Review of the Article: Estimation of Dynamic Discrete Choice Models by Maximum Likelihood and the Simulated Method of Moments by Philipp Eisenhauer, James J. Heckman, and Stefano Mosso, 2015

By Bahar Zafer

This paper summarizes the method used in Eisenhauer et al. (2015) to compare the performance of maximum likelihood (ML), and Simulated Method of Moments (SMM) estimations and points out their key findings. It focuses on the parts covering the numerical approach and the implementation of SMM by assuming that the details of the structural model of the dynamic discrete educational choice for which the comparison is made is known.

Comparison of ML and SMM in Eisenhauer et al. (2015) follows clearly specified steps. Authors first construct a dynamic structural model of binary educational choice and then, estimate the parameters using real data and ML. In the following steps, these estimates are assumed to be the *true values* of parameters as they are based on real data. Hence, the real data is treated as the population. Then, they use these estimates to simulate synthetic data. Using the synthetic data as sample, they estimate the structural model with two methods, ML and SMM. The comparison is made over the performance of ML and SMM to recover the *true parameter values.*

To estimate the model with ML and SMM, the authors apply Monte Carlo study using same equations and parameter distributions for both estimation methods. They use BFGS algorithm to maximize the likelihood when estimating ML model.

ML and SMM methods are compared across two main lines. The first one is the fit of estimates of the aggregate statistics (the average earnings at each state and state frequencies). This part concerns the parameters that reflect the static aspects of the model. Table 4 and 5 compares true parameter values, and ML, and SMM estimates for unconditional and conditional expectations of average earnings and state frequencies. ML is proven to be a consistent estimation in both Table 4 and 5. SMM performs almost equally well when estimation is unconditional. For the group of late college graduation, which has the smallest number of observations, the performance of SMM is worse than ML when the estimation of parameters is conditional on number of children (Table 5).

The second comparison concerns the dynamic aspects of the model. The authors compare the performance of ML and SMM to recover the true values of parameters that capture returns associated with transitioning from one state to another. This comparison also entails the methods' ability to distinguish the deterministic and stochastic components of the model as the stochastic part of cost ($\eta(\hat{s}', s)$, known to agents when they are giving their decision, but unobserved by the researcher) influences the decision making through the continuation value. Since the unobserved cost of transitioning from one state to the other ($U_C(\hat{s}', s)$), is known by the agents and the synthetic data is simulated from the true population values, a good estimation method (which is able to capture the true parameter values) must perform well when estimating standard deviations of random cost shocks at each state ($\hat{\sigma}_{\eta(\hat{s}', s)}$). Table 7 shows that SMM largely overestimates standard deviation of random cost shocks while ML provides accurate estimates. It is concluded that SMM is unable to recover the unobserved characteristics of agents (latent factors such as ability, θ) while agents' choices (hence, data) reflect them. As a result, while consistent estimation of moments with SMM is supposed to drive the difference, $\check{f} - \hat{f}(\psi)$, down towards zero by finding moment estimates close to real values of moments for a set of chosen parameters, ψ, random perturbations to the parameters in any direction increase this difference lowering the performance of SMM (Figure 6).

The article continues with a discussion of the choices that affect the performance of SMM estimation. They list what a researcher must decide before implementing SMM: (1) Which moments ($f(\psi)$), to use to estimate the vector of parameters, (2) the number of replications (R), (3) which weighting function (W) to use in criterion function ($\Lambda(\psi *)$), and (4) which optimization algorithm to use.

SMM targets to minimize the criterion function, $\Lambda(\psi)$, which is the weighted distance between the observed, $\check{f}$, and simulated moments, $\hat{f}(\psi)$, which are the functions of selected parameters, ψ. When choosing the moments, authors use an auxiliary model and implement the efficient method of moments (EMM) through a Wald approach instead of estimating the auxiliary model directly. This method is used to minimize the distance between the moments on the simulated and the observed data ($\check{f} - \hat{f}(\psi)$). The other component in the criterion function, $\Lambda(\psi)$, is the weighting function, W. Consistency and asymptotic normality of SMM depend on that auxiliary model, parameters, ψ, and weighting function are correctly chosen.

When choosing W, the authors seem to be concerned about smoothing the surface of $\Lambda(\psi)$ as they decide on the matrix of inverse variances of on the diagonal (see the equation at page 347). Nevertheless, comparing the matrix of inverse variances to the identity matrix (page 353, Figure 9) reveals that the choice of W cannot fix the problem discussed above: SMM estimators' inability to distinguish between the systematic and unsystematic cost components (Figure 6). The distance, $\check{f} - \hat{f}(\psi)$, increases with random perturbations to the structural parameters regardless of the choice of W. On the other hand, authors do not discuss how they choose the weighting function in detail and what might be the methods to choose the optimum W.

The choice of number of replications is relatively more straightforward. The idea of using many simulations is to reduce the effect of stochastic components on educational choices and states in the model by averaging out the errors. Hence, with many simulations the distance $\check{f} - \hat{f}(\psi)$ must ideally approach to zero. The choice of replications, R, seems to be model specific, as authors plot the changes in $\Lambda(\psi *)$ specific to the model and choose the lowest number of R where gains for model fit and accuracy from further increasing the number of replications is very minimal.

Finally, the choice of optimization algorithm is discussed comparing the performance of POUNDerS to the standard Nelder-Mead algorithm. Both algorithms can be used when there are discontinuities in the criterion function, $\Lambda(\psi)$, because they are derivative-free. The difference between them is their search strategy. POUNDerS algorithm minimizes the criterion function drastically faster than Nelder-Mead after 100th minute. (Figure 10). POUNDerS algorithm that utilizes a trust region for approximation is argued to be more efficient than the direct search algorithm of Nelder-Mead.

All things considered; for the given model, the article cannot implement a strategy to improve the performance of SMM to the accuracy level of ML when estimating the dynamic parameters which are used for economic and policy implications. While the model fit statistics suggests that SMM estimates are as good as ML estimates, the noisiness of criterion function as discussed above remains to be an issue irrespective of the choice of weighting function.