Chapter 8
Validation of Computable General Equilibrium Based Models

8.1 Introduction

Model validation in economics is more difficult than in many other disciplines, especially at the macroeconomic level. Controlled experiments are often inappropriate because macroeconomic modeling involves independent individual decision-makers and their interactions in the context of background conditions, such as changes business cycles and technological change, many of which are random or otherwise difficult to predict. Economics is more of an "observational" discipline like meteorology, astronomy, or sociology, and must therefore use approaches such as statistical analysis of data or simulation approaches.

Thus, the validation of E-CAT must be accomplished through indirect methods. In this chapter, we present some of these methods, discuss their relative merits and limitations, and apply two formal methods to one of the threats—an aviation system disruption.

This report is divided into 4 Sections. In Sect. 8.2, we summarize validation approaches and their application to CGE models in general. In Sect. 8.3, we discuss various model validation procedures, and in Sect. 8.4 apply two of them to the E-CAT Model.

8.2 Validation Criteria and Their Application to CGE Models

The following criteria have been used to validate economic models in general (see, e.g., Rose 2004; Dixon and Rimmer 2013):

Conceptual Soundness Does the model have a solid conceptual base? Is it based on established theory?

CGE models are generally considered to have a solid conceptual base because they represent an operational version of general equilibrium theory, or the interaction of individual decision-makers in multiple interconnected markets. The CGE model at the core of E-CAT is based on one major traditional approach to CGE model construction initially developed by Dervis et al. (1982) and Robinson et al. (1990) and that is closely related to another prevalent approach popular in the U.S. (Rutherford 1999).

Realism. Is the model reasonably realistic? Are its major assumptions too great a departure from reality? Of course, all models are an abstraction, but is the level of abstraction so great as to question its validity?

CGE models reflect standard behavior of representative producers and consumers in a multi-market context. The assumption of equilibrium is often criticized, but the CGE model at its core allows for disequilibria in the labor market, trade balances, fiscal balances, and, most importantly, imbalances in markets of produced goods and services due to external shocks (Rose 2015).

Applicability. Is the model appropriate to the case in point? Does it cover the requirements of the topic to be addressed?

CGE models represent the state-of-the-art approach to analyzing the economic consequences of disasters. They are especially adept at tracing economy-wide impacts of targeted shocks. More recently, they have been refined to include the 2 major categories of unconventional responses that distinguish economic consequence analysis of disasters from ordinary economic impact analysis: behavioral responses and resilience (see, e.g., Giesecke et al. 2012; Rose and Liao 2005). These 2 major categories of drivers are key components of the E-CAT analysis.

Comprehensiveness. Is the model broad enough to encompass key background conditions that could have a significant effect on the results?

CGE models are very comprehensive in several ways. They represent a full accounting of all inputs into production and all goods and services in consumption. They also include socioeconomic accounts and can include environmental variables, though there was no necessity for this to be done for E-CAT, except in limited cases such as oil spills. They also factor in many background conditions, such as unemployment rates, labor force participation rates, and factor constraints. However, most CGE models omit explicit consideration of inventories and excess capacity, as does ours, but these are relatively minor sources of resilience.

Data Quality. Are the data reasonably current? Are the data from a reliable source? Primary data are generally considered the most reliable, in part, because collection methods and assumptions are likely known, in contrast to secondary data, which refers to a compilation or aggregation of data, typically from published source, for which the origin is not as well known.

CGE models are based on a comprehensive set of input-output table accounts and their extension to social accounting matrix. Only national governments have the resources to collect the universe of data to compile these tables from primary sources. Even tables based on samples are prohibitively expensive. Hence, a

number of "data-reduction", or "non-survey", methods have been devised to generate input-output (I-O) tables. Similar methods have been devised to provide reasonable model updates as well (Miller and Blair 2007). We have used the most recent I-O data available to calibrate the USCGE Model. One area in which practically all CGE models can be criticized is the fact that their other major parameters, elasticities substitution in elasticities of demand are not based on primary data or inferential statistics as for the time and place which the model conforms. Rather they are "borrowed" from the most closely related context possible. Our selection of elasticity values was made after an extensive inspection of alternative elasticity values.

Model Construction. Is the method of model construction sound?

CGE Models are constructed on the basis of model calibration fitting many of the parameters to a single year of data by using "ratio" estimators, or simply the division of data on inputs by data on the gross output they produce in each sector. This estimator is considered to have less desirable properties than, for example, OLS or maximum likelihood estimators using regression analysis. The USCGE Model is also open to this criticism.

Track Record. Has the model or similar models been validated in related contexts?

CGE models are one of the most widely used tools of economic consequence analysis. They are held to be widely superior to input-output models, especially in the more complex context of disasters (Rose 1995, 2015). They are considered to be reliable for a broad range of public and private sector decisions. They have been validated in general by several methods (Dixon and Rimmer 2013). The US CGE Model and its several regional variants have been applied in more than a dozen major studies of the economic consequences of disasters published in major peer-reviewed journals (see, e.g., Rose and Liao 2005; Rose et al. 2007; Rose et al. 2009; Oladosu et al. 2013; Rose et al. 2015).

Accuracy. Are the results of the application of the model reasonable overall, or, better yet, considered accurate according to modeling or statistical criteria?

CGE models are considered to provide reliable results in many applications, especially where omitted variables are not likely to have a major influence and where assumptions are not too great of a departure from reality. Two validation tests are applied later in this chapter and help demonstrate the accuracy of the US CGE Model.

8.3 Model Testing Procedures and CGE Models

Several methods, or procedures, have been used to test economic models. Dixon and Rimmer (2013) note that there is not necessarily a one-to-one correspondence between the purposes of model validation (which we have labeled "approaches") and procedures. The following represents a list of such procedures, following Dixon and Rimmer (2013), and general practice in economics and other fields.

- Is the model consistent with an underlying set of statistical accounts? For example, does the base year or equilibrium version of the model replicate these accounts?

CGE models are based on an underlying table of double-entry book keeping accounts for an economy in a given year stemming from both input-output tables and social accounting matrices. The model is based on a transparent conversion of these annual inter-sectoral flows to normalized parameters (initial direct input values) by dividing each element in the table by its column sum (in most cases, the gross output of the good or service produced). Key parameters allow these input values to vary under different conditions. One of the standard consistency checks in constructing an I-O model is whether the equilibrium solution replicates the base accounts. This was done for our model as well.

- Estimation of Parameters. Do superior methods other than those on which the model is based, such as econometric estimation, yield parameter estimates close to those used in the model?

It has long been observed that the accuracy of many parameters used in CGE models could be improved by econometric estimation. However, the necessary time series data are generally not available to do so. Only one major US CGE model has been completely and consistently econometrically estimated (Jorgenson and Wilcoxen 1990).

- In-sample Tests. Can the model be used to accurately reflect the data and results of some of the inputs in its construction? For example, for regression analysis of model results, will that regression equation yield a close approximation to one of the sets of variable and parameter values used to estimate the regression in the first place?

CGE models have been found to pass this most basic test, which we apply below. We also applied a more sophisticated version known as the "cross-validation" test (Armine et al. 2013).

- Out-of-sample Tests. Do the model predictions conform to observed cases not in the sample?

This is a valuable test of CGE models when it is feasible. However, due to the lack of accurate estimates of out-of-sample cases, we do not apply it to our model in this volume. Note also that CGE models, as is the case for other modeling approaches, will perform better if background conditions remain relatively constant.

- Back-casting. Can the model simulate the historical record? This overlaps with the third procedure, if the historical case is within the sample, and overlaps with the fourth procedure if it is not.

This applies in a like manner to out-of-sample Tests.

- Sensitivity Tests. Do the predictions of the model swing wildly as a result of small changes in parameters? A more formal version of this procedure would generate confidence intervals surrounding the model outputs.

Given a large number of parameters (direct input or "technical" coefficients) in most I-O models, individual, or a small set of, parameter changes are unlikely to cause major swings in the results, except for limited cases where the parameters are a very high proportion of the sector's input requirements. Estimation of confidence intervals is not possible because CGE models lack, or have very limited, formal statistical properties.[1] We have performed several tests on the model parameters, such as input substitution elasticities and import (Armington) elasticities.

- Reduced-form Methods. The most basic version of this approach is what Dixon and Rimmer (2013) refer to as the "back-of-the-envelope" (BOTHE) approach, which translates the analysis into simple supply-demand, or equivalent "basic principles" diagrams. Another approach is regression analysis of multiple simulations from the model on the basis of the "synthetic" data generated by it. Typically ordinary least squares estimation is used, but additional insight can be developed by breaking the sample up in applying quantile regression analysis (see, e.g., Rose et al. 2011).

   This approach is at the core of the E-CAT Model.

- Consistency Checks. Is the model able to replicate outcomes for its endogenous variables given "true" values of exogenous ones?

   This represents an important step in the calibration of CGE models in general, and is satisfied in the construction of the US CGE Model. Specifically, the initial equilibrium solution of a CGE model must replicate its underlying social accounting matrix.

8.4   Model Validation Applications

We performed two formal tests of the validity of an Aviation System Disruption Scenario that is included in E-CAT. Below we present the results of both In-sample Validation and Cross-sample Tests.

8.4.1   In-Sample Validation

We first tested the reduced-form estimates from the E-CAT by subjecting them to two in-sample test cases. This involved comparing the GDP loss estimates of the Tool with the estimates from the studies by Rose et al. (2015) and Rose et al. (2009). The former was adopted as the lower-bound case, whereas the latter was adopted as the upper-bound case. Both studies were conducted independently under

---

[1]Econometric methods are often considered superior because they involve more extensive data (e.g., a time series, as opposed to data for a single year) and yield "goodness of fit" measures.

Table 8.1 Comparison of economic consequence estimates between E-CAT and the literature

Scenario	Reference case estimate[a]	E-CAT estimate[a] (High resilience)	Percent difference from reference case estimate	E-CAT estimate[a] (Low resilience)	Percent difference from reference case estimate
LAX bomb	23.1	16.5	−28.6	28.5	23.4
9/11 WTC	121.0	109.5[b]	−9.5	n/a	n/a

[a]Measured in billions of 2012 dollars
[b]The estimate is converted from $109 billion in 2006 dollars using the GDP deflator

somewhat different analytical frameworks, but under similar assumptions. This consistency and the broad range of outcomes make them useful benchmarks to validate E-CAT for this scenario.

The comparison of economic consequence estimates between E-CAT and the two in-sample test cases are presented in Table 8.1. For the case of a hypothetical bomb attack at the Los Angeles International Airport (LAX), the GDP loss estimates in the lower- and upper-bound resilience cases from E-CAT are $28.5 billion and $16.5 billion, respectively, and the reference case estimate of the national GDP loss is $23.1 billion (Rose et al. 2015), which falls very close to the midpoint of the range of the E-CAT estimates.

For the 9/11 World Trade Center (WTC) case, the estimate from E-CAT in the upper-bound resilience case is $109.5 billion, which is very close to the estimate of $121 billion by Rose et al. (2009). Only the E-CAT estimate with high resilience is included for the WTC attack scenario given the fact that resilience was actually found to be high in this case (Rose et al., 2009), so no bounding exercise was necessary. In the LAX Bomb Attack scenario, modeled in accordance with a TSA scenario, both low- and high-level resilience are included, as this scenario is purely hypothetical, and we have no specific knowledge of whether resilience would be closer to the lower or upper bound.

Overall, the in-sample validation tests support the contention that the E-CAT reduced-form tool is able to produce estimates of GDP losses consistent with the reference case estimates. The difference between the E-CAT estimate and the reference estimate is −9.5 % in the case of the 9/11 WTC Attack, whereas the estimates from the E-CAT in the case of the LAX Bomb Attack range between −28.6 and +23.4 % of the reference case estimate, and the average estimate ($22.5 billion) from E-CAT deviates from the reference estimate ($23.1 billion) by only −2.5 %.

8.4.2   Cross-Validation Test

The reduced-form CGE approach is validated using the cross-validation test with holdout samples based on the aviation system disruption scenario. The purpose is to evaluate whether the synthetic data generated from Latin-hypercube sampling

Table 8.2 Mean value comparison of variables among data sets

Variable	Training set	Testing set	Raw data
GDP change in value (billions of dollars)	−51.70	−31.81	−47.45
GDP change in percent	−0.32	−0.20	−0.29
Employment change	−0.32	−0.21	−0.30
Employment change in percent (millions of jobs)	−0.25	−0.17	−0.23
Magnitude (Percent of air service being disrupted)	14.59	14.72	14.65
Resilience lower bound (1=Yes/0=No)	0.45	0.40	0.44
Resilience upper bound (1=Yes/0=No)	0.23	0.35	0.25
Behavioral effect (1=Yes/0=No)	0.56	0.30	0.50
No. of observations	80	20	100

procedure and CGE analysis has an overfitting problem.[2] This is an important task as the validation helps to justify the predictive power of the reduced form equations.

The test with holdout samples was implemented in the following six steps:

- 80 % of the raw synthetic data is selected as the "training set" and the remaining 20 % as the "testing set".
- Training set, testing set and raw dataset are compared.
- Reduced form OLS regressions were conducted based on the training set for GDP and employment.
- Comparison of regression results for GDP based on the raw dataset and the training set suggests that the reduced form estimates for GDP and employment are consistent in both cases.
- Predicted estimates based on the training set are applied for the testing set to calculate goodness of fit of sub-reduced form model.
- Validation results are compared, suggesting the training set provides very similar estimates to those based on the raw data.

The descriptive statistics of the training set, testing set and the raw dataset are compared in Table 8.2, which shows that the mean values of all the variables for the training sets are higher than that for the original data set, whereas the mean values for the testing sets are relatively lower than that for the original set. The results of the reduced-form regression analysis are compared and summarized in Table 8.3. The deviations (in percent) of the training set estimates from the original set are generally within 5 %.

To further validate the data, the predicted estimates based on the training set are compared with the raw data set. The goodness of fit is measured through the multiple correlation coefficient, R-squared, which is calculated based on the

---

[2]"Overfitting is an issue that exists in in statistics and machine learning. It occurs when a statistical model describes random error or noise instead of the underlying relationship. Overfitting generally occurs when a model is excessively complex, such as having too many parameters relative to the number of observations. A model that has been overfit will generally have poor predictive performance, as it can exaggerate minor fluctuations in the data."

Table 8.3 OLS regression results for the aviation system disruption scenario

Dependent variable	Original data		Training set		Deviation (in percent)[a]	
	GDP	Employment	GDP	Employment	GDP	Employment
Magnitude	−7.32***	−0.05***	−7.43***	−0.05***	1.50	0.00
	(−10.16)	(−11.14)	(−9.42)	(−10.18)		
Resilience lower bound	48.13***	0.318***	50.55***	0.329***	5.03	0.03
	(−8.47)	(−9.60)	(8.11)	(−9.02)		
Resilience upper bound	60.14***	0.392***	61.96***	0.401***	3.03	0.02
	(−9.22)	(−10.33)	(8.39)	(9.26)		
Behavior impact	−73.14***	−0.400***	−74.56***	−0.408***	1.94	0.02
	(−14.93)	(−14.03)	(−13.59)	(−12.71)		
Constant	60.15***	0.350***	61.93***	0.357***	2.96	0.02
	(−5.08)	(−5.08)	(4.88)	(4.81)		
No. of observations	100	100	80	80		
adj. R-squared	0.795	0.803	0.804	0.802		

[a]Deviation is calculated using the difference of estimates between the training set and the original data divided by the original estimates
[b]T-statistics are included in parenthesis
[c]***represents statistical significance at the 1 % level

Table 8.4 Validation test based on the comparison between predictions of testing set

Content	Training set (80 observations)	Raw data (100 observations)
Reduced form equation		
GDP	GDP=61.93−7.43*Magnitude +50.55*Lower-Bound Resilience +61.96*Upper-Bound Resilience−74.56*Behavioral Effect	GDP=60.15−7.32*Magnitude +48.13*Lower-Bound Resilience +60.14*Upper-Bound Resilience−73.14*Behavioral Effect
Employment	Employment=0.35−0.05*Magnitude +0.32*Lower-Bound Resilience +0.4*Upper-Bound Resilience−0.40*Behavioral Effect	Employment=0.36−0.05*Magnitude +0.33*Lower-Bound Resilience +0.4*Upper-Bound Resilience−0.41*Behavioral Effect
R-squared[a]		
GDP	0.749	0.756
Employment	0.784	0.782

[a]The R-squared is calculated based on comparing the variability of the estimation errors with the variability of the original values for the testing sets, and is defined as: $R^2 = 1 - \dfrac{SS_E}{SS_T}$ where $SS_T = \sum_i (y_i - \widetilde{y})^2$ and $SS_E = \sum_i (y_i - \widehat{y}_i)^2$. y_i and $\widehat{y}_i$ are the original data values and predicted values based on either the training set or the original data set, respectively. Hence, the SS_T measures the total sum of squares, and SS_E measures the sum of squared errors

predicted values of GDP and employment and their corresponding values in the two different data sets. The validation results in Table 8.4 suggest that the training set provides very similar estimates to the estimates based on the raw data. Hence, it shows that the synthetic data are reliable for the reduced-form analysis.

References

Amine A, Mohamed O, Bellatreche L (2013) Modeling approaches and algorithms for advanced computer applications. Springer, Heidelberg

Dervis K, de Melo J, Robinson S (1982) General equilibrium models for development policy. Cambridge University Press, New York/Cambridge

Dixon P, Rimmer M (2013) Validation in computable general equilibrium modeling. In: Dixon PB, Jorgenson DW (eds) Handbook of computable general equilibrium modeling. Elsevier, Amsterdam/North-Holland

Giesecke JA, Burns WJ, Barrett A, Bayrak E, Rose A, Slovic P, Suher M (2012) Assessment of the regional economic impacts of catastrophic events: CGE analysis of resource loss and behavioral effects of an RDD attack scenario. Risk Anal 32(4):583–600

Jorgenson DW, Wilcoxen PJ (1990) Environmental regulation and U.S. economic growth. RAND J Econ 21(2):314–340

Miller RE, Blair PD (2007) Input-output analysis: foundations and extensions, 2nd edn. Cambridge University Press, New York/Oxford

Oladosu G, Rose A, Lee B (2013) Economic impacts of potential foot and mouth disease agro-terrorism in the United States: a computable general equilibrium analysis. J Biothreats Biodefense S12:001

Robinson S, Kilkenny M, Hanson K (1990) The USDA/ERS CGE model of the U.S. USDA, Washington, DC

Rose A (1995) Input-output economics and computable general equilibrium models. Struct Chang Econ Dyn 6:295–304

Rose A (2004) Economic principles, issues, and research priorities in natural hazard loss estimation. In: Okuyama Y, Chang S (eds) Modeling the spatial economic impacts of natural hazards. Springer, Heidelberg

Rose A (2015) Macroeconomic consequences of terrorist attacks: estimation for the analysis of policies and rules. In: Mansfield C, Smith VK (eds) Benefit transfer for the analysis of DHS policies and rules. Edward Elgar, Cheltenham

Rose A, Liao S (2005) Modeling resilience to disasters: computable general equilibrium analysis of a water service disruption. J Reg Sci 45(1):75–112

Rose A, Oladosu G, Liao S (2007) Business interruption impacts of a terrorist attack on the electric power system of Los Angeles: customer resilience to a total blackout. Risk Anal 27(3):513–531

Rose A, Oladosu G, Lee B, Beeler Asay G (2009) The economic impacts of the September 11 attacks: a computable general equilibrium analysis. Peace Econ Peace Sci Public Policy 15(2)

Rose A, Wei D, Wein A (2011) Economic impacts of the ShakeOut Scenario. Earthq Spectra 27(2):539–557

Rose A, Avetisyan M, Chan O, Rosoff H, Burns W, Slovic P (2015) The role of behavioral responses in the total economic consequences of terrorist attack on U.S. air travel targets. CREATE, University of Southern California

Rutherford T (1999) Applied general equilibrium modeling with MPSGE as a GAMS subsystem: an overview of the modeling framework and syntax. Comput Econ 14:1–46