ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439

Propagation of data error and parametric sensitivity in computable general equilibrium model forecasts

Joshua Elliott, Meredith Franklin, Ian Foster, Kenneth Judd, and Todd Munson

Mathematics and Computer Science Division

Preprint ANL/MCS-P1650-0709

July 2009, Revised November 2009

Propagation of data error and parametric sensitivity in computable general equilibrium model forecasts

Joshua Elliott* Meredith Franklin†
Ian Foster‡ Kenneth Judd§ Todd Munson¶

November 20, 2009

Abstract

We present a thorough sensitivity analysis of a computable general equilibrium model to both data and parameter errors. By examining model forecasts, we assess the impact of uncertainty in the parameters at the static core of the model and in the data set used to calibrate the model to a fixed base year. We also examine the behavior of the model due to the propagation of these uncertainties.

1 Introduction

The integrated assessment modeling (IAM) community seeks to explore complex interdisciplinary problems by synthesizing knowledge from a variety of areas into one framework. In assessing the economic dimensions of climate change, IAM researchers typically couple an economic model which forecasts energy demand, with a climate model which forecasts temperature increases due to a changing carbon cycle and anthropogenic forcing [4, 5, 7–9, 22, 24, 25]. However, a major challenge in developing an IAM framework with which policy-driven scenarios can be soundly examined is addressing and incorporating uncertainty. Computable general equilibrium (CGE) models have long been used to model the economy in IAMs and consist of equations that incorporate many parameters estimated from historical economic data. The two classes of such parameters in a typical static CGE model that based on nested constant elasticity of substitution (CES) type production and utility functions are elasticity of substitution and share parameters. As uncertainty is often overlooked, the sensitivity of CGE models to various parameter specifications is poorly understood. Furthermore, failure to incorporate uncertainty into an IAM framework could lead to misleading or erroneous conclusions.

Previous studies of parameter uncertainty have been limited to assessing a small subset of the relevant parameters or to CGE models of a single country's economy. For example, Webster et al. [26] and Hertel et al. [15] explored forecast sensitivity to variations in subsets of relevant elasticities, primarily those of substitution between capital and labor in the production functions. A study of calibration data uncertainty in the context of CGE models has never been undertaken, likely due to the prohibitively large sample sizes required to study parameter spaces at the relevant dimensionality. Approaches taken to quantify uncertainty include statistical sensitivity analysis [1, 14] and stochastic parameter control [17, 18]. In the statistical approach, distributions, typically Gaussian or uniform, are defined around uncertain parameters and the CGE is solved under random parameter draws from these distributions. Monte Carlo methods can be conveniently applied to sample the uncertainty distributions of exogenous parameters, but require significant computational power for large models. Stochastic parameter control factors in the temporal evolution of uncertain parameters and their relationships by incorporating uncertainty through the estimated variance-covariance matrices of the uncertain parameters.

In this paper, we report results from a study examining sensitivity to both parametric and calibration data uncertainties in the static CGE model that forms the core of the multi-sector and multi-regional CIM-EARTH model [13].

*University of Chicago Computation Institute and Argonne National Laboratory, email: `jelliott@ci.uchicago.edu`
†University of Chicago and Argonne National Laboratory email: `mfrankli@galton.uchicago.edu`
‡University of Chicago Computation Institute and Argonne National Laboratory, email: `foster@mcs.anl.gov`
§Hoover Institute, email: `judd@hoover.stanford.edu`
¶University of Chicago Computation Institute and Argonne National Laboratory, email: `tmunson@mcs.anl.gov`

In Section 2 we briefly review CGE models and CIM-EARTH. In Section 3, we detail the methodology used for model simulations, as well as the parameter distributions and tools used to conduct the uncertainty analysis. In Section 4.1 we report the results of ensemble simulations that we use to study the propagation of uncorrelated noise in the expenditure values of the model calibration data. In Section 4.2 we report on an ensemble simulation that we use to study the relative sensitivity of the model to uncertainties in the substitution elasticities, the primary parametric determinants of the static model. This study includes ensemble results for the full elasticity parameter space (71 parameters) and for one subset: the Armington international trade elasticities.

2 Background: the CGE Model

The basic calibrated share CES production function [6] has the form

$$\frac{y}{\bar{y}} = \left(\sum_i \theta_i \left(\frac{x_i}{\bar{x}_i} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{1}$$

where $\frac{y}{\bar{y}}$ is the ratio between the output of the industry in question (or of an intermediate bundle or aggregator function in a nest structure) at some time, and the base year value for this quantity from the calibration data set; and $\frac{x_i}{\bar{x}_i}$ are the ratios of the inputs for commodity i (capital, labor, coal, intermediate bundles) at that time with their respective base year values. The elasticity parameter, σ, controls to what degree these inputs can be substituted for one another as their relative prices change. At either extreme(σ=0 or σ=∞) we obtain special cases of the production function. For σ=0, we obtain the Leontief production function $\frac{y}{\bar{y}} = \min_i \left\{ \frac{x_i}{\bar{x}_i} \right\}$ implying that the inputs are perfectly complementary such that an increase in output requires an increase in *all* inputs. For σ=∞ we obtain the linear production function, $\frac{y}{\bar{y}} = \sum_i \left(\theta_i \frac{x_i}{\bar{x}_i} \right)$, implying that an increase in output minimally requires an increase in only *one* input. Another special case of the CES function used extensively in economics is the Cobb-Douglas function (σ=1)

$$\frac{y}{\bar{y}} = \prod_i \left(\frac{x_i}{\bar{x}_i} \right)^{\theta_i} \tag{2}$$

With this normalization of the production and utility functions, the share parameters, θ_i, are just the ratio of the base year industry expenditure on input i, $\bar{e}_i$, with the value of the function output, $\bar{r}_y$: $\theta_i = \frac{\bar{e}_i}{\bar{r}_y}$. The share parameters are used to calibrate the model so that the output is consistent with data from a base year or base period. The functions incorporating these share parameters are then nested to form representations of the various industries and consumers in the model. The nested function structure is typically represented by a graph; a basic production nest is shown in Figure 1.

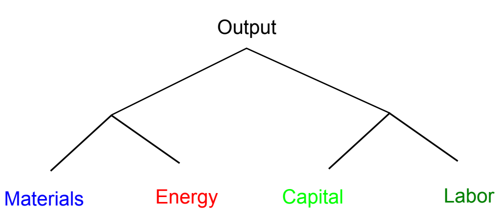

Figure 1: Basic nest for production

Each node of the tree is a CES function with a unique elasticity parameter that aggregates the inputs coming into it from below. The highest level node then aggregates the two intermediate bundles into the total industry output.

In this study, we synthesized econometric estimates of elasticity parameters from GTAP estimates [19], a recent estimate of elasticities in US industry from a historical BEA dataset [3], and reviews performed by the EPPA group [23, 26]. For ease of comparison with previous studies, and because Cobb-Douglas elasticities have long been used in studies of the substitutability of capital and labor, we use distributions centered on a Cobb-Douglas mean. The share parameters were calibrated exclusively with the GTAP version 7 database of global expenditure values [11]. The nested structure of the production and utility functions in this study are loosely based on that used by the EPPA group [2]. For more details, see the CIM-EARTH model documentation [13].

The static general equilibrium core is roughly the same as any other static CGE model [10,20,21], so the sensitivity results reported herein should apply to CGE modeling in general. The model configuration with which we work has moderate-scale spatial (16 regions), temporal (60 year horizon at one year time steps), and sectoral (16 production sectors plus 16 importers per region) resolutions. Dynamics in the CIM-EARTH prototype used in this work employ a recursive-myopic strategy in which most important drivers of economic growth and development (e.g., labor productivity and supply, energy efficiency, resource availability) are modeled with exogenous time trends roughly based on the dynamic equations of the EPPA model [2].

There are many ways to employ CGE models to forecast economic changes. For convenience, we restrict our analysis to single year (the horizon year, 2064) and time-series (2004-64) forecasts of specific model variables such as GDP and fossil fuel CO_2 emissions. We do not address sensitivity for variables relevant to comparison forecasts, ensemble simulations that test the differences between policy options (or sets of 'policy parameters' such as emissions prices or industry subsidies) or scenarios, which could have quite different responses to uncertainty. We leave such issues for future studies.

The results obtained from solving CGE models are highly dependent upon the choices of values for the elasticity and share parameters and thus on the data from which they are estimated. Because we have taken substitution elasticities from exogenous econometric estimations, we henceforth refer to their uncertainty as exogenous parametric uncertainty, or simply *parameter uncertainty*. This uncertainty is distinguished from the uncertainty due to error in the GTAP data set used to calibrate the share parameters, which we refer to as *calibration data uncertainty*.

3 Methodology

We performed two large sets of simulations to explore the sensitivity of model output to calibration data and parametric uncertainty, respectively. In both cases, we performed Monte Carlo sampling over uncorrelated Gaussian distributions. In the first set of 10,000 runs, we sample from the distributions of 16 expenditure data values to explore the implications of uncorrelated Gaussian noise in the calibration data set, and 5,000 runs to explore parametric uncertainty in elasticities. To handle the large scale parallelization of the CGE model, which enabled us to explore its parameter space, we developed a computational framework using the Swift parallel scripting system [28].

3.1 Calibration data uncertainty

We calibrate CIM-EARTH to a base year (2004) using the GTAP version 7 database [11]. That is, we tune the model share parameters so that the model gives results for 2004 that are consistent with this global expenditure database, as described in Section 1. The full database has 113 regions (R), 57 sectors (S), 5 base factors (F) and is summed to the $R \times S \times F$ aggregate database that is required for a particular question. A region can be anything from a single country to the whole globe, a sector can be as specific as raw milk or iron and steel, or as general as agriculture and industry, and base factors are skilled and unskilled labor, capital, land, and natural resources. For this study we use the $16 \times 16 \times 4$ aggregate listed in Figure 1. The total number of expenditure values in a particular $R \times S \times F$ aggregation of the GTAP database is

16 regions	**16 sectors** (per region)
Oceania	Agriculture and forestry
Southeast Asia	Coal
Japan	Oil
Rest of East Asia	Natural gas
India	Iron and Steel
Rest of South Asia	Chemicals
Russia, Georgia & Asiastan	Non-ferrous metals
Middle East & N. Africa	Cement/Mineral products
Sub-Saharan Africa	Other manufacturing
Western Europe	Refined petroleum
Rest of Europe	Electricity
Brazil	Land transport
Mexico	Air transport
Rest of Latin America	Sea transport
USA	Government services
Canada	Other services

Table 1: Aggregate regions and sectors for the 16×16 model used here.

$RS(2S+F)$ producer demand expenditures, R^2S importer demand expenditures and $R(2S+1)$ consumer demand expenditures (including consumer demand for savings). Thus, our selected aggregation reduces the number of expenditure values from about 1.5 million in the full database to just over 14,000 values in the aggregated database. Ideally, we would perturb the raw dataset before aggregation, but this capacity is not yet available due to both technical data formatting and processing issues and, more importantly, to the large number of expenditure values in the fully disaggregated database. About 15% of these 14,000 values are ignored for being negligibly small (less than \$100K) and another 8% are nearly so (less than \$1M), yet the parameter space is still large from the perspective of Monte Carlo uncertainty analysis. For this study, we perturb the 1600 largest (relative to the total expenditure of the relevant industry or consumer), which amounts to 3-5 values per sector (consumer and importer) per region. While we restricted the analysis to this smaller sample size to make the ensemble simulations more tractable, the parameter space of perturbed values still accounted for more than 75% of global expenditure. In Section 4.1 we address the feasibility of studies on the complete aggregated database and the full disaggregated database.

We set the width of the Gaussian distributions about the mean expenditure values to one of three different values representing our estimate of the reliability of economic data among various regions. Highly developed countries – the United States, Canada, Western Europe, Japan, Australia and New Zealand (most of the OECD countries) – are believed to have well established structures in place for consistent and accurate data gathering. Therefore, we assume that noise in reported expenditure values from these countries is small and set the standard deviation (s.d.) of each sampled distribution at 3% of the mean. In contrast, we assume that poorly developed countries and countries notorious for having data inconsistencies – China, South Asia, Asiastan (Central Asia), Sub-saharan Africa and much of Latin America – have a relatively large amount of noise in reported expenditure values, modeled as parameter distributions with standard deviations of 7% of the mean. All other regions – Mexico, Eastern Europe, Middle East/North Africa, South East Asia and Russia – were modeled with expenditure data distributions having 5% standard deviations around their means. The primary source of error in the GTAP database is likely reporting error, but there are other possible issues that could lead to noise in the dataset. For example, of the 87 regions and countries that were included in the GTAP 6 database (base year 2001), slightly less than half were updated with current datasets in GTAP 7. The others are updated through a simple rescaling of the data to the new base year. Unsurprisingly, the well established nature of data gathering agencies in OECD countries leads to more frequent data updating and thus little need to rescale. However, most countries in the high-error and many in the mid-error bracket defined for this study(including China, South Asia and most of Africa and Latin America) were not updated with new datasets in GTAP 7. We recognize the characterization of data errors is somewhat simplistic, but nevertheless believe it to represent a considerable improvement over current practice, which ignores this error source.

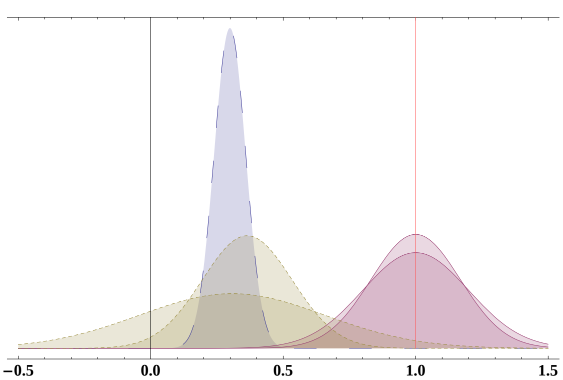

Figure 2: A comparison of parameter distributions for σ_{KL} in agriculture and forestry. Blue dashed: GTAP estimate with 20% s.d.; Yellow dotted: two estimates from Balistreri et al. 2003 [3]; Red solid: two distributions with Cobb-Douglas mean and s.d. estimate from EPPA [23] and s.d.=0.20.

We do not address the possibility of correlated errors in the calibration dataset. It is sometimes stated, for example, that local political agencies in China have a strong incentive to report higher numbers to inflate the successes of industry in their region [16]. Establishing the existence and extent of sources of correlated errors will require more detailed studies of the underlying data structures of our models. We leave both tasks for future studies.

3.2 Parameter uncertainty

We explore model sensitivity to parameter uncertainty in *all* substitution elasticities (71 values), and in one key subset, Armington trade elasticities (16 values), in order to explore relative sensitivities and assign priorities for future parameter studies based on sensitivities. Previous studies of the sensitivity of CGE model results to uncertainties in substitution elasticity parameters [23,26] only included subsets of substitution elasticity parameters, such as the elasticity in each industry between labor and capital.

Given the sizable discrepancies that exist in the econometrics literature between estimates of substitution elasticities, it is difficult to establish a consistent basis for producing uncertainty distributions for these parameters. Instead of trying to combine disparate and often contradictory estimates of means and standard errors, we chose to center the parameter distributions at relatively standard values with standard deviations set to 20% of the mean. For ease of comparison between this current study and other studies in the CGE literature, which predominantly use Cobb-Douglas production functions to model substitution between labor and capital, we have chosen σ_{KL}, the elasticity of substitution between capital and labor, to have mean one independent of region and sector. Figure 2 shows an example of elasticity parameter distributions. Two of the five distributions shown correspond to a recent study by Balistreri [3]: one for farms with mean 0.307 and a wide distribution, and another for agriculture, forestry and fishing services with mean 0.364 and a much tighter confidence. Two other distributions correspond to the input sample used in this study and the input sample described by Sokolov et al. [23], both with Cobb-Douglas mean. The final distribution shown has a mean estimate from GTAP [19] with 20% relative standard deviation. Figure 9 shows the capital-labor elasticity parameter for many other industries. For Armington elasticities we use estimates produced by the GTAP group [19] and for intermediate nest elasticities, we use values from the EPPA group [21], both with 20% standard deviations.

3.3 Computational studies

We performed two large sets of simulations to explore the sensitivity of model output to calibration data and parametric uncertainty, respectively. In both cases, we performed Monte Carlo sampling over uncorrelated Gaussian distributions. In the first set of 10,000 runs, we sample from distributions of over 1,600 expenditure data values to explore the implications of uncorrelated Gaussian noise in the calibration data set. The second large study contains 5,000 runs to explore parametric uncertainty in the full elasticity set and 1,000 runs to study the magnitude of the contribution from the Armington subset only. To handle the large scale parallelization of the CGE model, which enabled us to explore its parameter space, we developed a computational framework using the Swift parallel scripting system [27, 28]. All told, we employed roughly 30K CPU-hours (0.4-1.6 hours per run, depending on many factors) for the prototype ensemble studies, spread over several batches of jobs on TeraGrid and Open Science Grid computers: Firefly (U. Nebraska, OSG), QueenBee (Louisiana State University, TG), Ranger (Argonne National Lab, TG), TACC (U. Texas, TG) and TeraPort (Computation Institute, OSG). Simultaneous processor usage peaked at about 2K in these runs.

4 Results

We find stark contrasts in the relative sensitivity of different output variables of the CIM-EARTH model to calibration data and elasticity parameter noise. We find substantial differences in output sensitivity between variables in different regions and at different levels of aggregation, from which we construct a basic classification of forecast variable sensitivities in order to begin to answer the question of what can and cannot be learned from CGE modeling and estimate the ensemble sizes that are required to fully characterize the many uncertainties. One basic conclusion is that the robustness to parametric and data uncertainty of a model's conclusions depends strongly on the particular model outputs upon which the conclusion relies. We examine many regions and sectors as well as several levels of output variable aggregation:

1. Global aggregates: global GDP and global CO_2 emissions (Figures 6 and 5),
2. Region aggregates: regional GDP (Figure 10), regional emissions (Figures 13) and regional industry demands for electricity and refined petroleum products (Figures 14 and 15),
3. Sector specific aggregates: global industry production levels for steel and iron, chemicals, cement, and non-ferrous metals aggregated over regions, global consumer and industrial demand for electricity and refined petroleum products,
4. Region and sector specific revenue variables: regional industrial production levels for steel, chemical, cement, etc. (Figures 11 and 12),
5. Micro variables: regional consumer demand for electricity and refined petroleum products (Figures 16 and 17).

As expected, we find that output variables at a higher aggregation level display less sensitivity to both calibration data noise and elasticity parameter uncertainty. We collect most results in Tables 2 and 3.

4.1 Calibration data uncertainty

Table 2 gives the standard deviations of a selection of model output variables for a variety of regions and aggregation levels in the base (2004) and forecast (2064) years. The relative difference provides a basic measure of the strength of the model response to uncorrelated Gaussian noise in the calibration data. Not surprisingly, larger scale variables (aggregations of a larger number of perturbed variables) such as global GDP and emissions (vs. regional GDPs and emissions) are generally less sensitive to the calibration data noise. Comparisons among different variables (between global GDP and global emissions for example) are less obvious, though similar patterns are still apparent, especially in the forecast year. Variations between the rate of change in s.d. among output variables are more mixed, but a primary determining factor appears to be specific regional properties, with more developed countries generally having more stable responses to the perturbations.

In order to explore the model response to calibration data noise we apply a basic metric of linear variable response by measuring the correlation of selected variable's base year and horizon year values. Figure 3 shows an example of this metric for the model output of Western Europe's fossil fuel CO_2 emissions. In Table 2 we show this linearity metric for a wide range of model output variables and regions. Interestingly, this linearity appears to be degraded in aggregation; that is, global aggregates of variables – such as GDP, emissions and steel industry revenues – appear to have a significantly more nonlinear response to the input perturbations. This result is in some contrast to the behavior of distribution widths at varying level of aggregations, as described above.

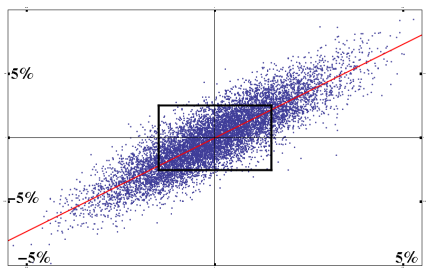

Figure 3: Relationship between model response (percent deviation from the mean) in the base year (x-axis) and horizon/forecast year (y-axis) for CO_2 emissions in Western Europe. The linear fit (red line) is a first-order estimate of the model's response to perturbed inputs in the calibration data. The linear correlation is presented in Table 2. The rectangle is one s.d. from the mean in each dimension.

The slope of the linear fit, s, (also reported in Table 2) provides a first-order estimate of the model response to perturbation. This metric describes how much growth of the distribution width can be assigned to this linearity. For correlation c we then have the relationship between the standard deviations

$$\frac{\sigma_{(64)}}{\sigma_{(04)}} = \frac{s}{c}\,.$$

The correlation is then interpreted naturally as the fraction of the increase in the standard deviation (the relative 'spread' of the input perturbation) that is due to this linear component.

It is tempting to assume that Gaussian noise can be characterized in larger parameter spaces than are considered here, due to the near linear response of many relevant model variables. For example, it may be possible to roughly characterize a model's response to calibration data noise throughout a forecast trajectory without simulating the entire model trajectory (perhaps even simply by solving the static base year model). Though the linearity of the response is not as strong for large scale aggregates such as global CO_2 emissions (an important quantity for obvious reasons), their relative sensitivity to the data noise is much less overall, ameliorating this potential concern.

It is also tempting to assume that the linearity of the model's response to calibration data noise implies that this error will compound with other sources of uncertainty in an approximately additive (as opposed to the usual multiplicative) way. If true, such a relationship would greatly simplify attempts to include this effect consistently with other parametric uncertainties, since ensembles of calibration data noise could be studied independently and combined with other uncertainties and since approximate characterizations of the response of a model (or class of models) to this noise could be calculated offline and applied to many applications. More studies are needed on the composite effects of uncertainty in different parameter sets (share/calibration parameters, elasticity parameters, dynamic equation parameters, etc.) to validate this assumption. However, the findings here are certainly a positive indication. Further, as a practical matter, we feel that the overwhelming challenge of doing true characterizations of model response to uncertainty in the full space of calibration data for each study independently, compels an approximate solution of this type.

Evaluating sensitivity to run set size. Next we construct a metric based on resampled distributions from the full ensemble to explore the extent to which a smaller run-set would have sufficiently characterized the uncertainty in each example model output variable explored here. Resampling techniques are used frequently in modern data analysis to explore the extent to which statistical measurables of a data set can be taken to be the correct values [12]. Briefly, the idea is to populate a new 'resampled' ensemble by pulling (or resampling) subsets of a fixed size from the original ensemble and calculating the statistical measurable of interest for the subset. The measurable from each resampled subset is then a single element in the resampled ensemble. The mean and standard deviation of the resampled ensemble then gives some information about

Table 2: Standard deviations and linear response measures for select variables in the global aggregate and for regions representative of the three data error classifications.

Variable	Aggr. Region	Standard dev.[a] 2004	Standard dev.[a] 2064	Linearity measures[b] Slope	Linearity measures[b] Correlation
GDP	Global	0.7%	0.8%	0.847	0.6944
	W. EU	1.2%	1.8%	1.297	0.8486
	Mexico	1.6%	2.5%	1.169	0.6827
	China	2.0%	4.2%	1.387	0.6295
CO_2 emissions	Global	0.7%	1.7%	1.300	0.5376
	USA	1.4%	2.0%	1.223	0.8790
	W. EU	1.5%	2.5%	1.464	0.8665
	Mexico	3.1%	4.1%	1.181	0.8813
	Russia	2.3%	4.1%	1.363	0.7604
	China	1.9%	3.7%	0.968	0.5079
Steel revenue	Global	0.9%	2.1%	1.128	0.4968
	USA	1.8%	2.5%	1.207	0.8373
	Russia	1.6%	4.7%	1.851	0.6381
	China	2.6%	3.5%	0.847	0.6221
Industrial electricity demand	Global	0.3%	1.5%	2.115	0.3624
	USA	0.4%	1.3%	2.349	0.7915
	Mexico	1.1%	2.6%	1.476	0.6175
	China	0.9%	3.4%	1.154	0.2991
Consumer electricity demand	Global	0.8%	2.2%	1.843	0.6750
	USA	2.0%	3.4%	1.691	0.9774
	Russia	2.3%	4.3%	1.461	0.7821
	China	3.4%	7.4%	1.507	0.6972

[a] relative standard deviation measured as % variation from mean
[b] linear correlation measure as in figure 3

additional forecast uncertainty due to the limited sample size. We perform resampling (with replacement) on the mean and standard deviation measurements for each variable listed in Table 2 using resample sizes (the size of the subset pulled from the original ensemble to calculate each measurable that populates the resampled ensemble) varying from 500 to 10,000. We use these resampled ensembles for two distinct purposes: to estimate the statistical uncertainty inherent in the forecast measurables with the sample sizes used in the study in order to show that the sample sizes were indeed large enough to fully characterize the forecast uncertainty resulting from the perturbations to the calibration data, and to estimate a minimum number of performed runs that would have given a mean and standard deviation measurement for the variable within an acceptable deviation from the result of the 10,000 runs, with high ($\simeq$95%) confidence.

For example, Figure 4 shows plots of the standard deviation of the resampled ensemble of the mean and s.d. of Chinese consumer demand for electricity (one of the most sensitive variables considered here). The standard deviation of the resampled ensemble of the mean and of the standard deviation is about 0.1% of the full sample mean and 1% of the full sample standard deviation (respectively) at a resample size of about 5,000 runs. Thus, if we had only performed $\simeq$5,000 runs instead of 10,000, we would have had a $\simeq$95% chance of measuring the same mean within $\pm$0.2% of the full sample mean and the same standard deviation within $\pm$2%. Together these results imply that, using half as many runs, we could predict, with 95% confidence the standard deviations of Table 2 to within 2-8% of the stated measurement from the full ensemble. This additional uncertainty is not substantial, though it is probably good to keep this additional error source below 10% to ensure that the forecast impact of data calibration noise are not being washed out by the statistical noise.

All other variables that we examined have the same or less additional statistical uncertainty when restricted to half the runs, so we conclude that the full model output could indeed have been produced with roughly half the runs. Indeed, if only large scale global aggregate variables such as global GDP and emissions are required, even fewer runs could be performed to give a satisfactory confidence interval. Along with the strong linearity in the model response to these perturbations, this result is further evidence that a characterization of the forecast sensitivity to uncorrelated noise in the full calibration data set can be accomplished

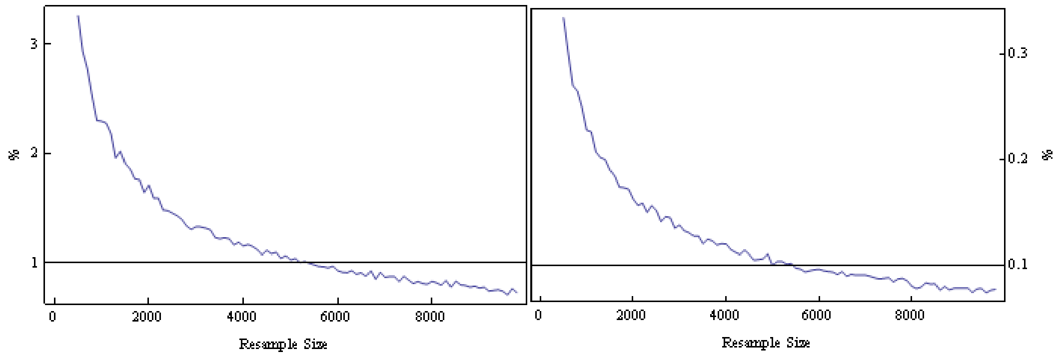

Figure 4: Bootstrap curves are used to determine the additional forecast uncertainty in statistical measurables caused by sample sizes that are not infinite. Each plot is generated by performing an independent bootstrap resampling measure with the resample size, the number of elements from the original ensemble that are resampled (with replacement) repeatedly to calculate the statistical measure that is in question, varied from 500 to 10,000. They are also used to estimate a minimum number of runs that can reproduce the statistical results of the full ensemble. Right: the resampled statistical measurable is the mean of the Chinese consumer demand for electricity in 2064. Left: the resampled statistical measurable is the standard deviation of the Chinese consumer demand for electricity in 2064. Both plots show the standard deviation of the resampled distribution of this measurable.

with a reasonably sized ensemble. Further, cancellations between noisy components that occur in the many levels of aggregation (in the aggregated expenditure values when we aggregate the calibration dataset from 113×57 to 16×16 and then in the aggregate of output variables from micro up to macro) means that it may be quite possible to get a robust understanding of the impacts of this noise, even for the extremely large parameter spaces described in Section 3.

4.2 Parameter uncertainty

Sensitivity of the model to perturbations in the elasticities is presented as relative variation from the mean over time for many of the model outputs described previously. Table 3 gives the uncertainty in the forecast year (2064) results for many variables at different levels of aggregation. Model response to uncertainty in elasticity parameters is very different from the response to calibration data noise (share parameter uncertainty). The calibration effectively tunes the static CGE in the base year to very near the 'correct' equilibrium solution (the one that reproduces the data) so that very few substitutions are made (and thus the output of the static model in the base year is very nearly independent of the elasticity parameters, modulo a few interesting examples that we consider in more depth below). Further, since elasticity parameters are taken to be the same across regions (industries are modeled the same way, independent of their locations), we do not see the noise cancellation in global aggregates (relative to regional variables) that we did for calibration data noise. However, the nonlinear response of model output variables to relative price fluctuations between commodities and factors as time goes forward is highly dependent on this parameter set.

Figure 5 shows global gross emissions and relative sensitivity to perturbations in substitution elasticities. The standard deviation of the global CO_2 emissions forecast grows to about 20% of its mean by 2064. Global GDP, on the other hand, (Fig. 6) can be forecast with substantially more confidence, with a standard deviation due to elasticity uncertainty that grows to only about 2.5%.

This stark distinction illustrates dramatically how much long-term forecast confidence differs among output variables. There are many reasons for the discrepancy in sensitivities to parameter noise between GDP and CO_2 emissions variables, primary among which are the facts that GDP is an aggregate of a larger assortment of micro model variables and is dominated by significantly more stable economies (USA, Western Europe, etc.) than those that dominate CO_2 emissions (primarily China). Output variables tend to be significantly less sensitive to uncertainty in substitution elasticities in these stable, high GDP economies, primarily because relative price fluctuations in these economies are significantly less pronounced. (Recall that elasticity parameters determine how easily industries and consumers can substitute between various commodities and primary factors as relative prices change.)

In fact, it turns out that GDP and fossil fuel emissions are not correlated at all for highly developed countries such as the United States and Western Europe, though for China and India (and other developing

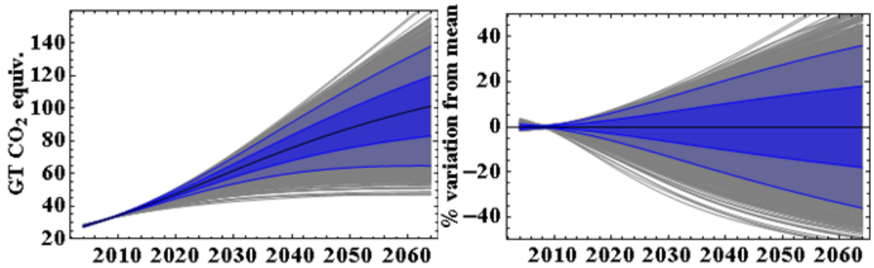

Figure 5: (Left) Global emissions from fossil fuel consumption for 5000 model runs with perturbed substitution elasticities, and (right) the relative sensitivity of global CO_2 emissions.

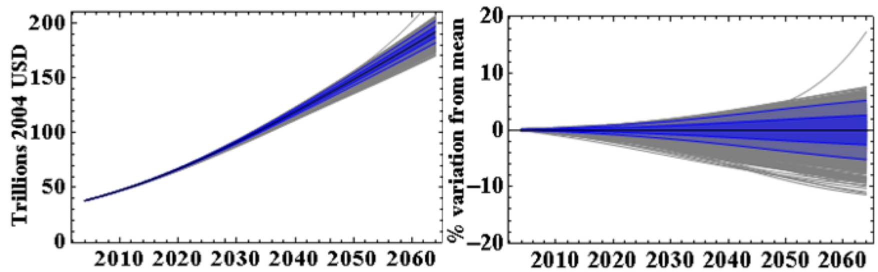

Figure 6: (Left) Global GDP for 5000 model runs with perturbed substitution elasticities, and (Right) the relative sensitivity of global GDP.

countries) there is a strong positive correlation between GDP and emissions (Figure 7).

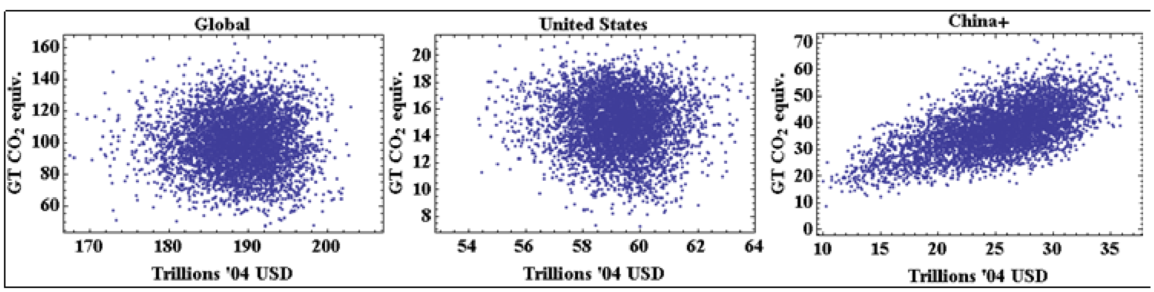

Figure 7: Forecast year GDP vs. fossil CO_2 emissions for world, the USA and China.

These differences probably account for most of the discrepancy in sensitivity between these two variables. As can be seen from Table 3 and the figures in Appendix B, the forecast sensitivity for GDP (Figure 10) and CO_2 emissions (Figure 13) in various regions falls generally into the expected hierarchy of aggregation levels. Figure 8 shows the model forecast for global and regional carbon intensity (fossil fuel CO_2 emissions per unit GDP) for the 5,000 runs in the ensemble. Regional carbon intensities have forecast year uncertainties largely comparable to the uncertainty in emissions themselves (i.e., we find little correlation between the response of GDP and emissions variables in nearly all regions).

4.3 The Armington subset

To begin to explore the major sources of this parametric sensitivity, we look at the model forecast response to perturbations of the 16 Armington trade elasticities in the model. These parameters control the substitutability of the domestic and imported versions of the 16 commodities (resp.). Parameters here are assumed to be the same in every region and each parameter realization is constrained to satisfy the so-called 'rule of two' [19], i.e. the elasticity between imports from various regions is set to twice the Armington elasticity between domestic and imports. We leave explorations of differences between countries and violations of the 'rule' for future work.

Variable	Aggr. Region	Forecast year s.d. Full set	Arm. subset
GDP	Global	2.62%	0.57%
	W. EU	4.46%	0.34%
	Mexico	6.16%	0.55%
	China	18.00%	2.25%
Steel revenue	Global	12.73%	3.08%
	USA	7.70%	2.01%
	Russia	10.14%	1.26%
	China	17.37%	5.38%
CO_2 emissions	Global	18.07%	1.06%
	USA	14.61%	0.56%
	W. EU	16.06%	0.46%
	Mexico	16.41%	0.39%
	Russia	24.10%	1.34%
	China	24.32%	3.11%
Industrial electricity demand	Global	11.81%	1.54%
	USA	4.35%	0.28%
	Mexico	11.71%	0.30%
	China	19.11%	3.95%
Consumer electricity demand	Global	16.61%	1.06%
	USA	8.66%	0.22%
	Russia	29.71%	1.20%
	China	23.48%	3.17%

Table 3: Forecast year (2064) standard deviations resulting from uncertainty in model elasticity of substitution parameters.

The results are collected in Table 3, and compared to the forecast year standard deviations from the full parameter set. Not surprisingly, variables are most sensitive in regions and sectors where international trade is essential. Variables in China show an especially substantial sensitivity to perturbations in Armington parameters, relative to other regions. Sizable sensitivities are also exhibited by industry revenue variables that have a sizable revenue component from global trade (as in the steel revenue variable from Table 3. While these sensitivities appear relatively small compared to the model response to perturbations in the full elasticity parameter set, it will be essential to account for this uncertainty in particular in a study of the impacts of climate policy on international trade and carbon leakage.

Acknowledgments

This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357.

References

[1] D. Abler, Rodriguez, A.G., and J. Shortle, *Parameter uncertainty in cge modeling of the environmental impacts of economic policies*, Environmental and Resource Economics, 14 (1999), pp. 75–94.

[2] M. B. Babiker, J. Reilly, M. Mayer, R. S. Eckaus, I. S. Wing, and R. C. Hyman, *The MIT emissions prediction and policy analysis (EPPA) model: Revisions, sensitivities, and comparisons of results*, Tech. Rep. 71, MIT, 2001.

[3] E. J. Balistreri, C. A. McDaniel, and E. V. Wong, *An estimation of u.s. industry-level capital-labor substitution*, Computational Economics 0303001, EconWPA, Mar. 2003.

[4] M. Berrittella, A. Bigano, R. Roson, and R. S. J. Tol, *A general equilibrium analysis of climate change impacts on tourism*, TOURISM MANAGEMENT, 27 (2006), pp. 913–924.

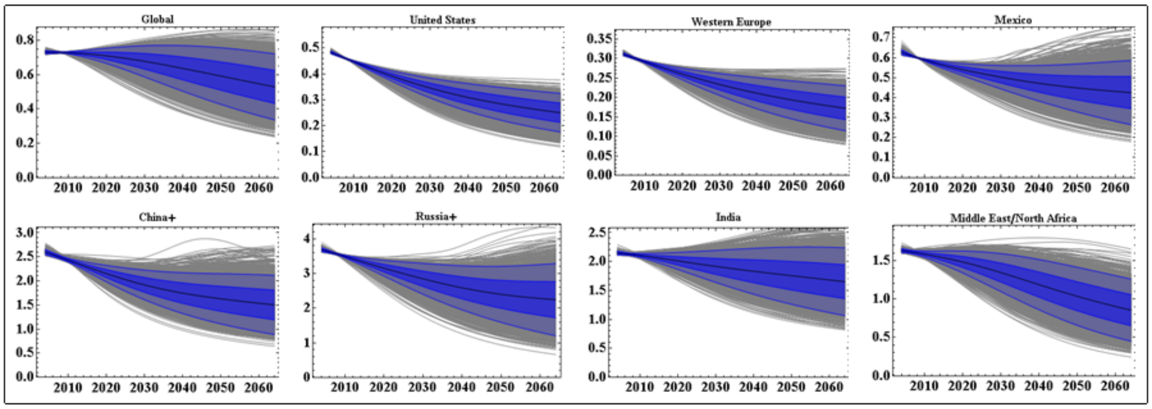

Figure 8: Carbon intensity forecast for the world and for 7 of 16 model regions in kg CO_2 equivalent fossil fuel emissions per '04 USD of GDP.

[5] M. Berrittella, A. Y. Hoekstra, K. Rehdanz, R. Roson, and R. S. J. Tol, *The economic impact of restricted water supply: A computable general equilibrium analysis*, WATER RESEARCH, 41 (2007), pp. 1799–1813.

[6] C. Boehringer, T. Rutherford, and W. Wiegard, *Computable General Equilibrium Analysis: Opening a Black Box*, Discussion Paper 03-56, ZEW, 2003.

[7] F. Bosello, R. Roson, and R. S. J. Tol, *Economy-wide estimates of the implications of climate change: Human health*, ECOLOGICAL ECONOMICS, 58 (2006), pp. 579–591.

[8] ——, *Economy-wide estimates of the implications of climate change: Sea level rise*, Environmental and Resource Economics, 37 (2007), pp. 549–571.

[9] ——, *Economy-wide estimates of the implications of climate change - a rejoinder*, ECOLOGICAL ECONOMICS, 66 (2008), pp. 14–15.

[10] J.-M. Burniaux and T. Truong, *Gtap-e: An energy-environmental version of the gtap model*, GTAP Technical Papers 923, Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University, 2002.

[11] B. Dimaran and R. M. ed, *The GTAP database- version 7*, tech. rep., Global Trade Analysis Center, Purdue University, 2007.

[12] B. Effron and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, 1993.

[13] J. Elliott, I. Foster, K. Judd, E. Moyer, and T. Munson, *Introducing CIM-EARTH*, Tech. Rep. ANL/MCS-TM-307, MCS Technical Memorandum, 2009.

[14] R. Gerlagh and B. Van der Zwaan, *A sensitivity analysis of timing and costs of greenhouse gas emission reductions*, CLIMATIC CHANGE, 65 (2004), pp. 39–71.

[15] T. Hertel, D. Hummels, M. Ivanic, and R. Keeney, *How confident can we be of CGE-based assessments of Free Trade Agreements?*, ECONOMIC MODELLING, 24 (2007), pp. 611–635.

[16] C. Holz, *"Fast, clear and accurate": How reliable are Chinese output and economic growth statistics?*, CHINA QUARTERLY, (2003), pp. 122–163.

[17] D. Kendrick, *Stochastic control for economic models: Past, present and the paths ahead*, Journal of Economic Dynamics and Control, 29 (2005), pp. 3–30.

[18] S. Kim, *Uncertainty, political preferences, and stabilization: Stochastic control using dynamic cge models*, Computational Economics, 24 (2004), pp. 97–116.

[19] J. Liu, C. Arndt, and T. Hertel, *Parameter estimation and measures of fit in a global, general equilibrium model*, Journal of Economic Integration, 19 (2004), pp. 626–649.

[20] H. Lfgren, R. L. Harris, and S. Robinson, *A standard computable general equilibrium (CGE) model in GAMS*, TMD discussion papers 75, International Food Policy Research Institute (IFPRI), 2001.

[21] S. Paltsev, J. Reilly, H. Jacoby, R. Eckaus, J. McFarland, M. Sarofim, M. Asadoorian, and M. Babiker, *The mit emissions prediction and policy analysis (eppa) model: Version 4*, Tech. Rep. 125, MIT, 2005.

[22] A. Sokolov, C. Schlosser, S. Dutkiewicz, S. Paltsev, D. Kicklighter, H. Jacoby, R. Prinn, C. Forest, J. Reilly, C. Wang, B. Felzer, M. Sarofim, J. Scott, P. Stone, J. Melillo, and J. Cohen, *The MIT Integrated Global System Model (IGSM) Version 2: Model Description and Baseline Evaluation*, Tech. Rep. 124, MIT, 2005.

[23] A. P. Sokolov, P. H. Stone, C. E. Forest, R. Prinn, M. C. Sarofim, M. Webster, S. Paltsev, C. A. Schlosser, D. Kicklighter, S. Dutkiewicz, J. Reilly, C. Wang, B. Felzer, J. M. Melillo, and H. D. Jacoby, *Probabilistic Forecast for Twenty-First-Century Climate Based on Uncertainties in Emissions (Without Policy) and Climate Parameters*, JOURNAL OF CLIMATE, 22 (2009), pp. 5175–5204.

[24] I. Sue Wing, *Computable general equilibrium models and their use in economy-wide policy analysis*, Tech. Rep. Technical Note No. 6, Joint Program on the Science and Policy of Global Change, 2004.

[25] R. Tol and S. Fankhauser, *On the representation of impact in integrated assessment models of climate change*, Environmental Modeling and Assessment, 3 (1998), pp. 63–74.

[26] M. Webster, S. Paltsev, J. Parsons, J. Reilly, and H. Jacoby, *Uncertainty in greenhouse emissions and costs of atmospheric stabilization*, Tech. Rep. 165, MIT Joint Program Report Series, November 2008.

[27] M. Wilde, I. Foster, K. Iskra, P. Beckman, Z. Zhang, A. Espinosa, M. Hategan, B. Clifford, and I. Raicu, *Parallel scripting for applications at the petascale and beyond*, Computer, 42 (2009), pp. 50–60.

[28] Y. Zhao, M. Hategan, B. Clifford, I. Foster, G. von Laszewski, V. Nefedova, I. Raicu, T. Stef-Praun, and M. Wilde, *Swift: Fast, reliable, loosely coupled parallel computation*, in 1st IEEE International Workshop on Scientific Workflows, 2007, pp. 199–206.

A Parameter distributions

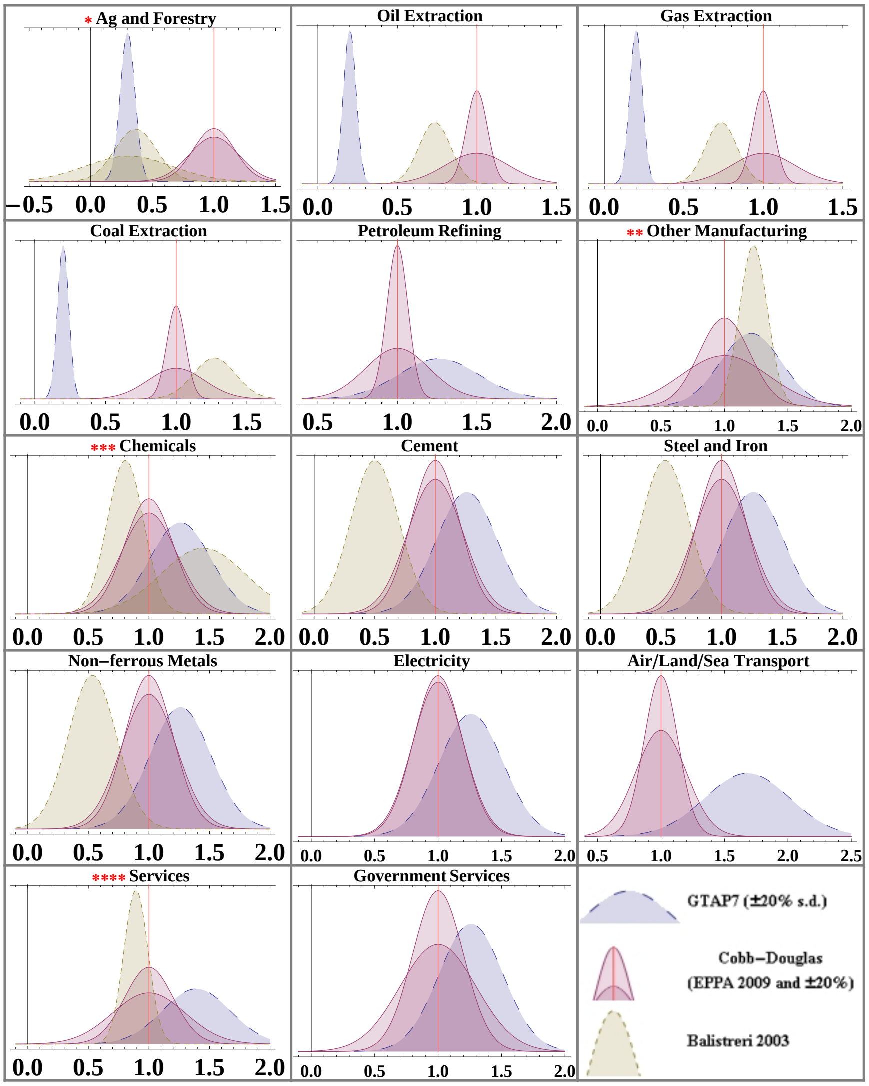

Figure 9: A comparison of parameter distributions for σ_{KL} for a variety of industries. The red line at σ=1 denotes the Cobb-Douglas point, and the black line at σ=0 denotes the Leontief or fixed-coefficients point. Some aggregate sectors have multiple estimates from [3] that are relevant: * Balistreri estimates two agriculture-related sectors: 'farms' and 'agriculture and forestry services'. ** Balistreri estimates many sectors relevant to generic manufacturing; the estimate for aggregated manufacturing and mining is shown here. *** Balistreri estimates two chemicals related sectors: 'rubber and misc. plastic products' and 'chemicals and allied products'. **** Balistreri only estimates one service sector: 'construction services', which should probably not be taken to be representative of aggregated services.

B Sensitivity measures

This appendix shows sensitivity plots for many variables in the aggregation hierarchy described in Section 4 and listed in Table 3, roughly ordered from the largest to the smallest scale aggregates.

Figure 10: Forecast sensitivity to elasticity uncertainty for **GDP** variables in 8 of 16 regions of the model.

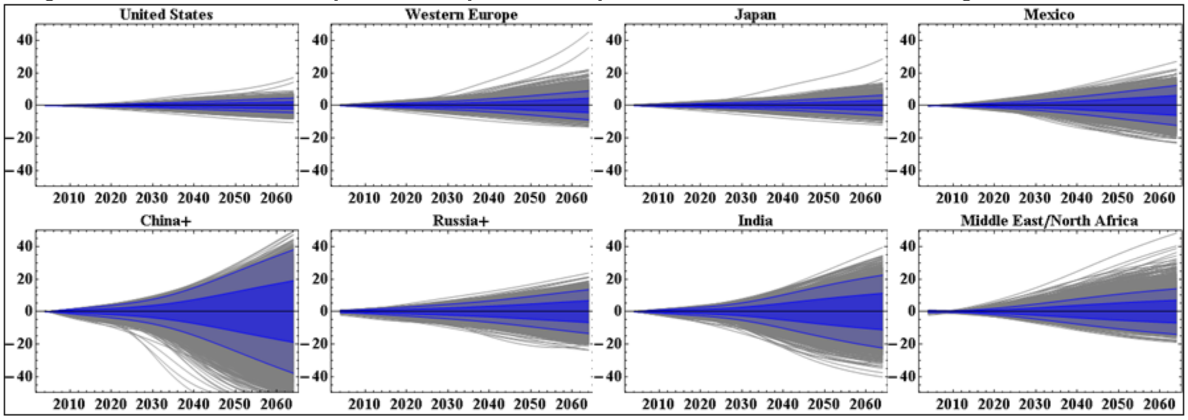

Figure 11: Sensitivity to elasticity uncertainty for the **revenue of the cement industry** in 8 of 16 regions.

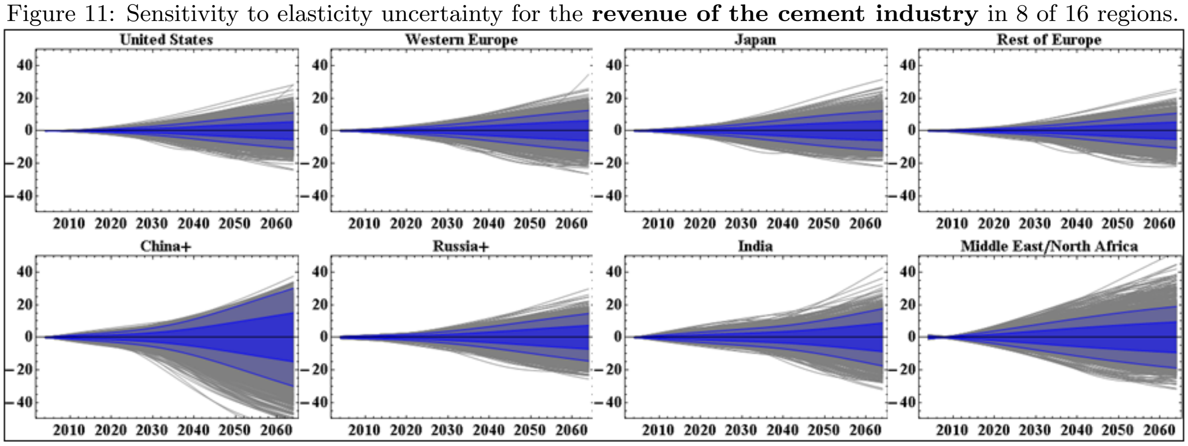

Figure 12: Sensitivity to elasticity uncertainty for the **revenue of the steel industry** in 8 of 16 regions.

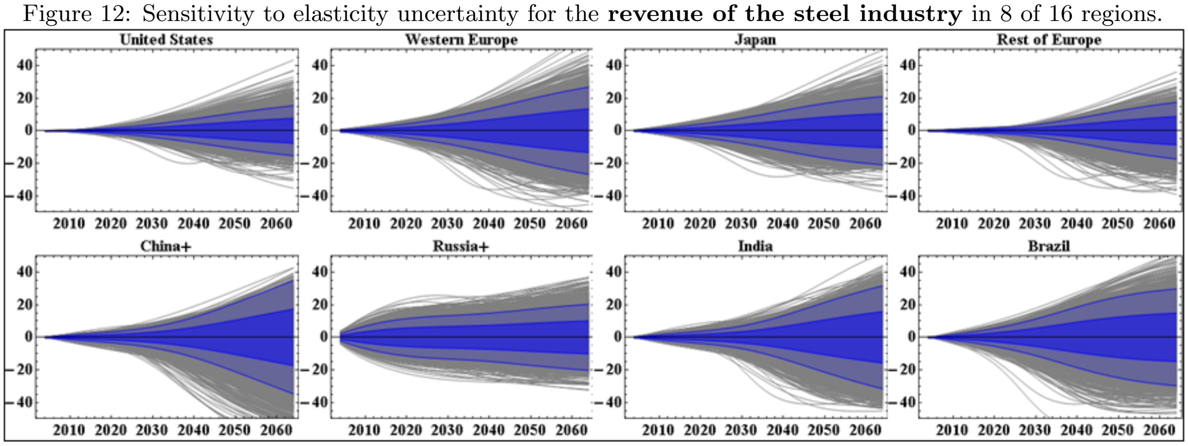

Figure 13: Sensitivity to elasticity uncertainty for $\mathbf{CO_2}$ **emissions** from fossil fuel consumption.

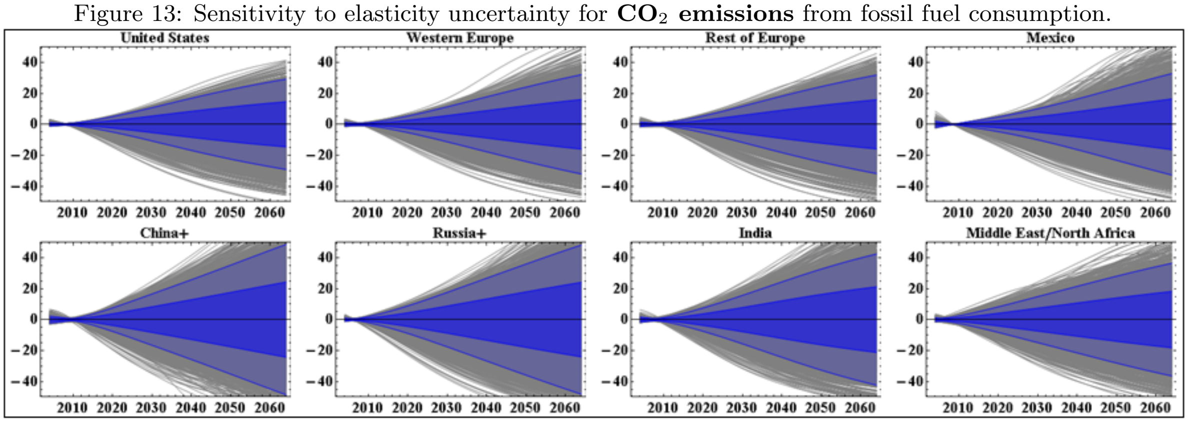

Figure 14: Output sensitivity of the **industrial demand for electricity** in 8 of 16 model regions.

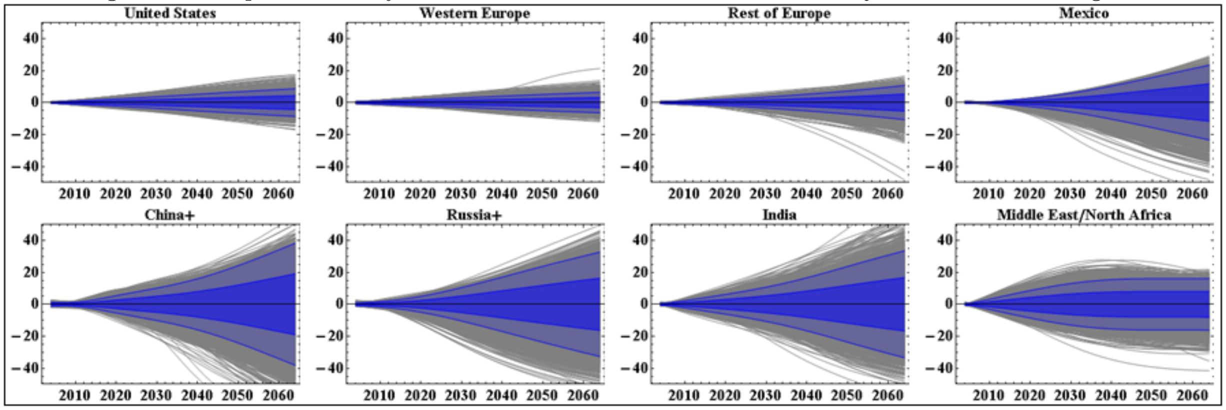

Figure 15: Output sensitivity of the **industrial demand for refined petroleum** in 8 of 16 model regions.

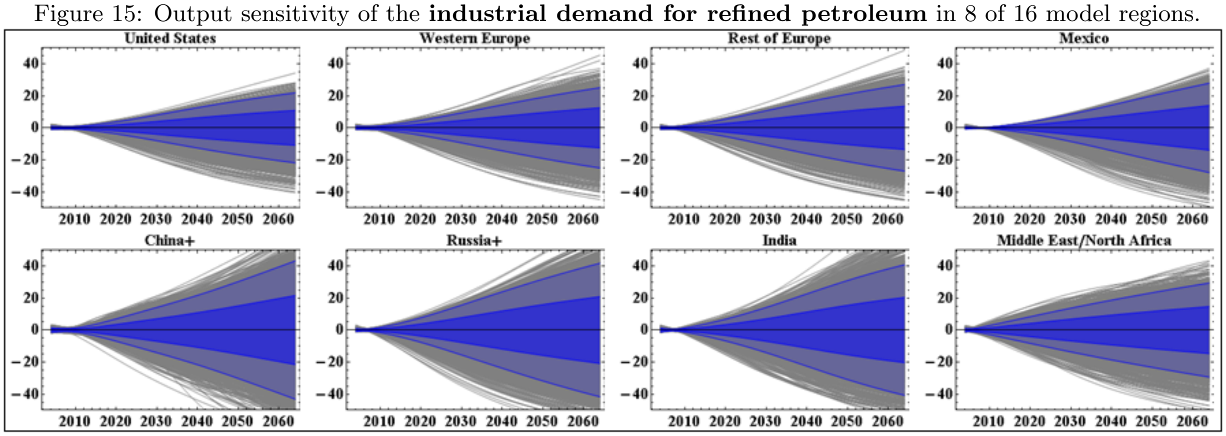

Figure 16: Output sensitivity of the **consumer demand for electricity** in 8 of 16 model regions.

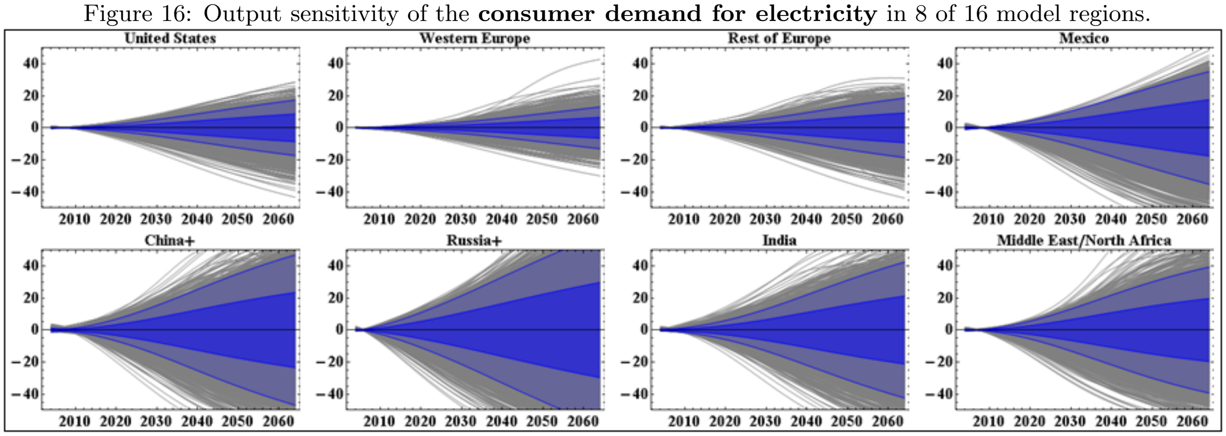

Figure 17: Output sensitivity of the **consumer demand for refined petroleum** in 8 of 16 model regions.