ECON G6905
Topics in Trade
Jonathan Dingel
Autumn 2023, Week 8

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Counterfactual outcomes by exact hat algebra in trade model

Recall exact hat algebra for Armington CES model from week 1:

$$\hat{w}_i \hat{L}_i = \hat{Y}_i = \sum_{j=1}^{N} \gamma_{ij} \hat{\lambda}_{ij} \hat{w}_j \hat{L}_j = \sum_{j=1}^{N} \frac{\gamma_{ij} \hat{T}_i \hat{w}_i^{-\epsilon} \hat{\tau}_{ij}^{-\epsilon}}{\sum_{l=1}^{N} \lambda_{lj} \hat{T}_l \hat{w}_l^{-\epsilon} \hat{\tau}_{lj}^{-\epsilon}} \hat{Y}_j$$

▶ The name refers to Jones (1965) "hat algebra" for comparative statics

▶ It's "exact" because it's global given whole demand and supply system

A system of equations characterizing counterfactual endogenous variables $\hat{w}_i$ in terms of initial equilibrium shares $\lambda_{ij} \equiv \frac{X_{ij}}{X_j}$ and $\gamma_{ij} \equiv \frac{X_{ij}}{Y_i}$, the trade elasticity ϵ, and counterfactual exogenous changes $\hat{L}_i, \hat{T}_i, \hat{\tau}_{ij}$

This week: Exact hat algebra and calibration

▶ Some counterfactual scenarios require more parameters than λ_{ij}, γ_{ij}, ϵ

▶ Exact hat algebra is not a calibration procedure

▶ A saturated model calibrated to observed shares may still be falsifiable

▶ Calibrating model parameters shares to perfectly reproduce observed shares risks overfitting the model

Counterfactual scenarios and required parameter values

Some counterfactual scenarios require (knowledge of) more parameters than just the information encoded in initial equilibrium shares λ_{ij} and γ_{ij} and the trade elasticity ϵ.

Eaton and Kortum (2002) conduct two thought experiments:

- Autarky: $\tau_{ij} \to \infty$ for $i \neq j$. $\hat{\tau}_{ij} = \infty$.
- "Zero gravity": $\tau_{ij} \to \tau_{jj}$ for $i \neq j$. $\hat{\tau}_{ij} = \frac{\tau_{jj}}{\tau_{ij}}$.

Knowing model parameters up to λ_{ij}, γ_{ij}, ϵ is *insufficient* to compute the zero-gravity scenario because that requires $\hat{\tau}_{ij} = \frac{\tau_{jj}}{\tau_{ij}}$.

Prior descriptions of exact hat algebra and calibration

Costinot and Rodriguez-Clare (Handbook 2014):
"We have described how to use gravity models to perform welfare and counterfactual analysis. An appealing feature of this approach, which we have referred to as the exact hat algebra, is that the impact of various counterfactual scenarios can be computed without estimating the full structure of the model. All the relevant information about preferences, technology, and trade costs can be inferred directly from the cross-section of bilateral trade flows and estimates of the trade elasticity."

Caliendo and Parro (2015):
"we can solve for the general equilibrium of the model without needing to estimate parameters which are difficult to identify in the data, [such] as productivities and iceberg trade costs." (c.f. Lewbel 2019)

Exact hat algebra is not a calibration procedure

Exact hat algebra concerns comparative statics, not calibration (blog post)

- ▶ The system of equations defines counterfactual outcomes for any model of the baseline shares λ_{ij} and γ_{ij} and the trade elasticity ϵ
- ▶ Model parameters matter (only) because they determine the baseline shares and the trade elasticity

Using observed shares is calibrating (combinations of) parameters

Finding (products of) parameters to match observed shares is calibrating them.

For purposes of computing counterfactual outcomes, using observed shares typically leverages two properties of the model and the counterfactual scenario:

- ▶ The model has enough degrees of freedom that there exist parameters that can rationalize any observed pattern of trade flows. ("Saturated" model)
- ▶ The baseline shares are sufficient to calibrate the combinations of parameters needed to compute counterfactual outcomes.

Corollaries to these two features when using observed shares:

- ▶ No pattern of baseline shares can reject the model specification. (Models with fewer degrees of freedom admit tests.)
- ▶ The set of admissible counterfactual scenarios is restricted. (e.g., zero gravity)

Are Armington model parameters identified by trade flows?

A statement equivalent to "one can calibrate the model to observed shares without estimating the full structure of the model" is "the model parameters are not identified by observed trade shares"

- If $\{T_i, L_i, w_i\}$, $\{\tau_{ij}\}$, and ϵ satisfy

$$w_i L_i = \sum_j \frac{T_i (w_i \tau_{ij})^{-\epsilon}}{\sum_\ell T_\ell (w_\ell \tau_{\ell j})^{-\epsilon}} w_j L_j$$

then $\{\lambda_i^\epsilon T_i, L_i, w_i\}$, $\{\lambda_i \tau_{ij}\}$, and ϵ also satisfy this system for any $\lambda_i > 0$. With iceberg trade costs, productivity and a shifter of export costs are the same.

- If $\{T_i, L_i, w_i\}$, $\{\tau_{ij}\}$, and ϵ satisfy this system, then $\{T_i, L_i, w_i\}$, $\{\xi_j \tau_{ij}\}$, and ϵ satisfy this system for any $\xi_j > 0$. Expenditure shares reflect relative prices.

- Similarly, $\{\mu_i^\epsilon T_i, \mu_i^{-1} L_i, \mu_i w_i\}$, $\{\tau_{ij}\}$, and ϵ also satisfy this system for any $\mu_i > 0$. (Isomorphic endowment economy has $Q_i = T_i^{1/\epsilon} L_i$)

Are Armington model parameters identified by trade flows?

Suppose you want to identify *relative* trade costs, τ_{ij}/τ_{jj}. Normalize $\tau_{jj} = 1 \ \forall j$.

Using two prior results, if $\{T_i, L_i, w_i\}$, $\{\tau_{ij}\}$, and ϵ satisfy this system, then $\{\lambda_i^\epsilon T_i, L_i, w_i\}$, $\{\lambda_i \xi_j \tau_{ij}\}$, and ϵ also satisfy this system for any $\lambda_i > 0$ and $\xi_j > 0$.

$$Y_i = \sum_j \frac{T_i (w_i \tau_{ij})^{-\epsilon}}{\sum_\ell T_\ell (w_\ell \tau_{\ell j})^{-\epsilon}} Y_j = \sum_j \frac{X_{ij}}{X_{jj}} X_{jj} = \sum_j \frac{T_i (w_i \tau_{ij})^{-\epsilon}}{T_j (w_j \tau_{jj})^{-\epsilon}} X_{jj}$$

$$= \sum_j \frac{\lambda_i^\epsilon T_i (w_i \lambda_i \xi_j \tau_{ij})^{-\epsilon}}{\lambda_j^\epsilon T_j (w_j \lambda_j \xi_j \tau_{jj})^{-\epsilon}} X_{jj}$$

If one chooses $\xi_j = \lambda_j^{-1}$, then this preserves $\tau_{jj} = 1$.

We cannot recover *directed* trade costs from directed trade flows.

Factory-gate prices do not discipline the Armington model

▶ Recall that bilateral trade costs τ_{ij} are isomorphic to bilateral preference shifters β_{ij} in the Armington model (week 1).

▶ Conditional on $Y_i = p_i Q_i$, any observed p_i can be rationalized by introducing preference shifters.

▶ Similarly, in Melitz (2003), firm-level productivity φ might shift TFPR by lowering costs (TFPQ) or raising demand (quality). Given this isomorphism, firm-level prices can do little to discpline the model.

▶ If you've committed to the Eaton-Kortum model with symmetric preferences, then the price index may be more informative. But an isomorphic "quality" interpretation of T_i seems easy.

A saturated model may still be falsifiable

▶ We cannot use baseline shares (nor prices) to evaluate the model

▶ The CES model of trade flows is a demand system and we know how to test demand systems (at least, the one-elasticity straightjacket of CES/IIA)

▶ One approach: Estimate an encompassing specification and test the parameter restrictions imposed by CES (e.g., Hausman and McFadden (1984) test IIA logit by estimating a nested-logit specification and testing whether $\lambda = 1$)

▶ Other tests specifically leverage IIA: Estimate on subsets of countries and see if elasticities are common (the other Hausman and McFadden (1984) test)

▶ One can also test the complete passthrough from tariffs to prices (Head, Mayer 2023)

▶ One could test the triangle-inequality on trade costs

The CES/IIA model with one elasticity does not fit the data

Adao, Costinot, and Donaldson (2017) study reduced exchange economies:

- ▶ "the mixed CES demand system allows data to speak to whether this independence of irrelevant alternatives embodied in CES holds empirically"
- ▶ "the deviations from IIA...are a systematic feature of the data...related to the similarity of competitors in terms of per capita GDP"

Lind, Ramondo (2023) develop a Ricardian model with richer productivity patterns:

- ▶ "By relaxing the independence assumptions used in the literature, the model generates import demand systems spanning the entire generalized extreme value (GEV) class (McFadden 1978, 1981)."
- ▶ "significant sharing of technologies across countries and sectors...manifests in considerable heterogeneity in correlation in productivity, which, in turn, changes the answers to standard counterfactuals"
- ▶ "our [latent factor model] estimates capture quite accurately departures from IIA within and across sectors"

Overfitting

▶ One overfits a model by using a more flexible parameterization that improves in-sample fit but worsens out-of-sample performance.

▶ Adding more covariates always improves the R^2 of OLS, but "use as many variables as possible" is not the preferred specification of forecasters (Perhaps more relevant: Consider 2SLS with N observations and N instrumental variables. The first-stage R^2 of 1.0 is not good news: $\beta^{\text{OLS}} = \beta^{\text{2SLS}}$. Now imagine using N white-noise draws as IVs.)

▶ Hastie, Tibshirani, and Friedman (2009): "Unfortunately training error is not a good estimate of the test error, as seen in Figure 7.1. Training error consistently decreases with model complexity, typically dropping to zero if we increase the model complexity enough. However, a model with zero training error is overfit to the training data and will typically generalize poorly."

A saturated model is overfit if there's noise

Belloni, Chernozhukov, Hansen (2014) come close to describing the calibrated-shares procedure:

The key concept underlying the analysis of high-dimensional data is that dimension reduction or "regularization" is necessary to draw meaningful conclusions. The need for regularization can easily be seen when one considers an example where there are exactly as many variables (plus a constant) as there are observations. In this case, the ordinary least squares estimator will fit the data perfectly, returning an R^2 of one. However, using the estimated model is likely to result in very poor forecasting properties out-of-sample because the model estimated by least squares is overfit: the least-squares fit captures not only the signal about how predictor variables may be used to forecast the outcome, but also fits the noise that is present in the given sample, and is not useful for forming out-of-sample predictions. Producing a useful forecasting model in this simple case requires regularization; that is, the estimates must be constrained so that overfitting is avoided and useful out-of-sample forecasts can be obtained.

A more critical perspective on calibrated shares

Antras and Chor (Handbook, 2022):

" practitioners of this approach often praise how parsimonious it is relative to CGE models, which involve the estimation of thousands of parameters. An often glossed-over fact, however, is that the hat-algebra approach requires the model to fit the data *exactly*, which amounts to calibrating all parameters of the model (or combinations of them) to values that ensure this exact fit . . .

although quantitative work often requires strong assumptions on functional forms, calibrating thousands of parameters to fit the data exactly can be problematic for the validity or reliability of the counterfactual predictions of those models. The problem is similar to overfitting in regression analysis leading to poor out-of-sample performance. As recently shown in Dingel and Tintelnot (2020), this is a particularly severe problem in spatial environments in which the data the model is fitted to contains a significant number of zeros. Note that even in the WIOD – a WIOT focusing on relatively rich countries – the share of zeroes is 13.7% in the matrix of input-use coefficients and 46.8% in the matrix of final-use column vectors."

Dingel and Tintelnot (2023)

The calibrated-shares procedure can perform poorly in high-dimensional settings because matching (noisy) observed shares amount to overfitting the model parameters.

Go to Dingel and Tintelnot (2023) slidedeck.