on absolute trade costs the benchmark general equilibrium is unidentified, and the regression loses its ability to inform the structure. The estimating system, equation (15) subject to (13), is consistent with the general equilibrium at the benchmark, but this system alone cannot inform us about the economic variables away from that benchmark. Moving from regression coefficients to an operational economic model requires a normalization that defines the absolute sizes of the $\hat{t}_{ij}^0$. Our normalization is based on the illogic of negative distance-related costs.

[24]Anderson and van Wincoop (2003) do not report information obtained on the estimated constant k, which is inevitably inconsistent with its structural interpretation given direct measures of benchmark income. Empirical estimates of k from the least-squares problem imply an aggregate income that is over 3.5 times larger than observed (by summing across the Y_i^0). If one were interested in testing the theory rather than calibrating the theoretic model, the model is easily rejected based on a hypothesis test of the similarity

of the items in $\hat{\gamma}$ are thus identified.

3.3 Counterfactual Analysis

Counterfactual simulations are computed by evaluating the system with different endowments or trade costs.[25] Specifically, we compute the general equilibrium with border frictions eliminated, such that $t_{ij} = d_{ij}^{\hat{\rho}}$. This simulates integration of the Canadian and US economies. Table 3 presents the numeric benchmark equilibrium and percent changes in each of the variables using A-vW's central elasticity ($\sigma = 5$), A-vW's coefficient estimates ($\beta_1 = -0.79$ and $\beta_2 = -1.65$), and choosing Alabama output as the numeraire. Utility is numeraire independent, but changes in income and the prices depend explicitly on our arbitrary choice of numeraire (the model is homogeneous degree zero in prices).

To analyze the effect of border removal on international versus intranational trade we recover the disposition of each region's output (q_i) using the demand equations. The endowed quantity for each region is either consumed in the United States, consumed in Canada, or it *melts* in transit. Table 4 presents these results.

Before moving to further analysis we point out a general inconsistency between our counterfactual results and those that appear in Anderson and van Wincoop (2003). Possibly because they do not view their estimation as a method of calibrating the general equilibrium, A-vW do not control the price normalization in their measures of counterfactual multilateral resistance, and do not account for the effects of substantial income changes on nominal trade flows. A-vW calibrate the taste parameters correctly from the benchmark fitted values but apply them to the reduced-form import demand and expenditure functions [see Appendix B of Anderson and van Wincoop (2003)]. We highlight the importance of returning to

between k and its theory consistent value, $-\ln \sum_i Y_i^0$. The implications of these inconsistencies are further explored in section 4.

[25]The relatively small non-linear system (of only 164 equations) is solved using PATH, a complementarity problem algorithm, available in the GAMS software. The code is available upon request.

Table 3: Benchmark Equilibrium and Simulated Border Removal ($\sigma = 5$)

	Income (Y_i)		Cost of Living Index (E_i)		Price of Output (P_i)		Utility (U_i)	
	Benchmark ($US billion)	Percent Change	Benchmark	Percent Change	Benchmark	Percent Change	Benchmark	Percent Change
Province								
Alberta	56.3	14%	2.1	-16%	1.0	14%	26.5	36%
British Columbia	62.9	10%	1.9	-12%	1.0	10%	32.9	26%
Manitoba	16.7	19%	2.1	-20%	1.0	19%	7.8	49%
New Brunswick	9.5	16%	2.1	-18%	1.0	16%	4.6	42%
Newfoundland	6.4	19%	2.4	-20%	1.0	19%	2.7	50%
Nova Scotia	12.4	16%	2.1	-17%	1.0	16%	5.9	39%
Ontario	194.3	14%	1.8	-16%	1.0	14%	107.3	35%
Prince Edward Island	1.6	17%	2.2	-19%	1.0	17%	0.7	45%
Quebec	107.1	12%	1.9	-14%	1.0	12%	57.5	29%
Saskatchewan	15.3	18%	2.2	-19%	1.0	18%	7.1	46%
State								
Alabama	83.0	0%	1.6	-1%	1.0	0%	53.2	1%
Arizona	85.0	0%	1.8	-1%	1.0	0%	48.4	1%
California	843.1	0%	1.4	-1%	1.0	0%	595.1	0%
Florida	300.7	0%	1.5	-1%	1.0	0%	195.1	1%
Georgia	170.9	0%	1.5	-1%	1.0	0%	111.9	1%
Idaho	22.4	1%	1.8	-2%	1.0	1%	12.7	2%
Illinois	312.3	0%	1.5	-1%	1.0	0%	210.3	1%
Indiana	129.7	0%	1.5	-1%	1.0	0%	86.7	1%
Kentucky	79.9	0%	1.5	-1%	1.0	0%	52.8	1%
Louisiana	94.7	0%	1.6	-1%	1.0	0%	59.0	1%
Massachusetts	174.0	0%	1.5	-1%	1.0	0%	119.6	1%
Maryland	124.6	0%	1.4	-1%	1.0	0%	88.0	1%
Maine	25.1	1%	1.6	-2%	1.0	1%	15.2	3%
Michigan	217.3	0%	1.5	-1%	1.0	0%	142.6	2%
Minnesota	114.6	0%	1.6	-1%	1.0	0%	71.1	2%
Missouri	118.3	0%	1.5	-1%	1.0	0%	76.4	1%
Montana	16.1	1%	1.8	-2%	1.0	1%	9.0	3%
North Carolina	168.6	0%	1.5	-1%	1.0	0%	110.3	1%
North Dakota	12.7	1%	1.7	-2%	1.0	1%	7.4	2%
New Hampshire	27.2	1%	1.6	-2%	1.0	1%	17.5	2%
New Jersey	243.9	0%	1.4	-1%	1.0	0%	168.2	1%
New York	541.1	0%	1.4	-1%	1.0	0%	386.7	1%
Ohio	256.6	0%	1.5	-1%	1.0	0%	174.0	1%
Pennsylvania	283.1	0%	1.4	-1%	1.0	0%	196.7	1%
Tennessee	116.7	0%	1.5	-1%	1.0	0%	75.9	1%
Texas	453.0	0%	1.6	-1%	1.0	0%	283.7	1%
Virginia	170.0	0%	1.5	-1%	1.0	0%	114.1	1%
Vermont	13.0	1%	1.6	-2%	1.0	1%	8.2	3%
Washington	136.4	1%	1.7	-2%	1.0	1%	82.6	3%
Wisconsin	117.7	0%	1.6	-1%	1.0	0%	75.3	1%
Rest-of-US	988.6	0%	1.8	-1%	1.0	0%	552.5	2%

Table 4: Disbursement of Endowments ($\sigma = 5$)

	Endowment q_i	Consumed in Canada: Bench-mark	Consumed in Canada: Counter-factual	Consumed in the US: Bench-mark	Consumed in the US: Counter-factual	Transport *Melt*: Bench-mark	Transport *Melt*: Counter-factual
Province							
Alberta	56.3	19.3	5.7	4.1	15.7	32.9	34.9
British Columbia	62.9	33.0	12.8	3.9	17.6	26.1	32.5
Manitoba	16.7	4.6	1.0	1.7	5.4	10.5	10.3
New Brunswick	9.5	3.0	0.8	0.9	3.0	5.6	5.7
Newfoundland	6.4	1.5	0.3	0.6	1.8	4.3	4.2
Nova Scotia	12.4	4.4	1.2	1.0	3.7	7.0	7.5
Ontario	194.3	79.4	24.0	16.5	64.7	98.5	105.7
Prince Edward Island	1.6	0.5	0.1	0.2	0.5	1.0	1.0
Quebec	107.1	47.7	16.7	7.3	31.4	52.1	59.0
Saskatchewan	15.3	4.3	1.0	1.4	4.7	9.6	9.6
State							
Alabama	83.0	0.4	1.6	37.7	36.3	45.0	45.1
Arizona	85.0	0.4	2.0	32.9	31.4	51.6	51.6
California	843.1	2.1	9.8	548.4	537.4	292.6	296.0
Florida	300.7	1.2	5.2	147.0	142.4	152.6	153.1
Georgia	170.9	0.8	3.4	81.1	78.2	89.0	89.3
Idaho	22.4	0.2	0.8	8.4	7.8	13.8	13.7
Illinois	312.4	1.5	6.8	153.8	148.1	157.0	157.5
Indiana	129.7	0.8	3.4	61.5	58.8	67.5	67.5
Kentucky	79.9	0.5	2.1	36.8	35.2	42.7	42.7
Louisiana	94.7	0.4	1.8	42.3	40.8	52.0	52.1
Massachusetts	174.0	1.3	5.7	105.6	100.2	67.2	68.1
Maryland	124.6	0.7	3.1	75.9	72.9	48.0	48.5
Maine	25.1	0.4	1.6	10.4	9.4	14.3	14.1
Michigan	217.4	1.9	8.6	99.3	93.2	116.1	115.6
Minnesota	114.6	0.8	3.6	48.6	45.9	65.2	65.1
Missouri	118.3	0.6	2.7	53.6	51.4	64.2	64.2
Montana	16.1	0.2	0.7	5.8	5.4	10.1	10.0
North Carolina	168.6	0.9	4.2	80.6	77.2	87.0	87.2
North Dakota	12.7	0.1	0.5	4.7	4.4	7.9	7.8
New Hampshire	27.2	0.3	1.3	12.8	11.9	14.1	14.0
New Jersey	243.9	1.6	7.4	132.3	126.0	109.9	110.5
New York	541.1	3.9	17.7	323.3	307.4	213.9	216.1
Ohio	256.6	1.8	7.9	126.8	120.6	128.1	128.2
Pennsylvania	283.1	2.1	9.2	152.5	144.9	128.6	129.0
Tennessee	116.7	0.6	2.6	52.6	50.5	63.5	63.5
Texas	453.0	1.7	7.7	205.1	198.5	246.2	246.9
Virginia	170.0	1.0	4.5	85.9	82.1	83.1	83.4
Vermont	13.0	0.2	0.8	5.6	5.1	7.2	7.1
Washington	136.4	1.7	8.1	69.3	63.1	65.4	65.1
Wisconsin	117.7	0.8	3.6	51.4	48.7	65.5	65.4
Rest-of-US	988.6	6.8	30.0	348.9	328.6	632.9	630.1

the extensive form for counterfactual analysis. Unlike A-vW's method, our extensive-form method requires the choice of a numeraire and clearly delineates the *parameter*, benchmark income, from the endogenous *variable*, income.

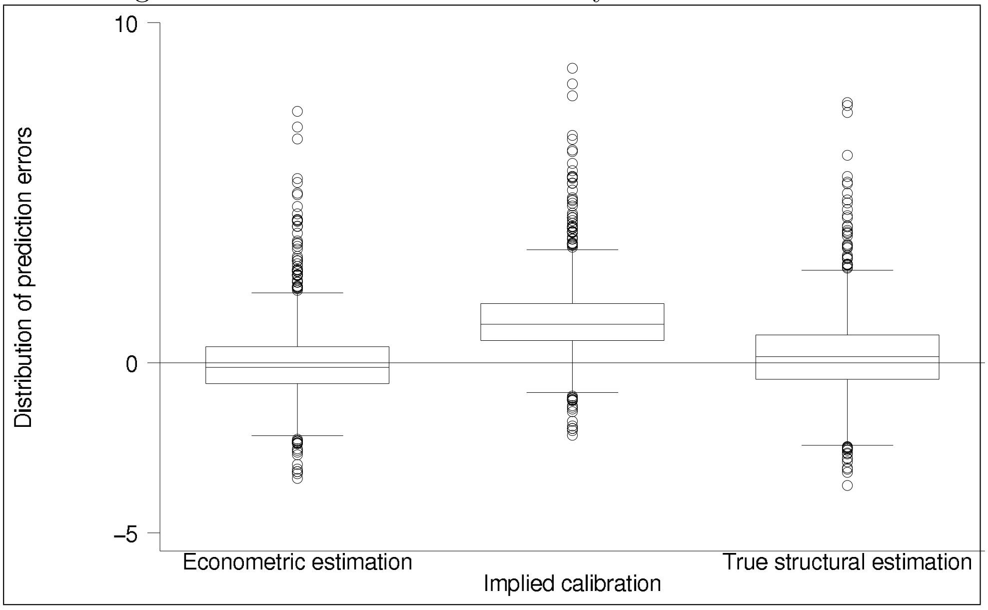
4 Implications for the Calibrated Model and Alternative Calibrations

In the context of the econometric application, we highlight three lessons illuminated by our calibration perspective. First, structural integrity must be maintained throughout the calibration. We show that a lack of discipline concerning the regression intercept generates a bias in the structural fitted flows and taints the structural interpretation of other regression coefficients. Second, we take a partial step toward traditional calibration techniques by operationalizing A-vW's proposed inclusion of home bias in tastes. This consumes at least one degree of freedom (outside of the regression analysis) and generates an unidentified first-order approximation. Information on trade costs must be added for an operational system. Our third point is that, if some estimates of trade costs are required (because the regression does not identify these), the econometric method uses the available information inefficiently. A more efficient calibration is achieved using the traditional calibration method to achieve an exact fit on observed trade flows.

4.1 Calibration Bias and Stochastic Efficiency

Equations (14) and (15) highlight a key difference between the economic model and the econometric procedure proposed by A-vW. The economic theory motivating the regression equation imposes a restriction on the regression constant—it should equal $-\ln \sum_i Y_i^0$. Failure to impose this restriction biases econometric estimates relative to the theory, which requires equality between the value of aggregate demand and aggregate income. The bias manifests itself in the different implied benchmark trade flows depending on whether we compute them

Figure 1: Tradeoff between efficiency and structural bias



using equation (14) or (15). The theory cannot accommodate a free intercept, k, without violating a key adding up requirement.[26]

Figure 1 illustrates the tradeoffs between calibration bias and stochastic efficiency. We generate the log difference between the fitted and actual trade flows for three sets of fitted flows.[27] Columns one through three show the error distributions for the econometric procedure, a calibration based on estimates generated by the econometric exercise, and a theory consistent estimation/calibration, respectively.[28]

The figure demonstrates two points. First, structurally consistent estimation can reduce

[26]Unlike A-vW's method, some traditional econometric methods of estimating demand systems (e.g., the Linear Expenditure System, or the Almost Ideal Demand System) automatically impose adding up [see Deaton and Muellbauer (1980)]. Adding up is only achieved in the A-vW estimation by restricting the regression intercept.

[27]It is important to note that the log difference in trade flows is not the residual being minimized in the regression, see equation (15). The log difference in trade flows is the more common residual in the broader gravity literature.

[28]The plots follow the conventions of STATA software. The central line within each box represents the median value of the distribution. The box includes estimates within the interquartile range (those between the 25th and 75th percentile). The whiskers extend beyond the box in each direction at a distance of 1.5 times the interquartile range. Observations outside of the whiskers are outliers represented as individual data points.

econometric efficiency when theory imposes restrictions on the parameter estimates. The distribution of errors under structurally consistent estimation (column 3) is larger than in econometric estimation (column 1). It is notable, however, that the econometric efficiency loss that occurs when the regression constant is restricted to its theory consistent value is small.

Second, parameter estimates taken from estimation procedures inconsistent with the theoretic structure of an economic model generate bias in subsequent calibrations that use the parameter estimates as inputs. The distribution of error terms in the implied calibration (column 2) is identical to that of the econometric estimation, it is simply shifted higher. By allowing the estimation model to overstate aggregate income, the econometric procedure understates the trade costs necessary to fit trade flows within the model. A calibration based on trade costs from the estimating procedure substantially overstates interregional trade.

In contrast, the error terms from theory-consistent estimation (a constrained regression constant) are replicated in calibration.[29] The error distribution in column 3 is identical, whether one uses a structurally consistent estimation procedure or a calibration using structural parameters generated by structurally consistent estimation. It is for this reason that we advocate estimation procedures that are fully consistent when the goal is parameterizing a model for subsequent counterfactual analysis. We also emphasize that structural estimation is not a new procedure, it is simply calibration renamed.

4.2 Systematic Taste Bias

A-vW caution that we should not accept their estimates of $\hat{E}_i^0$ in their literal interpretation as consumer price indices Anderson and van Wincoop (2003, page 176, and footnote 17), because measured trade resistance, attributed to borders and distance, need not indicate a

[29]The regression coefficients change substantially when structure is imposed on the constant term: $\beta_1 = -1.44$ and $\beta_2 = -1.85$ under true structural estimation.

price markup.[30] A-vW propose home bias in preferences as a source of trade resistance. We adopt A-vW's suggestion of structural taste bias, which is consistent with their regression analysis. In this context the first order calibration is underidentified. The ambiguity in the system is only resolved through a direct measure of the trade frictions that identifies these frictions independent of geographic regularities in the CES distribution parameters.

To illustrate the importance of our assumption about taste bias consider a new benchmark parameter, $\hat{\lambda} \in [0,1]$, that represents the proportion of measured border resistance that is *not* due to taste bias.[31] The only modifications are to our definition of the reference distortions. Equation (12) becomes

$$\hat{t}_{ij}^0 = [1 + \hat{\lambda}(\hat{b}_{ij}^0 - 1)]d_{ij}^{\hat{\rho}}. \tag{16}$$

At the benchmark the instruments must be set to their estimated local values ($t_{ij} = \hat{t}_{ij}^0$), and, as before, the counterfactual involves a computation of the equilibrium under $t_{ij} = d_{ij}^{\hat{\rho}}$.

Using (16) to compute the benchmark distortions fully operationalizes the structural taste bias suggested by Anderson and van Wincoop (2003). The advantages of the calibrated-share form should be noted: no modifications are required in our functional representation of the equilibrium (the benchmark value shares are unchanged and there is no need to change the scaling). The benchmark distortions and value of the instruments change, but the calibrated-

[30]The $\hat{E}_i^0$ are always interpreted as the price of the composite good (the composite good is regional units of utility), but its value relative to the benchmark f.o.b. prices of endowments depends on the arbitrary scale of utility. To interpret this measure literally as a consumer price index the benchmark units of the composite good need to be comparable with endowment units (another special cardinalization of utility). In this case when trade frictions that cause an f.o.b. to c.i.f price wedge (pecuniary costs as A-vW call them) are removed the consumer price indices revert to unity. Like other issues relating to the cardinalization of utility, the interpretation of $\hat{E}_i^0$ as a consumer price index, as opposed to the relative price of utils, is largely irrelevant in the more general discussion of regularities in trade resistance due to borders and distance. A-vW's point is that some portion of the resistance embodied in the measured $\hat{t}_{ij}^0$ might be due to things other than transport or border policy that vary with distance and country respectively. Our conclusions from an earlier paper support this, more general, interpretation of trade resistance [Balistreri and Hillberry (2006)].

[31]We only examine border-related taste bias, but it is also reasonable, and prudent, to think that distance related resistance also includes a taste component. We concentrate on the border bias because of its relevance to welfare effects in the counterfactual simulation of border removal. For this illustration we also make a stark simplification that the portion of border costs that are due to taste bias are constant across regions.

share form allows us to directly input these changes to the first-order information without the tedious recalibration of the scale dependent distribution parameters in the standard form represented in (8).

It is apparent from (16) that $\hat{\lambda}$ (or equivalently, a direct measure of the border cost) will determine the magnitude of the price changes embodied in the counterfactual, and σ determines the responses to those price changes. As a method of second-order approximation, the A-vW econometric method is silent on the two most important pieces of information for welfare analysis ($\hat{\lambda}$ and σ). Welfare analysis depends critically on these parametric assumptions, which cannot be informed by the proposed regression.[32]

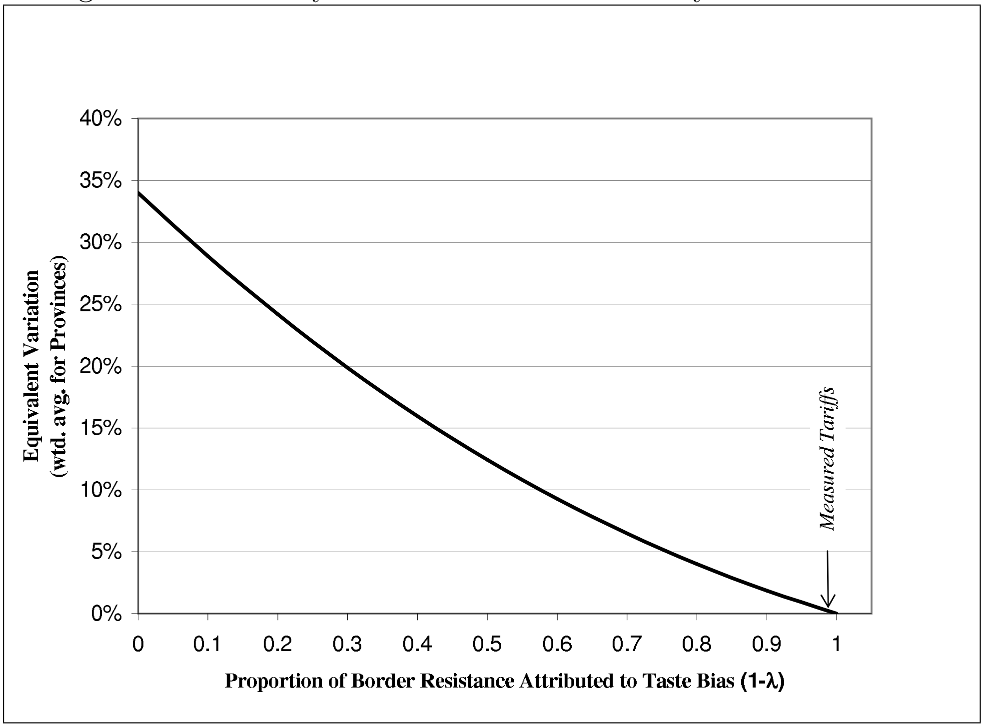
We illustrate the importance of good measures of $\hat{\lambda}$ in Figure 2. If one assumes no taste bias the welfare effects of border removal are maximized conditional on a given σ. At the values preferred by Anderson and van Wincoop (2002), of $\sigma = 5$ and $\hat{\lambda} = 1$ (no taste bias), the tariff equivalent of the border charge is 51%. On the other extreme, if we simply measure duties collected divided by trade flows the ad valorem rate in 1993 was less than 0.4%.[33] Using this rate as the border charge (and $\sigma = 5$) the proportion of border resistance attributed to taste bias is 99% (marked in the figure as *Measured Tariffs*). Using measured tariffs the welfare impact on Canada of removing the border is only 0.1%. Including non-tariff barriers in the calculated border charge will likely place $\hat{\lambda}$ in the middle of its range. Our point is that freeing $\hat{\lambda}$ implies that trade flow data cannot be used in the gravity framework to infer trade costs.

Thus, we contend that the econometric calibration outlined in Section 3 does not complete even the first-order approximation to the general equilibrium because it fails to quantify key first-order information on the benchmark trade costs (or equivalently, it fails to quantify

[32]The large welfare benefits of integration advertised by Anderson and van Wincoop (2002) follow from the assumption that all measured trade resistance at the border is a border charge.

[33]Measured tariffs are Calculated from the USITC Interactive Tariff and Trade DataWeb. This illustrative calculation only measures the ad valorem rate on U.S. imports from Canada.

Figure 2: Sensitivity of Canadian Welfare Analysis to Taste Bias



the taste bias). The reduced-form estimation is consistent with the broad class of models covered on the taste-bias interval, $\hat{\lambda} \in [0, 1]$. All of these calibrations are equally as efficient as the A-vW procedure, when measured by the sum of squared residuals.

4.3 Improved Stochastic Efficiency in a Generalized Calibration

Once one admits that taste-bias must be considered—in order to allow the model to better approximate a world with production specialization and non-traded goods, for example—there is no particular need to add the restrictions imposed by the stochastic form of the regression. The regression no longer uniquely identifies the $\hat{t}^0_{ij}$. If the taste parameters are free to be asymmetric then we can fit the trade flows exactly. That is, we use the observed trade flows rather than the fitted trade flows to compute the structural parameters. Effectively, this reallocates the assumed cross-sectional error terms in the econometric model to an idiosyncratic taste bias. From an econometric perspective this is akin to adding pairwise fixed effects. With the restriction of identical tastes (or structural taste bias) lifted there is at least one taste parameter for each observed trade flow. With no degrees of freedom it is trivial to calibrate the system to a consistent equilibrium that produces an R-squared of one for the observed trade flows.

To fit the data more efficiently we simply compute the value shares based on the observed trade flows:

$$\hat{\theta}_{st} = \frac{X^0_{st}}{Y^0_t}. \tag{17}$$

We make the nontrivial assumption that the trade flows contain no measurement error (as we did with income). This directly accommodates the observed pattern of trade. The cost of this direct approach is that summary measures of how trade reacts to distance and borders, on average, are not directly reported. The benefit of this approach is that the benchmark replicates the (first-order) observations.

The first-order calibration is not complete, however, without trade flows for the unobserved pairs, and a measure of the trade costs. For simplicity and comparability with the econometric calibrations we assume that trade costs are those implied by the regression analysis (with $\hat{\lambda} = 1$). Thus, we use the regression coefficients as a source of descriptive information on average costs.[34] Obviously, it would be more appropriate to find information on each of the bilateral trade costs, however, using the regression coefficients to compute the $\hat{t}^0_{ij}$ offers a consistent point of departure to compare the more efficient calibration with the results presented above.

The unobserved trade flows are problematic, because there might be any number of combinations that are consistent with an equilibrium that includes the observed flows (satisfying the R-squared equals one criteria). To solve this problem we again appeal to the descriptive properties of the original regression. Using the regression coefficients as given, we mini-

[34] An econometric analogue to this procedure is conducted by Cheng and Wall (2003). They estimate a model with countrypair fixed effects, and then regress those fixed effects on distance and other geographic variables to measure the average effect of the geography variables on trade.

mize the sum-of-squared deviations between the fitted value of the unobserved flows and the right-hand side of equation (14) choosing values for the unobserved pairs and subject to the adding-up constraints. The programming problem is given by;

$$\begin{array}{lll}
\min_{\{\hat{X}^0_{ij}\}} & \sum_{i\notin S}\sum_{j\notin T} & \left[\ln\left(\frac{\hat{X}^0_{ij}}{Y^0_i Y^0_j}\right) - k - \beta_1 \ln d_{ij} - \beta_2(1-a_{ij}) \right. \\
 & & \left. + \ln \hat{E}_i^{0(1-\sigma)} + \ln \hat{E}_j^{0(1-\sigma)}\right]^2 \\
\text{subject to:} & & Y^0_i = \sum_s X^0_{si} + \sum_{j\notin S} \hat{X}^0_{ji}, \forall i; \text{and} \\
 & & Y^0_i = \sum_t X^0_{it} + \sum_{i\notin T} \hat{X}^0_{ij}, \forall i. \qquad \text{(NLP2)}
\end{array}$$

Any number of modifications might be added to this programming problem to accommodate additional information. For example, some of the unobserved pairs might be restricted to zero flow if we believe that there is actually no trade (this is no longer inconsistent with the theory because some of the idiosyncratic share parameters might be zero). Once the fitted values from the programming problem are used to compute the remaining value shares the calibration is complete.

Using the implied trade costs from the original model (a 51% tariff equivalent), Table 5 shows a comparison between the welfare impacts of border removal from the econometric calibration and the welfare impacts of border removal from the superior direct calibration. The econometric calibration, based on A-vW's proposed econometric model, has significant implications on the size of the welfare impact on Canada and masks many of the inter-provincial distributional impacts of border removal.

As one might expect, the dispersion in welfare effects is greatly reduced in the econometric calibration because the model is calibrated to the fitted (or average) flows. When we calibrate to the observed trade flows a rich story about the pattern of trade emerges. For example, the simulated impact on Quebec's welfare of removing the US-Canada border drops from 29% to 19% when we use the actual data.

The differences are directly attributable to errors in the fitted values implied by the

Table 5: Welfare impacts of economic integration under alternative calibrations

Province	Benchmark GDP ($US billions)	Equivalent Variation: Econometric Calibration	Equivalent Variation: Exact-Fit Calibration
Alberta	56	36%	28%
British Columbia	63	26%	14%
Manitoba	17	49%	20%
New Brunswick	9	42%	24%
Newfoundland	6	50%	78%
Nova Scotia	12	40%	14%
Ontario	194	35%	36%
Prince Edward Is.	2	45%	64%
Quebec	107	29%	19%
Saskatchewan	15	46%	19%
GDP Weighted Average		34%	27%

econometric calibration. In the fitted values the proportion of Quebec's GDP that is exported across the border is 25.9%, but in the actual observations the proportion of Quebec's GDP that is exported across the border is 15.7%. Ontario in contrast has fitted border flows that are almost identical to the observed flows (30.3% versus 30.4%). Therefore, the alternative approaches to calibration produce comparable welfare impacts for Ontario but not Quebec. The general pattern in Table 5 is an overstatement of welfare impacts under the original calibration, and this is attributed to the biased fitted flows generated by the atheoretic intercept (illustrated in Figure 1).

A further point to take away from Table 5 is that the welfare impacts of economic integration are sizeable in the last column despite our use of traditional calibration techniques. Anderson and van Wincoop (2002) contend that their approach is more plausible because it estimates non-tariff barriers to be many times larger than formal trade barriers (of course, this requires the identical-taste assumption). They go on to critique traditional calibrated models as understating the effects of NAFTA. We find the explanation of the differences trivial: the critiqued computational studies focused on experiments that removed much smaller barriers. The traditional approach is to directly measure trade barriers (tariff and non-tariff, if available) and attribute any unexplained border resistance to idiosyncratic taste parameters. In contrast, A-vW assume all trade resistance at the border to be a border charge. Table 5 shows that given comparable border charges (and comparable response parameters) the traditional calibration technique yields comparable aggregate results.

5 Conclusion

Contemporary economic analysis includes two broad traditions of fitting economic models to data. While the analytical objectives of calibration and estimation have traditionally differed, recent applications highlight the need to consider them in a unified framework. Under consistent identifying assumptions, both approaches generate the same structural parameters necessary to relate exogenous changes to endogenous outcomes.

For many structural econometric studies, the core assumptions of the economic model are without question. The goal is to identify an estimable reduced form of the equilibrium system. We emphasize the importance of returning to the model's extensive form if reduced form estimates are to be given their structural interpretation. Operationalizing an economic model requires identification of a full set of exogenous structural parameters.

In contrast to the unified framework proposed by Dawkins et al. (2001), Anderson and van Wincoop (2002) argue that estimated models are superior to calibrated simulation models. Our illustration that estimation is calibration moves the focus of analysis onto the key identifying structural assumptions that make the fitting procedure possible, and beyond the particular label placed on the fitting procedure. The econometric procedure proposed by Anderson and van Wincoop generates results that seem to contradict similar calibrated models, but we show that this is due to specific structural and parametric restrictions not found in traditional simulation models. Our contribution, however, is broader than our specific example. Viewing empirical investigations from the unified perspective, that structural estimation is calibration, will aid in applying structural assumptions consistently and in directing assumptions toward more efficient use of limited data.

References

Anderson, James E. (1979) 'A theoretical foundation for the gravity equation.' *The American Economic Review* 69(1), 106–116

Anderson, James E., and Eric van Wincoop (2002) 'Borders, trade, and welfare.' In *Brookings Trade Forum: 2001,* ed. Susan M. Collins and Dani Rodrik (Washington D.C.: Brookings Institution Press)

_ (2003) 'Gravity with gravitas: A solution to the border puzzle.' *The American Economic Review* 93(1), 170–191

Balistreri, Edward J. (2005) 'A note on the local and global consistency of Anderson and van Wincoop's gravity model.' *mimeo*, Colorado School of Mines

Balistreri, Edward J., and Russell H. Hillberry (2006) 'Trade frictions and welfare in the gravity model: How much of the iceberg melts?' *Canadian Journal of Economics* 39(1), 247–265

Brown, Drusilla K., and Robert M. Stern (1989) 'Computational analysis of the U.S.-Canada free trade agreement: the role of product differentiation and market structure.' In *Trade policies for international competitiveness,* ed. Robert C. Feenstra (Chicago: University of Chicago Press)

Cheng, I-Hui, and Howard Wall (2003) 'Controlling for heterogeneity in gravity models of trade.' Federal Reserve Bank of St. Louis, Working Paper: 1999-010D

Dawkins, Christina, T. N. Srinivasan, and John Whalley (2001) 'Calibration.' In *Handbook of Econometrics,* ed. J. J. Heckman and E. E. Leamer, vol. 5 (Amsterdam: Elsevier) chapter 58

de Melo, Jaime (1988) 'Computable general equilibrium models for trade policy analysis in developing countries: A survey.' *Journal of Policy Modeling* 10(4), 469–503

Deaton, Angus, and John Muellbauer (1980) *Economics and Consumer Behavior* (Cambridge: Cambridge University Press)

Feenstra, Robert C. (2004) *Advanced International Trade: Theory and Evidence* (Princeton: Princeton University Press)

Francois, Joseph (2001) 'Flexible estimation and inference within general equilibrium systems.' Conference Proceedings: The Fourth Annual Conference on Global Economic Analysis, vol.1, GTAP

Harrison, Glenn W., and H.D. Vinod (1992) 'The sensitivity analysis of applied general equilibrium models: completely randomized factorial sample designs.' *Review of Economics and Statistics* 74, 357–362

Harrison, Glenn W., Richard C. Jones, Larry J. Kimbell, and Randall M. Wigle (1992) 'How robust is applied general equilibrium analysis?' *Journal of Policy Modeling* 15, 99–115

Hertel, Thomas, David Hummels, Maros Ivanic, and Roman Keeney (2004) 'How confident can we be in CGE-based assessments of free trade agreements?' NBER Working Paper #10477

Hillberry, Russell H., Michael A. Anderson, Edward J. Balistreri, and Alan K. Fox (2005) 'Taste parameters as model residuals: Assessing the 'fit' of an armington trade model.' *Review of International Economics* 13(5), 973–984

Hoover, Kevin D. (1995) 'Facts and artifacts: Calibration and the empirical assessment of real-business-cycle models.' *Oxford Economic Papers* 47(1), 24–44

Hummels, David (2001) 'Toward a geography of trade costs.' *mimeo*, Purdue University

Judd, Kenneth L. (1996) 'Approximation, perturbation, and projection methods in economic analysis.' In *Handbook of Computational Economics,* ed. Hans M. Amman, David A. Kendrick, and John Rust, vol. 1 (Amsterdam: Elsevier) chapter 12

Leamer, Edward E., and James Levinsohn (1995) 'International trade theory: The evidence.' In *Handbook of International Economics,* ed. Gene M. Grossman and Kenneth Rogoff, vol. 3 (Amsterdam: Elsevier) chapter 26

Liu, Jing, Channing Arndt, and Thomas W. Hertel (2004) 'Parameter estimation and measures of fit in a global, general equilibrium model.' *Journal of Economic Integration* 19(3), 626–649

Lòpez-de-Silanes, Florencio, James R. Markusen, and Thomas F. Rutherford (1994) 'The auto industry and the North American Free Trade Agreement.' In *Modeling trade policy: Applied general equilibrium assessments of North American free trade,* ed. Joseph F. Francois and Clinton R. Shiells (Cambridge: Cambridge University Press) chapter 8

Mathiesen, Lars (1985) 'Computation of economic equilibria by a sequence of linear complementarity problems.' *Mathematical Programming Study: 23* pp. 144–162

Perroni, Carlo, and Thomas F. Rutherford (1995) 'Regular flexibility of nested CES functions.' *European Economic Review* 39(2), 335–343

_ (1998) ‘A comparison of the performance of flexible functional forms for use in applied general equilibrium analysis.’ *Computational Economics* 11(3), 245–263

Rose, Andrew K., and Eric van Wincoop (2001) ‘National money as a barrier to international trade: The real case for currency union.’ *American Economic Review* 91(2), 386–390. Papers and Proceedings of the Hundred Thirteenth Annual Meeting of the American Economic Association

Rutherford, Thomas F. (1995a) ‘Constant elasticity of substitution functions: some hints and useful formulae.’ *mimeo*, University of Colorado

_ (1995b) ‘Extensions of gams for complementarity problems arising in applied economic analysis.’ *Journal of Economic Dynamics and Control* 19(8), 1299–1324

Shoven, John B., and John Whalley (1984) ‘Applied general-equilibrium models of taxation and international trade: An introduction and survey.’ *Journal of Economic Literature* 22(3), 1007–1051

Trefler, Daniel (1993) ‘Trade liberalization and the theory of endogenous protection: An econometric study of U.S. import policy.’ *Journal of Political Economy* 101(1), 138–160