Estimation of Unknown Parameters in Dynamic Models Using the Method of Simulated Moments (MSM)

Abstract: We introduce the Method of Simulated Moments (MSM) for estimating unknown parameters in dynamic models. The MSM is useful when there are empirical data related to the behavior of different entities and error terms do not follow any well-established distribution. Statistical moments such as mean and variance of empirical data can be matched against the moments of model-generated data in order to estimate some structural parameters of the model. The major value of the MSM for estimating dynamic models is in its flexibility to be used with any type of data, including cross-sectional data, to estimate dynamic models.

Keywords: Method of Simulated Moments, parameter estimation, cross-sectional data, time series data

Introduction

Increasingly, dynamic modelers face problems where estimating model parameters from numerical empirical data is a requirement. This trend is partly motivated by increasing availability of numerical data from a large number of ongoing and one off data collection projects that survey different concepts of interest to dynamic modelers, from individuals and firms to disease incidences and measures of economic performance, just to name a few. Another driver of this trend is the increasing application of dynamic models, beyond case specific corporate projects, to theoretical and academic problems (Repenning 2003, Sterman 2006). In these cases generic models for a category of objects (e.g. individuals, firms, countries) are desired. Parameterizing such models requires specifying the different parameters that quantify similarities and differences across different objects, a goal often dependent on using robust and replicable parameter estimation procedures. In fact, in light of rapid growth and dissemination of improved parameter estimation methods for model calibration, hypothesis testing, and policy recommendation in social and behavioral disciplines, continued relevance of any modeling sub-discipline may partially be tied to its ability to remain up-to-date with the best available tools in this domain.

When formal parameter estimation procedures are used, modelers typically compare time series data against the same variables in a model, and minimize the weighted sum of a function of the error term by changing the uncertain parameters until best fitting estimates are found through a nonlinear optimization algorithm (Oliva 2003). The error function is frequently defined as the squared error but absolute error and absolute percent error terms are also common (Sterman 2000). Weights for different data points are usually given based on the confidence the researcher has in the accuracy of the data and its relevance to the problem at hand. When reported, confidence intervals are calculated using normality and independence assumption for error terms which, with weights proportional to the reciprocal of error variance, would turn least squared error estimates into maximum-likelihood estimates (MLE). Bootstrapping methods are also sometimes used for estimating confidence intervals (Dogan 2007). While these approaches cover many important estimation challenges, they each include some shortcomings. Ad hoc selection of the error term and the weights for different data points reduces the consistency of the methods and their ability to provide confidence intervals. Normality and independence may regularly be violated which negate the benefits of MLE when using squared errors. Bootstrapping, while flexible, increases the computational costs significantly and as a result may prove infeasible for many realistically-sized problems. Finally, all these methods rely on having time series data, and cannot extract from distributions in cross-sectional data the dynamics that have led to those distributions.

In this paper we offer an introduction to the Method of Simulated Moments (MSM) for application to dynamic modeling problems. The basic idea of this method is to define appropriate moments of data and, by changing uncertain parameters, minimize the difference between those moments and their simulated counterpart coming from the model. Moments could be any function of data points available. However for analytical confidence intervals to be available, one needs these moments to be normally distributed, often meaning that each moment is an average across a function of multiple independent observations coming from the same underlying distribution (then normality follows from the central limit theorem). In practice those observations (that feed into the moments calculations) are picked either from time series data when a system is in steady state (e.g. stock prices over time), or at similar points in the life of similar units of observations (e.g. all 5-year old individuals in a country).

From Method of Moments to Method of Simulated Moments (MSM)

The MSM is an offspring of a classical estimation method in statistics, the Method of Moments (MM). The MM is based upon finding unknown parameters of a certain distribution by relating these parameters to the moments of the distribution and then using empirical moments (obtained from data) to back up the unknown distribution parameters. While for certain probability distributions the MM can be used to recover parameter values through analytical expressions, it faces two major challenges:

- We need to know the true functional form of the distribution of outcomes.
- We should be able to express the parameters of the distribution in terms of the data moments, a task only feasible for a small set of probability distributions.

For many distributions we cannot find an analytical (close-form) solution to relate moments to parameters. Structural models in general and systems dynamics models in particular usually do not have an analytical solution to relate the output of the model to its structural parameters.

Mcfadden (1989) was the first who proposed using simulation instead of trying to solve the moment conditions analytically. His paper was focused on discrete-response models (multi-nomial Probit) however he provided theoretical foundations for more general models. Mcfadden (1989) believed that an unbiased simulator is used in the MSM where the simulation errors are independent across observations, and the variance introduced by simulation will be controlled by the law of large numbers operating across observations. Lee and Ingram (1991) and Duffie and Singleton (1993) extended the framework and provided a rigorous treatment of the MSM estimators for time-series and panel-data cases and provided relevant statistics for making tests. Since then the MSM has been widely used in various sub-fields of economics such as finance (both asset-pricing and corporate finance), macroeconomics, Industrial Organization (IO), international trade and labor economics.

An example from the dynamic modeling literature

While the MSM has become a major econometrics tool for the past two decades, it has been rarely applied in the system dynamics literature. Barlas (2006), in the design of Behavior Pattern Testing (BTS II) approach and software, uses some of the basic ideas of the MSM, to match moments of model against data, but does not draw on the MSM literature or discuss issues related to confidence levels. Rahmandad and Sabounchi (2011) adapt the MSM to estimating the parameters of an individual weight gain and loss model. In this section we provide a brief overview of this application to provide a more concrete example of MSM use. A simple model of individual's body mass, consisting of fat mass and fat free mass, is developed. The model included a few uncertain parameters. In absence of time series data, those parameters were estimated from cross sectional data on individual weights coming from the National Health and Nutrition Examination Survey (NHANES). NHANES 2005-2006 population of 5,971 subjects was categorized into 110 subpopulations based on different ethnicities (5 ethnicities), genders (2 genders) and age (11 age groups). For each population group two moments, average body weight and variance of body weight, was calculated as the moments to be matched, leading to a total of 220 moments to match.

On the other hand, the model was replicated (using subscripts in Vensim software) for 5971 instances that matched the demographic characteristics (Age, Gender, Ethnicity) of the NHANES sample in year 2006. Initial body weight and fat fraction for these individuals was drawn from distributions of another NHANES sample in 1999-2000. Note that each round of NHANES uses a sample different from other rounds, thus we cannot track the same individual over time and the data is cross-sectional. The model was then simulated to grow this synthetic population from their initial age in year 2000 to their final age (consistent with NHANES sample) in year 2006. Mean and variance of weight for different subpopulations in the simulated population was calculated in year 2006, and compared against the 220 moments coming from the data. Weighted sum of squared errors was calculated using weights of reciprocal of variance in each moment, itself calculated using variance and kurtosis of different moments. This error was minimized by changing 17 uncertain parameters using the Vensim internal optimization engine. The estimated parameters provided the minimum error. As a result, the authors were able to estimate a dynamic model, including individual growth mechanisms, from cross sectional data with individuals in different age groups.

While this application follows the basic ideas of the MSM, it has some differences from the canonical MSM procedure. First, in this application the number of moments (220) is larger than many typical applications, in which the number of moments and parameters to be estimated are in the same order of magnitude. Second, given the computational costs in this setting, each moment was only simulated once whereas typically multiple simulations, using different noise seeds, shall provide the estimation for the moment, before it is compared with data.

Basics of the MSM

Suppose you have built a model which captures the dynamics of people's body weight as a function of their initial weight, eating and physical exercise habits, genetics, age, gender and other fixed and time-varying characteristics. People differ both in terms of their idiosyncratic characteristics (genetics, initial weight, etc) and their environmental factors (e.g. quality of food, cost to exercise, social eating habits, etc). By changing initial conditions and model parameters one will get different dynamic paths for the agent's (individual's) weight as a function of her age.

Suppose we have data on the weight of several children of age 10 (our initial value) as well as at ages 11 and 12. Further assume that we are interested in estimating a structural parameter (e.g. average weight growth per year) which determines the weight path as a function of initial weight.

By fixing this (unknown) parameter to an initial value and simulating the model with all empirical values for the initial weight (age 10), we will generate different paths of weight-age for a simulated population the same size as the number of subjects in our dataset. Now we can compare the distribution of model-predicted weight profiles at ages 11 amd 12 against the empirical distributions. Specifically we can compare the mean and variance of weight for simulated population at ages 11 and 12 against the mean and variance at the same ages observed in the data. It is likely that our initial choice for the structural parameter leads to mean and variance weights different from those observed in the data. However these simulated moments are a function of the parameter. By changing the structural parameter of the model we will change both the mean and the variance of simulated weight values. We can therefore use an optimization method to search for the parameter value that minimizes the difference between model generated mean and variance and their empirical values over all available moments (i.e. mean and variance of age at ages 11 and 12). This is the core idea behind the method of simulated moments: we simulate the moments of the model to find simulated counterparts for observed data, then change the structural parameters until the simulated moments match the observations as closely as possible.

Formal Definitions

Consider a fully-specified model, i.e. a model that can be simulated given a set of parameter values. Assume that there are d unknown parameters which we are interested in estimating. Let's assume that our empirical data $\{x_t\}$ are observed for T different agents. There are P moments functions (sometimes called descriptive statistics) that are available in the data for each agent and $m_i(t), i \in \{1,...,P\}$ is the ith moment of the data for agent t. The i^{th} element of vector of data moments $\tilde{M}_i$ is defined by $\tilde{M}_i = \frac{1}{T}\sum_{t=1}^{t=T} m_i(t)$.

Notice that since we only have access to a sample of data for estimating moments, the true moments of population from which the data sample is collected are approximated by empirical moments $\tilde{M}_D$. The true functional form of the system's dynamics which lead to output $g(.)$ is approximated by the model's output $\hat{g}(.)$. The output of the model is a function of known parameters vector Z, unknown parameters vector θ (to be estimated) and random inputs u. Choosing different values for u will generate different values for g. We assume that the model is correctly specified so that $\hat{g}(.)$ is an unbiased estimator of $g(.)$: $\mathrm{E}(\hat{g}(Z,\theta,u)) = g(Z,\theta)$

This ensures that if we generate a large enough sample of outputs using a true random stream of inputs u_t the arithmetic average of the model output should generate a reasonable approximation of the real-world processes that generate the observations.

$$m_i(\hat{g}(.)) \approx \frac{1}{N}\sum_{j=1}^{N}\frac{1}{T}\sum_{t=1}^{T} m_i(\hat{g}(Z,\theta,u_t)) \tag{1}$$

Equation 1 can be understood as following. The component $\sum_{t=1}^{T} m_i(\hat{g}(Z,\theta,u_t))$ represents the fact that moments may need a number of observations to be calculated. For example, think of variance as a popular moment. The standard formula of variance $\mathsf{E}(X-\mathsf{E}(X))^2$ suggests that we should calculate the empirical average of $(X_t-\mathsf{E}(X))^2$ for all observations $t \in 1,\ldots,T$. As a convention the notation of $\sum_{t=1}^{T} m_i(\hat{g}(Z,\theta,u_t))$ represents all types of averaging operations which are required for calculating different moments. A necessary (but not sufficient) condition for being able to identify the model is to have more moment conditions than unknown parameters.

The core of the MSM is to minimize the (weighted) difference between the empirical and simulated moments by changing the unknown parameters. The estimated parameter set is the value of parameters that minimizes this difference. Specifically, with vector of simulated moments M_S consisting of $m_i(\hat{g}(.))$ elements and the $< P \times P >$ matrix W for weighting the moment conditions:

$$\theta^* = \operatorname{argmin}(M_S - M_D)W(M_S - M_D) \tag{2}$$

If an estimate moment is very sensitive to the random input path, it will generate possibly diverse values across different rounds. On the other hand, those moments that are more robust against the choice of the sample will show smaller dispersion. Using the inverse of variance for matrix W helps us give more weight to more robust moments and reduce the importance of those that change a lot from one round to another.

Choose the Moment Conditions

Usually the first and second moments of model's outcomes (mean and variance) are good candidates to use. Remember that the number of moment conditions should be (equal to or) larger than the number of unknown parameters. Thus, depending on the number of parameters you should decide to use informative moment conditions. Figure 1 shows examples of informative and non–informative moments.

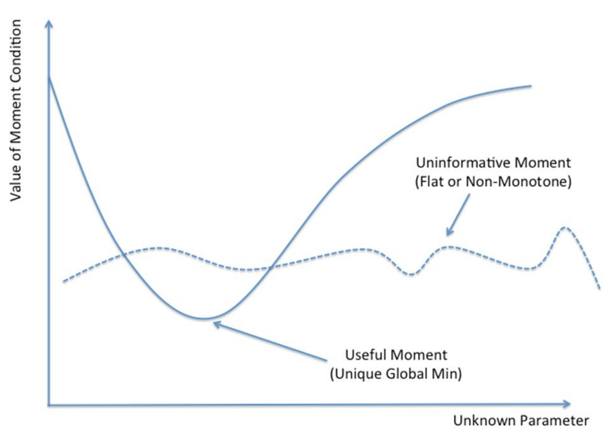

Figure 1: Informative and non–informative moments

The moment specified by the solid line moves smoothly as the unknown parameter changes and has a unique well–defined extreme point (minimum in this). Therefore, minimizing the distance between this function and the empirical moment will generate a unique parameter value. On the other hand, the moment represented by the dashed–line is not informative. It is not very sensitive to changes in parameter value. We can not even be sure that these small changes are due to true response of the model to various parameter value or are the artifact of computational or sampling errors (though if the graph is smooth these conjectures will be less valid).

Optimization Routine and Iteration

This is the most computationally-challenging step of the MSM procedure. We need to minimize the weighted distance of model-generated moments from empirical moments. More formally:

$$\theta^* = \operatorname{argmin}(M_m - M_D)W(M_m - M_D) \tag{3}$$

We need to use numerical optimization routines to find the minimum of the total error function. A smart choice of initial values for parameters may facilitate the quicker convergence of the optimization routine significantly. Any numerical optimization method requires a tolerance rule to stop. This will be given as the error tolerance for the objective function $\| Q^i - Q^{i-1} \| \quad \varepsilon_Q$ as well as for the parameters $\| \tilde{\theta}^i - \tilde{\theta}^{i-1} \| \quad \varepsilon_\theta$.

Similar to any non-linear optimization routine, the MSM estimator may fall into the trap of a local maximum. Moreover, if some of moment conditions are not very informative, hence have low sensitivity to parameter values the problem may face a flat value function which makes it very difficult to progress and converge. To avoid introducing sampling error into rounds of simulation we should work with the same random sample of shocks in each period to make sure that changes in results are due to changes in structural parameters and not the random sample.

Robustness Checks

The MSM uses numerical methods to find the minimum of objective function. Therefore, the results might be sensitive to initial values, the precision of the search algorithm (the level of error tolerance), and the quality of algorithm to distinguish local and global extreme points. We recommend to re-run optimization using distant initial values to check if the results are sensitive to the choice of initial value.

Conclusion

over the last three decades the research in system dynamics has largely focused on diverse applications of the original toolbox, with limited methodological expansions in parameter estimation domain. While formal parameter estimation may not be feasible for many modeling problems, expert dynamic modelers should be equipped with the relevant tools when numerical data is available, model purpose requires reliable parameter estimates, or the audience requires formally estimated parameters.

Given that most dynamic models do not follow a fixed structural form (e.g. linearity), estimation procedures that are independent of model structure are most beneficial. Moreover, independence and distributional assumptions on error terms for dynamic models are not always easy to justify, so methods such as the MSM with fewer such assumptions are preferred. The MSM is especially useful When error terms do not follow any well-established distribution. It could be a good choice when models include stochastic processes that drive the model, and their impact on the model behavior is reflected in the data against which the model is to be calibrated, e.g. when we are trying to match the variance observed across multiple units.

The MSM is also applicable to diverse data types, including both time series and cross-sectional data. It may be the only viable choice for estimating dynamic models when data is cross-sectional as it allows us to extract the information about the historical trajectories of units hidden in their cross-sectional distributions.

References

Barlas, Y. (2006). Model validity and testing in system dynamics: Two specific tools. 24th International Conference of the System Dynamics Society, Nijmegen, The Netherlands, The System Dynamics Society.

Dogan, G. (2007). "Bootstrapping for confidence interval estimation and hypothesis testing for parameters of system dynamics models." System Dynamics Review **23**(4): 415-436.

Duffie, D. and K. J. Singleton (1993). "Simulated Moments Estimation of Markov-Models of Asset Prices." Econometrica **61**(4): 929-952.

Lee, B.-S. and B. F. Ingram (1991). "Simulation estimation of time-series models." Journal of Econometrics **47**(2–3): 197-205.

Mcfadden, D. (1989). "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical-Integration." Econometrica **57**(5): 995-1026.

Oliva, R. (2003). "Model calibration as a testing strategy for system dynamics models." European Journal of Operational Research **151**(3): 552-568.

Rahmandad, H. and N. Sabounchi (2011). Building and estimating a dynamic model of weight gain and loss for individuals and populations. the 29th International Conference of the System Dynamics Society. Washington, DC.

Repenning, N. P. (2003). "Selling system dynamics to (other) social scientists." System Dynamics Review **19**(4): 303-327.

Sterman, J. (2000). Business dynamics: systems thinking and modeling for a complex world. Boston, McGraw-Hill/Irwin.

Sterman, J. D. (2006). "Learning from evidence in a complex world." American Journal of Public Health **96**(3): 505-514.