

# Semiparametrics Done Wrong

*null*

2018-12-29

Colleagues, friends, acquaintances, television personalities, and random people on the street often ask me “What is your deal with simulated method of moments?” So, instead of just making a face like I’ve just drunk spoiled milk when I get to the end of a promising abstract and read the phrase “We estimate the model using the simulated method of moments” like I usually do, I figure it would be worthwhile to lay out some more coherent thoughts, with some simple strawman examples, of why I find most analyses using the simulated method of moments (along with related methods like indirect inference<sup>1</sup>) to be unreliable tools for learning anything useful about the economy.

To a large extent, the problems with SMM are *not* with its statistical properties. For the most part, it does what it says on the tin (err, let’s say in Gourieroux and Monfort’s monograph): given a correctly specified parametric model from which you can simulate, and some set of sample averages from the data, you can recover the parameters with reasonable asymptotic precision, without going through a lot of computationally intensive or intractable likelihood evaluations, which can be especially nasty if you have a model for which each likelihood evaluation requires operations like high dimensional integration (as in nonlinear state space models and other latent variable models) or optimization (as in economic models of the behavior of optimizing agents). Simulation error adds some factor (inversely proportional to the number of simulated data sets) to the standard errors relative to usual generalized method of moments estimation, which itself may have larger standard errors than MLE, but these can be measured and quantified, and even reduced down to the efficient level with enough simulations and proper choice of moments. In graduate school, for my Qualifier Exam in Econometrics, I chose for my topic simulation estimation (see the notes I prepared here) so this was some of the first technical econometric literature I had exposure to, and for 1980’s-1990’s style work, most of it holds up fine.<sup>2</sup>

Beyond these properties, SMM has another “desirable” property: you don’t need to use every conceivable moment from your data to estimate your model, so long as you have a set of moments sufficient to identify the parameters. This is closely related to the ability of the generalized method of moments to perform semiparametric inference. Given a set of moments implied by a model  $E[g(z, \theta)] = 0$ ,  $g(\cdot, \cdot) : \mathcal{Z} \times \Theta \rightarrow \mathbb{R}^K$ , you can estimate the parameter value  $\theta$  which satisfies these conditions. Equivalently, one is implicitly defining a set of probability distributions  $\mathcal{F}(\theta) := f(z) : \int g(z, \theta) f(z) dz = 0$  where the set need not be single-valued, and may even be infinite-dimensional in some cases, in which case the model is called “semiparametric.” As an example, if  $g(z, \theta) = z - \theta$ , then  $\mathcal{F}(\theta)$  is the set of all measures with mean  $\theta$ . One can always recover the parameters (which are said to be “identified”) so long as  $\{\mathcal{F}(\theta) : \theta \in \Theta\}$  form a partition of a set of measures  $\mathcal{F} := \cup_{\theta \in \Theta} \mathcal{F}(\theta)$  over data  $z \in \mathcal{Z}$ , that is, if different parameters correspond to distributions with different values of the moments. Researchers outside of econometrics usually have a preference for defining a set of measures explicitly rather than implicitly, but the modeling principles are the same as any other way of defining a likelihood. Among other things, this means that one can do things like test the model by testing whether the distribution of the data is in  $\mathcal{F}$ , or putting a prior distribution over this set, which under the identifiability condition may be done by having a prior over  $\Theta$  augmented with a prior over the fiber bundle over  $\mathcal{F}(\theta)$ <sup>3</sup>

---

<sup>1</sup>My critiques also hold for primitive ancestors of these techniques like calibration. I want to emphasize that my criticism has little to do with the well-trodden criticisms of calibration on statistical grounds, and encompasses features which are shared by methods which provide proper frequentist uncertainty quantification. Heckman and Hansen’s discussion of calibration covers all of these classes of criticisms, and you should probably read them rather than me.

<sup>2</sup>Both tastes and standards of rigor have changed over time, as the project of trying to shoehorn every conceivable data analysis procedure into a Taylor-expansions-of-sample-averages framework has reached its limits, but the approximations mostly work well in practice even when not all the i’s are dotted and t’s crossed the way one would expect now. Attribute this to a predilection among econometric theorists for testing results by simulation, I suppose.

<sup>3</sup>This is the idea behind Bayesian GMM and related methods, along with a wide variety of Bayesian semiparametric models (R package) for particular applications

The problem with SMM is that it is frequently used as a way to use only a smaller set of moments for estimation, but almost never used for genuine semiparametric inference, to define a model which identifies a parameter which may correspond to a large set of possible probability distributions. Instead of starting with a moment and finding the set of probability distributions that correspond to it, one starts with a set of probability distributions (given by a simulated model) and finds some moments  $M(\theta) = \int m(z)f(z, \theta)dz$  which it generates, which then produce a set of moment functions  $g(z, \theta) = m(z) - M(\theta)$  which themselves generate a set of probability distributions  $\mathcal{F}_m(\theta) := f(z) : \int g(z, \theta)f(z)dz = 0$ .

By construction, it must be the case that  $f(z, \theta) \in \mathcal{F}_m(\theta)$ . However, it is often the case that  $\mathcal{F}_m := \cup_{\theta \in \Theta} \mathcal{F}_m(\theta)$  is a strict superset of  $\mathcal{F} = \{f(., \theta) \theta \in \Theta\}$ , the model from which one is simulating. In words, the set of probability models which satisfy the moment conditions can be much larger than the set of distributions compatible with the model. For example, with the same number of moments as parameters (and a rank condition),  $\mathcal{F}_m$  is all possible distributions, even if the simulation model is highly restrictive. In principle this can be fine, if one is explicitly thinking about SMM as a semiparametric method in which  $\theta$  corresponds to the set  $\mathcal{F}_m(\theta)$  rather than the measure  $f(., \theta)$ . Except when this set is a singleton, it seems to rarely have structure of interest, like decomposability into a parameter of interest and a nuisance parameter, as in more explicit semiparametric constructions.<sup>4</sup> In these cases, the use of a limited set of moments serves only to reduce or eliminate the power to detect model misspecification and to confuse the set of models  $\mathcal{F}_m(\theta)$  with the particular element  $f(., \theta)$ .

This is particularly attractive if you have a model which is misspecified and which you have no interest in changing or if you are not interested in distinguishing between  $f(., \theta)$  and a set of models which may not even be observationally equivalent, but whose differences cannot be detected using the moments chosen. There are scientifically legitimate reasons to want each of these. The former rests on desire to provide “reasonable” simplifications to ensure tractability, interpretability, and all of these other things people speak of as benefits of parsimony, whatever that means. In my experience, pressing this point usually results in the quoting of George Box. The latter is of course the point of semiparametric inference, to the extent that the equivalence class of models is equivalent in a way meaningful to the application. These can also be used, however, as underpowered studies with weak construct validity in other fields might be,<sup>5</sup> to construct sequences of studies all “supporting” some claimed class of proposition, building up a large literature containing negligible content.

To give a concrete example of how this might play out, let’s introduce a simple economic model, and think about how different estimation methods might result in more or less reliable results. To build a straw man, I’ll pick a model I will call “Supply and Demand.”

Suppose we have data on  $N$  spatially or temporally separated markets  $i$ , each iid, in which a quantity  $Q_i$  of a homogeneous good is sold by producers to consumers at price  $P_i$ . The causal response of consumers in market  $i$  to an intervention setting the price to  $P_i$ , is parameterized by a linear function, which I will call a “demand curve”

$$Q^i(P) = a_0 + a_1 P$$

and similarly, the causal response of producers in market  $i$  to purchases of a quantity  $Q$  by consumers will be to adjust the price to a level set by a linear function which I will call the “supply curve”

$$P^i(Q) = b_0 + b_1 Q$$

Completing the model, let’s make the controversial assumption that  $P_i = P^i(Q_i)$  and  $Q_i = Q^i(P_i)$ .

The resulting distribution, which we can simulate from or characterize in closed form in this case, is

$$(P_i, Q_i) = \left( \frac{b_0 + a_0 b_1}{1 - a_1 b_1}, \frac{a_0 + a_1 b_0}{1 - a_1 b_1} \right) \forall i$$

<sup>4</sup>See Dridi, Guay, and Renault (2007) for explicit conditions under which one can treat such a model as semiparametric. To some extent, I see Nakamura and Steinsson’s advocacy for the use of “identified moments” as a suggestion to choose  $m$  in such a way that the set of distributions in  $\mathcal{F}_m(\theta)$  has common scientific or policy implications. While this is a laudable goal, guaranteeing this requires some explicit thought into the structure of  $\mathcal{F}_m(\theta)$ . To achieve this, their suggestion involves using “credibility revolution” style reduced form semiparametric estimates as moments, so that at least a subset of the implications of the model hold under the same conditions that the reduced form estimates do.

<sup>5</sup>Cf Shalizi’s “Neutral Model of Inquiry”, also here and here

Statisticians will notice several problems with this model in practice.