Chapter 8
Simulation Methods in Econometrics

Thierry Kamionka

Abstract This chapter is devoted to presenting estimation methods based on simulations in econometrics. The techniques used are based on concepts that were discussed very early on, but the first significant advances emerged as a result of the availability of computing resources. Developments specifically for econometrics appeared mainly with simulated likelihood and simulated moments. Other methods then emerged at the same time as the first ones were being improved.

8.1 Introduction

The use of simulations to approximate in mathematics is not recent. In 1733, Georges Louis Leclerc, Comte de Buffon, proposed the needle problem. The "Mémoire sur le jeu du franc carreau", presented in 1733, was published in 1777 (see Leclerc, 1777). Let us consider a parquet on which are traced parallel grooves regularly separated by L units of length. Let's throw the needles at random, the needles measuring $L/2$ units. The probability that a needle cuts a groove is equal to $1/\pi$. Pierre Simon Laplace, 1749-1827, in the Théorie analytique des probabilités (cf. Laplace, 1886), describes a similar experience. The parquet is replaced by a surface divided into parallel lines, and the needle is replaced by a very narrow cylinder. The idea is then to approximate this probability by the proportion of needles which intersects the grooves to obtain an estimation of π. Pierre Simon Laplace indicates that *"Finally, we could use probability theory to adjust the curves or square their areas"*[1]. If you consider, for example, a disk of radius r, you can then use the estimate of π obtained to estimate the area of the disk as $\hat{A} = \hat{\pi} r^2$.

Thierry Kamionka, CNRS, CREST and Institut Polytechnique de Paris,
5 avenue Henry Le Chatelier, 91120 Palaiseau, France, e-mail: kamionka@ensae.fr

[1] "Enfin on pourrait faire usage du calcul des probabilités, pour rectifier les courbes ou quarrer leurs surfaces", Théorie analytique des probabilités (Laplace, 1886).

A Monte Carlo simulation is an algorithm used to approximate a numerical value based on random samples. Monte Carlo methods are useful to deal with the multidimensional integration problems. In particular, Monte Carlo methods are numerical techniques that allow to approximate integrals using random draws (Robert & Casella, 2005)

The term 'Monte Carlo method' comes from researches that were conduced at Los Alamos National Laboratory during the second word war (1939-1945) for the Manhattan project related to the development of the first atomic bomb. The principle of the use of simulation was proposed by John von Neumann and Stanislaw Ulam to solve the problem of neutron diffusion in fissionable material (Metropolis, 1987). The name itself was proposed by N. Metropolis. Interestingly, Monte Carlo Markov Chain (MCMC) algorithms were developed by the same group of scientists who proposed the Monte Carlo method (see Robert & Casella, 2001).

In Los Alamos, numerical calculations were performed using desktop calculators and punch-card tabulating machines. During the second world war at the University of Pennsylvania in Philadelphia, was elaborated the first electronic computer that could be considered a general-purpose machine (see Metropolis, 1987). The ENIAC (Electronic Numerical Integrator And Computer) began operations in December 1945. The ENIAC is considered the first general-purpose electronic computer.

Until the 1960s, research in econometrics relied on methods that yielded estimators with analytical forms. The estimation methods used were based on maximum likelihood estimation, the linear least squares estimator, and the instrumental variables method. The linear least-squares estimator can be linked to the maximum-likelihood estimator when the normal distribution is used to model the distribution of the error terms in the equations. The distributions used to model for maximum likelihood estimation were chosen so that these distributions allow for analytical forms of the estimator.

In structural econometrics, modeling is based on assumptions about the behavior of economic agents. The structural models used during this period were also influenced by the need to develop models capable of generating estimable features based on available datasets. Random utility models are based on the assumption that when individuals must choose between different alternatives, they select the one that provides them with the greatest utility (Marschak, 1960). Random utility models have been used since this period because if it is assumed that the random terms in the utility functions are iid and follow a Type I extreme value distribution; the probability of choosing a given alternative follows a multinomial logistic distribution. This model has become widely used even though it is characterized by the property of independence from irrelevant alternatives. In other words, the probability of choosing one alternative over another does not depend on the list of other alternatives available to the economic agent. This structural choice model is a prime example of the interaction between economic theory and advances in econometric methods. In fact, advances in the field of simulation-based estimation techniques have made it possible to extend this model to cases where the parameters of choice probabilities are random for the observer, leading to the mixed logit models (Revelt & Train, 1998, McFadden & Train, 2000). Mixed logit models could not have been estimated in the

1960s because there is no analytical form for a Gaussian mixture of a conditional logistic probability. In addition, the number of parameters considered to be random can be very large. Mixed logit models do not rely on the assumption of independence of irrelevant alternatives.

It is also worth noting that, in Bayesian econometrics, until the 1970s, posterior estimates of the parameters of an econometric model were based on assumptions that allowed for the derivation explicit formulas for the posterior expectations of the parameters. This Bayesian approach involves jointly modeling the parameters and the sample. To do this, we selected distributions for the model parameters that we knew would combine with the sampling probability to produce known distributions of the parameters given the sample, or posterior distributions. Thus, the integrals corresponding to the posterior moments of the parameters had an analytical form. In econometrics, Kloek and van Dijk (1978) consider the estimation of the parameters of a system of equations corresponding to a macro model. They use Monte Carlo techniques in order to estimate expectations corresponding to posterior moments.

Starting in the late 1970s, the use of numerical optimization algorithms, such as the Newton-Raphson method, combined with ever-increasing computing power, made it possible to consider nonlinear models. To illustrate, this period corresponds to the development of the use of qualitative variable econometrics, ARCH and GARCH models for modeling time series, and nonlinear models in panel data econometrics. The estimation of parametric models involves numerically optimizing an objective function which is assumed to have an analytical form or that includes integrals with a reasonable order of integration. The estimators used are known as M-estimators (see Gouriéroux & Monfort, 1995b).

8.2 Importance Sampling

This is a cross-cutting technique in the field of simulation-based estimation methods. It is used for the estimation of posterior moments (Geweke, 1989a, Fougère & Kamionka, 1992, 2003), in methods of simulated moments (M. Keane, 1994), as well as for simulated maximum likelihood estimation. An important algorithm in this field is based on this technique (GHK algorithm[2]). This technique is used to address classification error issues associated to the dependent variable (M. Keane & Sauer, 2009, 2010) and to estimate models based on incomplete observation schemes (Kamionka, 1998). This technique is used in the context of sequential Monte Carlo methods, specifically particle filters (see Blevins, 2016).

The practitioner must evaluate an integral. This integral can be viewed as the expected value of a certain function of interest with respect to a probability density function. One possible application is where the function of interest is a conditional probability density function with an unobservable heterogeneity term (see Gouriéroux & Monfort, 1991). In another case, the function of interest may be a parameter of

[2] Geweke-Hajivassiliou-Keane.

interest, in which case the probability density function with respect to which the integration is performed will be an a posteriori distribution.

An integral of a function of interest depending on a variable with respect to the probability density of that variable will be replaced by an empirical mean of the function of interest of the variable (or of the parameter according to) calculated for i.i.d. draws in the distribution characterized by the density function of the variable. This is an unbiased approximation of the integral. According to the strong law of large numbers, this average converges almost surely towards the integral that we are trying to evaluate. If the realizations follow an ergodic stationary process, this average converges almost surely towards the integral. If it is difficult to draw samples from the distribution of the variable under study, importance sampling involves drawing samples from a different distribution. In this case, we weight the function of interest by the ratio of the probability density of the variable to the value of the density function used for the draws.

If the function of interest is a parameter, if the sampling probability is known only up to a multiplicative constant, the posterior expectation of the parameter can be expressed as the ratio of two integrals, each of which is known up to the same multiplicative constant. The integral in the denominator, which represents the integral of the posterior density, and sums in principle to the value of this unknown multiplicative constant. Each of these two integrals can then be estimated by importance sampling (Geweke, 1989a). In this case, we have two integrals to estimate, one in the numerator and one in the denominator. We can then use the importance sampling technique for each of these integrals.

M. Keane and Sauer (2009, 2010) study female labor supply behavior using a dynamic panel Probit model. The dependent variable is binary. The endogenous variable is equal to one if the woman works during the period considered, and to zero otherwise. They distinguish between true choice and reported choice for the endogenous variable. The conditional probability to report a choice given the true choice and the lagged value of reported choice is assumed to have a logistic form. They replace each individual contribution with an empirical average, constructed from simulated labor market histories of the latent variable. Since each history for each individual is generated from the true distribution of latent variables for a fixed value of the parameter vector, each component of the average is divided by the density of that history for that parameter value.

Sauer and Taber (2021) propose an estimation method based on indirect inference and importance sampling adapted to the case where the model incorporates at least one discrete dependent variable.

Geweke (1989a) considers the use of Monte Carlo integration techniques with importance sampling for bayesian estimation of econometric models. Interestingly, he considers the impact of the choice of importance sampling density on the numerical standard error of the estimator of the posterior expectation of a function of the parameters.

8.3 Simulated Maximum Likelihood Estimation

8.3.1 Presentation

Historically, in econometrics, researchers have been led to consider how to estimate or evaluate probabilities that are complex to calculate. They first addressed the problem of estimating such an object, then incorporated this approximation into early empirical studies, and subsequently realized that for large-scale integrals, the methodology presented accuracy challenges, which led several econometricians to propose a solution.

One of the earliest simulation-based estimation methods in econometrics is simulated maximum likelihood. A common problem is estimating a probability for which an analytical expression in terms of the model's parameters cannot be obtained. This probability of an event occurring is an integral that has no closed-form solution. The event under consideration is true if the value of a random variable falls within a certain interval, for example. The probability of the event we wish to calculate can be approximated by conducting a Monte Carlo simulation. In this process, random samples of the variable under study are drawn independently, and the probability of interest is replaced by the proportion of these samples that fall within the target range. In econometrics, this approach was proposed by Lerman and Manski (1981). These two authors illustrate the approach using a discrete-choice model with K ($K \in \boldsymbol{N}^*$) alternatives based on a stochastic utility model. An individual chooses the option whose associated utility is higher than the utilities provided by all other options. If the additive error terms in the utility functions are distributed according to a multivariate normal distribution, the probability of choosing one of the alternatives does not have a closed-form solution. The proposed estimator for the probability of choosing one of the alternatives is the simulated frequency method. This method requires a large number of draws when the total number of alternatives is high.

Pakes (1986) addresses a similar problem of estimating the probability of an event occurring when that probability has no analytical expression. Interestingly, the complexity of the problem studied by Pakes (1986) stems from the fact that he considers structural modeling. He is considering whether to renew a patent each year when the renewal is costly. An economic agent will decide to renew a patent for another year if the expected discounted value of holding the patent for one more year exceeds the cost of renewal. What the economic agent does not know for certain is the income that the patent will generate after a certain number of years have passed. For a given number of years, there is an income threshold such that it is optimal to renew the patent if the income is greater than or equal that threshold. What can be observed is the proportion of patent holders who have stopped paying the annual maintenance fee after a certain number of years. The probability of not renewing a patent after a given number of years depends on the model parameters and the income threshold that characterizes the agent's optimal choice. There is no analytical expression for this probability. Therefore, it is estimated based on assumptions about the Markov process that generates the future revenue from maintaining the patent for another year.

These assumptions are chosen such that the revenue thresholds at each stage of the patent's life have an explicit form. These probabilities are estimated using a method similar to that employed by Lerman and Manski (1981) by implementing a Monte Carlo technique.

The asymptotic properties of the Simulated Maximum Likelihood (SML) estimator were studied by Pakes and Pollard (1989), L.-F. Lee (1992) for discrete response models, Hajivassilou and McFadden (1990) for limited dependent variable models, Gouriéroux and Monfort (1993) for panel data models. One of the notable aspects of the paper by Pakes and Pollard (1989) is that the authors study the asymptotic properties of a class of estimators - notably those of the simulated maximum likelihood estimator - without assuming continuity of the objective function used to characterize this estimator.

Gouriéroux and Monfort (1991) consider models that incorporates unobservable heterogeneity. In these models, the distribution of endogenous variables can be described conditional on observable exogenous characteristics and unobserved heterogeneity. An individual's contribution to the likelihood function is expressed as the integral of the conditional density with respect to the density function of the unobserved heterogeneity. Many econometric models incorporate such heterogeneity: linear and nonlinear panel models, count models, duration models, choice models, and linear models with random coefficients. The simulated likelihood function involves replacing the integrals in each individual contribution with a mean obtained from independent draws from the distribution of the omitted heterogeneity. In this context, the asymptotic properties of the simulated maximum likelihood estimator are given in the context of a i.i.d. sample. An important point is highlighted: the dependence of these asymptotic properties on the choice made regarding the draws of unobserved terms - whether to use samples specific to each individual or identical samples for all individuals. Laroque and Salanié (1993) propose a simulation based estimation method that is related to simulated maximum likelihood. This is known as simulated pseudo-maximum likelihood. This method includes simulated nonlinear least squares. They intend this estimator for dynamic nonlinear models that may include lagged endogenous variables, lagged latent variables, and error terms with an autoregressive structure. As an illustration, they consider a canonical disequilibrium model. One of the advantages of the approach adopted by Laroque and Salanié (1993) is that it does not assume that the observations are independent and identically distributed.

The simulated maximum likelihood estimation method is widely used in econometrics. One area of application in economics where this estimation technique is employed is the analysis of consumer choice. The consumer must choose among a number of alternatives. The consumer's satisfaction with a given alternative is described by a stochastic utility function (see Revelt & Train, 1998). The additive stochastic process is assumed to be iid and distributed according to an extreme value distribution. In this case, the conditional probability that a consumer will choose a particular option follows a logistic distribution. We refer to this as a conditional Logit model when there are only two choice options, and as a multinomial Logit model in the more general case. These conditional choice probabilities depend on a parameter vector that is unknown to the econometric model and varies from one consumer to

another. We assume that this parameter vector is independently distributed among consumers according to a distribution characterized by its parametric probability density function. An individual contribution to the likelihood function is the integral of the conditional probability of selecting one of the alternatives, with respect to the probability density of the model's parameter distribution. Once again, we are faced with the problem of evaluating an integral that may be of high dimension. To estimate the model parameters, Revelt and Train (1998) model each individual's successive choices and, consequently, the associated conditional probabilities. The model parameters are assumed to be normally distributed. The resulting model is known as a mixed Logit model. These two authors have realized an application to study household choices among different household appliances. The asymptotic properties of the simulated maximum likelihood estimator are studied by McFadden and Train (2000) in the context of the mixed multinomial Logit model, assuming that the number of samples used tends to infinity. McFadden and Train (2000) show that the simulated maximum likelihood estimator (SMLE) is asymptotically equivalent to the maximum likelihood estimator (MLE) as the number of draws approaches infinity. Train (2016) considers a mixed logit model in which the mixture distribution is approximated by a discrete logistic distribution. When describing consumer choices, the available information may be more detailed than the choices themselves. Thus, Train and Winston (2007) also take into account how individuals rank the various alternatives. They also consider brand preference, which effectively introduces a temporal correlation between purchasing decisions. Train and Winston (2007) operate within the framework of a stochastic coefficient choice model and consider conditional choice probabilities that take the form of multinomial logistic distributions so that parameter estimation requires the use of the SML estimator. Simulated maximum likelihood has been used in many areas of econometrics. For example, to study individual transitions between labor market states, Gilbert, Kamionka and Lacroix (2011) employ a multi-state, multi-episode model that incorporates sophisticated forms of unobservable heterogeneity, allowing them to correlate the distributions of durations of stay in labor market states.

The simulated maximum likelihood estimation, because it is so widely used and despite having emerged very early in the history of econometrics, aged well.

8.3.2 The GHK Algorithm

In the case of a multinomial Probit model, when the number of components of the endogenous variable is high, one cannot use a simulator based on the frequency of draws belonging to a given quadrant. What we need to find is a method that yields samples from the joint normal distribution, all of which belong to the quadrant in question. This quadrant represents all values of the latent variables that result in a given value of the endogenous variable. The method commonly used is the GHK algorithm (Geweke-Hajivassiliou-Keane), which was originally proposed by Geweke (1989b). The method is based on the Cholesky decomposition of the covariance

matrix. The algorithm involves sequentially drawing each component of the latent variable vector given the values of the preceding components. Each draw is made from a truncated normal distribution over an interval. A draw from the truncated normal distribution over an interval can be obtained using the cumulative distribution function of the standard normal distribution, its inverse function, and i.i.d. uniform draws. Since the samples are drawn from truncated normal distributions rather than from a multivariate normal distribution, an unbiased estimator for each integral must also be provided. This will serve as the estimator of the individual contribution to the likelihood function (cf. Gouriéroux & Monfort, 1995a). The GHK algorithm can also be applied to a situation in which one of the endogenous variables is discrete and ordinal.

L.-F. Lee (1997), for example, considers the estimation of a dynamic discrete choice model in the case of a relatively long panel. Each latent dependent variable depends on exogenous individual characteristics, past realizations of the endogenous variable, and an idiosyncratic error term that follows an ARMA process. The contribution to the likelihood function is an integral whose dimension equals the number of observation periods.

8.3.3 An Extension to the GHK Algorithm

In a multinomial Probit model, each component of the endogenous variable vector is a binary variable. It is common practice in econometrics to view these components as resulting from the discretization of continuous latent variables. In the case of this multinomial Probit model, these latent variables share a common distribution, which is a multivariate normal distribution. However, one may need to model a set of endogenous variables that combines both discrete and continuous variables. To illustrate, consider an endogenous variable that takes the value 0 for some observations and a strictly positive value for the others. Such a variable can be modeled using a Tobit model (Type 1, Amemiya, 1984). Another example is a situation where wages or the logarithm of wages are observed only when a person is employed. This situation can be described by two endogenous variables: the first, which is binary, indicates employment status, and the second, which is either positive or zero, represents wages (this is a Type 2 Tobit model). In this case, when the person is without employment, the log of the salary they would have earned if employed ranges from negative infinity to positive infinity. As with the GHK algorithm, to generate samples from a truncated multinomial distribution, the algorithm involves sequentially sampling the error terms from the latent variable equations. On the other hand, for this variant of the algorithm, when the value of one of the latent variables is observed, the value of the error term in its equation can be precisely determined. An unbiased estimator for each integral can be obtained which involves a probability density function for the components whose latent variable values are observed. This algorithm was proposed by S.-K. Chang (2011) to study the labor supply of married women. S.-K. Chang (2011) considers two complementary models. A conditional Tobit model (Type 1),

based on panel data and including an individual effect, treated as a correlated random effect component and an AR(1) error term. This model is dynamic because it includes the lagged value of the endogenous variable in the latent variable equation. The second model used by S.-K. Chang (2011) is a two-tiered dynamic panel Tobit model In this Tobit model, there are two latent variables. Only the sign of one of these two latent variables is observed. The value of the other latent variable is observed only if the first is positive. Xun and Lubrano (2016) examine the sensitivity of the results obtained by S.-K. Chang (2011) to the model specification. Kamionka (2022) uses panel data to examine the decision to participate in sports, employment, and the logarithm of wages. This is a dynamic model with correlated random effects (CRE) and error terms for the three latent variables, which may be autocorrelated and correlated with one another. The model is estimated using an SML estimator based on an extended GHK algorithm, insofar as the exact wage value can be observed when the individual is employed in a given year. Brunet, Kamionka and Lacroix (2025) use panel data to examine the relationship between home ownership and individual labor market behavior. Since earnings are one of the endogenous variables and the number of observation periods is relatively high, the model is estimated using simulated likelihood maximization based on an extended GHK algorithm. In any case, the assumptions made about the distribution of the error terms of the latent variables in the model under consideration will determine the structure of the covariance matrix associated with the multivariate normal distribution. The error terms may result from a combination of individual effects on panel data and idiosyncratic terms, the latter of which may exhibit dynamics.

8.3.4 Sofware Packages

Let us remark that there exist now many software packages that allow to estimate commonly used models using SML estimation. In Table 8.1, we present some of these software commands. It is worth noting that while there are many software packages available in the field of SML estimation, they primarily focus on the mixed Logit model and the GHK algorithm. For the least frequently used econometric models, one may have to program the estimation.

8.4 Simulated EM Algorithms

8.4.1 Introduction to EM Algorithm

The expectation-maximization (EM) algorithm facilitates the search for the maximum likelihood estimator in the presence of incomplete observations. In the case of an incomplete observation scheme, the observed variable is a function of a latent variable. For at least some realizations of the endogenous variable, the latter is not equal to

Model	Authors	Year	Software	Command
Bivariate RE Probit	Alexander Plum (see Plum, 2016))	2016	Stata	bireprov
Multinomial Logit model with unobserved heterogeneity	Haan and Uhlendorff (see Haan & Uhlendorff, 2006)	2006	Stata	mlogit_sim_d0
Mixed Logit model	A.R. Hole (see Hole, 2007)	2007	Stata	mixlogit
Mixed Logit	K. Train	2006	Matlab	mxlmsl
Mixed logit	StataCorp		Stata	cmmixlogit
Panel-data mixed logit	StataCorp		Stata	cmxtmixlogit
RE Dynamic Probit with autocorrelated errors	Mark Stewart (see Stewart, 2006)	2006	Stata	redpace
Models with random parameters	Mauricio Sarrias (see Sarrias, 2016)	2016	R	Rchoice
Mixed Multinomial Logit	J. Molloy & F. Becker B. Schmid & K.W. Axhausen	2021	R	mixl
Mixed Multinomial Logit	Y. Croissant	2025	R	mlogit
GHK multivariate normal simulator	StataCorp (see Gates, 2006)	2025	Stata	ghk
GHK multivariate normal simulator*	A. Genz et al.	2026	R	mvtnorm
GHK multivariate normal simulator	L. Cappellari and S. Jenkins (see Cappellari & Jenkins, 2003)	2003	Stata	mvprobit
GHK multivariate normal simulator	Samsiddhi Bhattacjarjee	2016	R	tmvnsim

* Log-likelihood for multivariate normal distribution for interval-censored and exact data.

Table 8.1: Software packages to implement SML estimation

the latent variable. One can consider that the latent variable would be the realization of the dependent variable if the observation scheme had been complete. There is a loss of information between the complete observation scheme and the incomplete observation scheme corresponding to the observed sample. This loss of information can affect the parameters that can be identified from the observed sample.

One of the most commonly used models in econometrics is the Probit model. The endogenous variable is dichotomous, taking on values of 0 or 1. It is typically modeled by assuming a continuous latent variable that is assumed to be normally distributed. When the latent variable is positive or zero, the value of the dependent variable in the sample is assumed to be 1. Conversely, when the latent variable is

negative, the value of the endogenous variable in the observed sample is assumed to be 0. The conditional expectation of the latent variable is assumed to be equal to the product of a vector of exogenous explanatory variable realizations and a vector of parameters, plus a residual term that is normally distributed with a mean of zero and a standard deviation equal to a positive scalar parameter. It can be verified that, in such a context, only the ratio of the parameter vector appearing in the conditional expectation of the latent variable to the standard deviation of that same latent variable is identifiable.

A right-censored duration model can also be viewed as an example of an incomplete observation design. When the duration is less than the censoring threshold, the duration is observed: the dependent variable is then equal to the duration, which is also the latent variable. When the duration is greater than or equal to the censoring threshold, the observation is equal to the censoring threshold. In this case, there is a loss of information because the exact duration is unknown, but we know that it is at least equal to the value recorded in the sample.

Another example of an incomplete observation design is one in which the conditional distribution of the endogenous variable depends on an unobservable characteristic, and the distribution of this omitted heterogeneity variable is known up to a parameter vector. In this case, the contribution of an observation to the likelihood function is equal to the integral of the conditional density given the exogenous explanatory variables and the omitted heterogeneity term, with respect to the density function of the omitted variable. In this example, the latent variable can be thought of as the couple consisting of the value of the dependent variable and of the omitted heterogeneity variable.

The EM algorithm was proposed by Dempster, Laird and Rubin (1977). It is an iterative algorithm that can be viewed as a first-order deterministic Markov chain. We start the EM algorithm by specifying an initial value for the parameter vector we want to estimate. Each iteration of the Markov chain consists of two steps: a step involving the calculation of an expectation, and a step involving the maximization of that expectation with respect to the parameter vector being estimated. The expectation step (E-step) consists of calculating the conditional expectation of the log of the contribution to the likelihood in the context of full information. Note that the conditional expectation is calculated with respect to the distribution of the latent variable, given the value of the endogenous variable in the sample and the values of the exogenous explanatory variables, for a parameter vector set to the value obtained in the previous iteration of the deterministic Markov chain. We continue iterating until the resulting parameter vector changes only slightly. The EM algorithm is useful for obtaining the maximum likelihood estimator (MLE) of the model parameters (see Wu, 1983). In particular, the EM algorithm can be used in context such that the expectation to be evaluated in the expectation step do not have an analytical expression. Once again, we are faced with the problem of calculating an integral -one for each observation in the sample at our disposal.

Ruud (1991) provides examples of econometric models that were estimated using an EM algorithm.

8.4.2 Simulated EM Algorithms

When the integral to be calculated at each expectation step (E-step) is difficult to compute-and, in particular, when it does not have an analytical form-it can be replaced by an empirical average obtained from draws obtained from the distribution of the latent variable. These simulations must be performed under the conditional distribution of the latent variable, given the exogenous explanatory variables of the model, the value of the dependent variable available in the sample, and a parameter vector value fixed at the last iteration of the algorithm. Note that the value of the modeled variable is a function of the latent variable, which may result in the loss of some information about the latent variable. We are in a situation where the empirical mean of these draws converges to the expected value, which corresponds to the integral we are trying to evaluate. In an EM-type algorithm, the empirical mean considered is the mean of the logarithms of the observations' contributions to the likelihood function. This is therefore an unbiased simulator because, for any fixed number of draws from the latent variable, the expected value of the empirical mean corresponding to the simulator is equal to the desired integral.

In practice, in many applications, the draws of the latent variable can be obtained using independent uniform random numbers. This is the case when draws from the latent variable can be obtained using the inversion theorem (also known as inverse transform sampling). In this case, when $F(x)$ is the cumulative distribution function of the latent variable and u is a draw from the uniform distribution on the interval $[0,1]$, then the inverse of the cdf calculated at u is a draw[3] from the distribution with cumulative distribution function $F(.)$[4].

To illustrate, in the case of a Probit model -whether it involves cross-sectional data or panel data- the latent variable is distributed according to a conditional distribution that is a truncated normal distribution. In this case, let y^* be the latent variable. Then $y^* = \Phi^{-1}((\Phi(b)-\Phi(a))u+\Phi(a))$ is distributed according to the distribution of the latent variable, where $\Phi(.)$ is the cdf of the standard normal distribution. In this example, $[a,b]$ is the interval to which the latent variable belongs and which is determined by the value of the dependent variable. If the dependent variable is equal to 1, a is equal to the negative of the conditional expectation of the latent variable, and b is equal to $+\infty$. In this case, we use uniform draws that belong to the interval $[\Phi(a),\Phi(b)]$ as the normal distribution is truncated on the interval $[a,b]$.

We can have two different Simulated EM algorithms. If we draw new independent random numbers u in each iteration of the algorithm, the sequence of parameter vectors obtained from the algorithm is an ergodic Markov chain. It is the StEM algorithm (Nielsen, 2000a and Nielsen, 2000b). If we use the same independent

[3] For algorithms for drawing according to probability distributions, see Devroye, 1986.

[4] This property is fundamental to the field of simulation. In particular, when used for the SML estimator, these uniform draws should be fixed when seeking to optimize the logarithm of the simulated likelihood function with respect to the parameter vector. This can be done either by storing the results of these uniform draws or by reusing the uniform draw generator after re-initializing it to the same value (the SEED parameter).

random numbers u for all iterations, the resulting algorithm is the the SimEM algorithm (Nielsen, 2000a).

In the econometric literature, the SimEM model has been used notably by Ruud (1991) and Train (2008). The StEM algorithm has been implemented, for example, by Broniatowski, Celeux and Diebolt (1983), by Celeux and J. (1985) and by Kamionka (2025).

For the StEM algorithm, let us consider as estimator a vector of parameters drawn from the stationary initial distribution of the Markov chain described by the algorithm. Nielsen (2000a) shows that, under regularity assumptions, when the number of observations becomes large, this estimator is consistent and is asymptotically normally distributed. He provides the expression for the asymptotic covariance matrix of the estimator. Interestingly, this expression depends on the variance matrix of the score function under incomplete information-that is, based on the sample-and on the variance matrix of the score function under complete information-that is, when the information from the sample is supplemented by the latent variable. This asymptotic property of the estimator is established for a fixed number of draws from the latent variable. It should therefore be noted that the estimator converges for a fixed number of simulations, but these simulations are specific to each observation in the sample. As the number of draws increases, the asymptotic variance of the estimator converges to that of the maximum likelihood estimator.

Assuming regularity conditions, Nielsen (2000a) shows that the parameter estimator obtained from the SimEM algorithm is consistent and asymptotically normally distributed. This property of asymptotic convergence also holds for a fixed number of draws from the latent variable. Incidentally, as with the StEM algorithm, this number of draws can be set to just 1. The asymptotic variance of the parameter estimator obtained using the SimEM algorithm also converges to that of the maximum likelihood estimator as the number of draws per observation in the sample becomes very large.

In applied econometrics, Arellano, Blundell and Bonhomme (2017) used the estimator derived from the StEM algorithm to estimate the parameters of a quantile-based panel data model. They use this model to study income persistence and how consumption responds to income shocks. In this case, the M-Step consists of several quantile regressions. The asymptotic properties of the estimator obtained in the case of such an M-step are studied by Arellano and Bonhomme (2016). Similarly, Arellano, Blundell, Bonhomme and Ligh (2024) use a similar StEM algorithm to study the heterogeneity of consumption reactions to income shocks.

Users may then wonder which simulated EM algorithm to choose between StEM and SimEM. Nielsen (2000a) compares the properties of the two estimators derived from the two algorithms for a fixed number of draws. He shows that for the same value of the number of draws of the latent variable, the estimator we obtain from the StEM algorithm has a smaller asymptotic variance than the asymptotic variance of the estimator resulting from the SimEM algorithm.

Estimation methods based on simulated or stochastic EM algorithms are grounded in the pioneering work of Dempster et al. (1977), as well as in extensions incorporating simulations conducted during the 1990s and theoretical results obtained in the early

2000s. These techniques are particularly well-suited to models incorporating discrete unobservable heterogeneity. Applications with this characteristic are still numerous. In this respect, these methods aged well.

8.4.3 Sequential Likelihood

This section is particularly suited to models with unobservable features that follow a multinomial distribution. There are many examples of such models in applied econometrics. Kennan and Walker (2011) consider a model with optimal migration. In this model, the main factor driving migration is expected income. To account for unobserved heterogeneity, they use a discrete mixture of conditional models. Cameron and Heckman (1998) consider a dynamic model of schooling attainment. They approximate the distribution of the latent factor using a discrete distribution. Eckstein and Wolpin (1999) consider a model of dynamic discrete choice. They model high school attendance and work decisions considering a discrete distribution of types for young individuals. Heckman and Singer (1984) model duration data with a discrete distribution of unobserved heterogeneity. M. Keane and Wolpin (1997) examine the sequence of choices made by young men and rewards, based on a discrete distribution of person types. Kiefer (1980) assumes that the dependent variable is generated by switching regressions. Information about the regime that generated the observation is unknown. He considers a discrete mixture of linear regression models. Mroz (1999) considers the effect of a binary endogenous variable on a continuous dependent variable. Both the continuous outcome variable and the binary variable depend on an unobservable characteristic with a discrete distribution. Train (2008) uses a mixed logit model with discrete mixing distribution to model consumer's choice among alternative-fueled vehicles. Provencher, Baerenklau and Bishop (2002) assume that every day, a fisherman decides whether to go salmon fishing. They use a multinomial distribution to model the group to which each angler belongs.

When the model incorporates a discrete distribution of unobservable heterogeneity, one contribution to the likelihood function is the mean of the density function of the dependent variable conditional on both observable and unobservable heterogeneity. This average is weighted by the probability that the omitted heterogeneity equals a given value, and the sum is calculated for all possible values of that variable. Two properties can be derived. The first is that the parameter vector we seek to estimate apart from the probabilities associated with the mixture is the solution to a maximization problem involving the mean of the logarithms of the conditional contributions of the dependent variable, weighted by the probabilities that the omitted heterogeneity takes on its various values given the observed value of the dependent variable, the values of the explanatory variables, and for a parameter vector fixed at the maximum likelihood estimator. It is this first property that allows us to implement an EM-type algorithm, in which the maximization step consists of maximizing this average of the logarithms of the conditional densities, by weighting these quantities with the conditional probabilities associated to the omitted heterogeneity calculated

from the parameter vector obtained in the previous step of the algorithm. The second property is that the probabilities associated with the mixture can be updated using the sample-averaged conditional probabilities that the omitted heterogeneity takes on its various values, given the value of the endogenous variable, the explanatory variables and a parameter vector set to its most recently calculated value obtained during the execution of the algorithm.

When the contribution of an observation to unobserved heterogeneity can be factored into a first density that depends on a subset of the parameters and another density that may depend on all the parameters, it is then possible to decompose the maximization step of the EM algorithm into several successive sub-steps. This sequential way to implement of the EM algorithm is called Expectation-Conditional Maximization (ECM), see Meng and Rubin (1993) and Arcidiacono and Jones (2003). In this case, the maximization step can be broken down into a first substep of optimization with respect to a first set of parameters, followed by a second substep of maximization with respect to a second set of parameters, with the first set of parameters being fixed at their most recent values obtained during the execution of the algorithm. This decomposition property is useful when the model has a relatively large number of parameters to be estimated. Arcidiacono and Jones (2003) show, under standard regularity conditions, that the estimator associated with the ECM algorithm is consistent and asymptotically normally distributed (CAN). It should be noted, however, that the estimator obtained in this sequential manner is not as efficient as the maximum likelihood estimator. It should be noted that Bonhomme (2006) provides an estimate of the parameter covariance matrix for the estimator obtained using an ECM algorithm.

This property of factorization of the individual contributions to the likelihood function can be used in a similar way in the context of a simulated EM algorithm.

8.5 Method of Simulated Moments

In econometrics, one often encounters situations where conditions on empirical moments are available to characterize an estimator of a model's parameters. This is particularly true of the linear least-squares estimator. It is also true of the maximum-likelihood estimator, insofar as the likelihood equations characterize the estimator of interest. In this case, the vector of parameters to be determined is the solution to the equation $M(\theta) = E[g(y;\theta)] = 0$, where the expected value is calculated based on the true distribution-the one that generated the data in the observed sample. In other words, there are restrictions on the population moments, and we assume that these conditions hold only for the true values of the parameters. In practice, these conditions regarding population moments are replaced by an empirical average obtained from an observed sample drawn from the population of interest. An estimator of the parameter vector θ is the value of that vector such that the sample moments are as close to zero as possible. This estimator is the one used in the method of moments (MoM). When the function $g(y;\theta)$ is difficult to compute, particularly when its expression

includes a multivariate integral, it is difficult, if not impossible, to precisely compute the sample moments for a given parameter vector.

The case we will discuss in detail is one in which the function g(.), which appears in the sample moments, is an integral of a function $G(y, \nu; \theta)$ of the dependent variable y, a latent variable ν, and the parameter vector θ with respect to the conditional density of the latent variable given the dependent variable. In this case, the function $g(y; \theta)$ can be estimated by drawing the variable ν from its conditional distribution and calculating the empirical mean of the function $G(y, \nu_k; \theta)$ over the resulting draws. Pakes and Pollard (1989) study the properties of the estimator of the simulated moments method (MSM) obtained by choosing the value of the parameter vector that makes the simulated empirical moments as close to zero as possible. The articles by Pakes (1986) and McFadden (1989) are two examples of the application of such a simulation-based estimation method. The article by Pakes (1986) examines the decision to renew a patent or not. McFadden (1989) considers the estimation of the parameter vector of a multinomial probit model. Pakes and Pollard (1989) consider some regularity conditions such that the MSM estimator converges in probability to the value of the vector of parameters unique solution of the asymptotic problem. They also provide conditions for regularity, such as if the MSM estimator is convergent, then its asymptotic distribution is that of the normal distribution, for which they provide the expression for the covariance matrix. Gouriéroux and Monfort (1995a) also consider the conditions under which the MSM estimator is consistent and asymptotically normally distributed.

McFadden and Train (2000) consider a multinomial logit (MNL) model with random coefficient. Each individual must choose one alternative from a set of options. Given the observable characteristics of the individual and the coefficient vector, the conditional probability that the individual will choose one of the alternatives has a logistic form. The coefficient vector is assumed to follow a distribution characterized by its probability density function. The conditional probability, given the observable explanatory variables, that the individual will choose one of the alternatives is equal to the integral of the choice probability conditional on those same exogenous variables and the coefficient vector, with respect to the coefficient density function. McFadden and Train (2000) consider a specification in which the model's coefficient vector is the sum of the parameter vector and the quantity resulting from the product of a factor loading matrix and a vector of factor levels. In this formulation of the model coefficients, only the factor levels are random; the factor loading matrix is deterministic, and its components are themselves model parameters. They then consider that the distribution of factor levels is characterized by its probability density function. In particular, they note that the probabilities that an individual will choose an alternative can be simulated by drawing factor levels from their distribution and replacing the integral that defines these probabilities with an empirical mean.This mean is calculated by substituting the vector of coefficients with the one obtained for each draw of factor levels. The objective function of the method of simulated moments is constructed by considering, for all possible alternatives, the differences between the value of the dependent variable associated with the alternative-which is 1 if the individual chose that option and 0 otherwise-and the probability that the individual

did so. It is this latter probability that is simulated. Each of these differences between the binary variable associated with an option for the individual and the probability that the individual will choose it can be weighted by any instrument vector that has the same dimension as the model's parameter vector. The asymptotic properties of the MSM estimator are studied by McFadden (1989) and by Pakes and Pollard (1989).

The MSM estimator can be useful in many areas of applied econometrics. This is particularly true for estimating economic models based on behavioral assumptions. For example, Duffie and Singleton (1993) propose a Markov model of asset prices. In this model, the authors consider the production of the consumer good. The production function takes into account the capital stock used and a technology shock for the period. A portion of the capital stock depreciates each period and must be replaced. The company pays a rental rate to use the capital and distributes dividends to shareholders. The consumer consumes the goods produced by the firm. The consumer also rents the capital used by the firm. To determine the level of consumption for a given period, the consumer maximizes the expected value of the discounted utility stream over an infinite horizon. The consumer's utility function is of the relative risk aversion (CRRA) type. Consumer utility in each period may be affected by a taste shock. The consumer takes his or her budget constraint into account in each period. Consumer income includes dividends and income from the rental of capital. The portion of income not used for consumption is allocated to the purchase of stocks. Shocks to the production function and consumer preferences follow a Markov process, which depends on the values of these shocks in the previous period and an i.i.d. stochastic process. For each value of the parameter vector and for an i.i.d. sequence of error terms, the model can be solved to determine the optimal level of capital employed in each period, consumption in that period, the dividends the company pays and the firm's market capitalization. The MSM estimator is a parameter value that matches the sample moments associated with the process realizations and the sample moments of the simulated process.

The model proposed by Michener (1984) is very similar to the model proposed by Duffie and Singleton (1993), with the exception that there is no shock to consumer utility. In Michener (1984)'s model, if the error terms are assumed to be independent and identically distributed random variables following a normal distribution, if the depreciation rate is 100%, and if the utility function is logarithmic, then the model has an analytical solution. In other words, for each period, we can derive the expression for the equilibrium asset prices, the level of the capital stock, and the dividend that is distributed.

The data generated by such a Markov process cannot be considered independent. Therefore, Duffie and Singleton (1993) examine the properties of the MSM estimator under the assumption that the process generating the data is stationary and ergodic. They show that, for a fixed number of simulations, the MSM estimator is convergent and asymptotically normal. B.-S. Lee and Ingram (1991) study the asymptotic properties of the MSM estimator under slightly different conditions for a fully specified stochastic equilibrium model. For a consistent estimation of the asymptotic covariance matrix, one can refer to Newey and West (1987).

Boyd, Lankford and Wyckoof (2013) analyze the determinants of the assignment of public school teachers to jobs in the Albany-Schenectady-Troy, Buffalo, Rochester, Syracuse, and Utica-Rome metropolitan areas. They use a game-theoretic two-sided matching model and the MSM estimator to study the factors that influence the match between elementary school teachers and their first jobs.

The simulated moments method emerged in the 1980s. It is both very general—the maximum likelihood estimator is based on moment conditions—and is particularly widely used in the field of demand modeling. For these two reasons, the method aged well.

8.6 Indirect Inference

Some econometric models can be particularly difficult to estimate because the likelihood function does not have an analytical form. Gouriéroux, Monfort and Trognon (1984) studied the asymptotic properties of the maximum likelihood estimator when the model is ill-specified but the distribution under consideration belongs to a linear exponential family. This approach nevertheless requires that we be able to give the form of the conditional expectation of the dependent variable, which is not always possible. This raises the question of how to estimate the parameters of the distribution that generated the data when the distribution used for the estimation is not the true distribution, and the expected value associated with that distribution is not equal to the true conditional expectation. This issue was first addressed by Gallant and Tauchen (1996), who proposed using a simulation-based estimation technique to establish a link between the pseudo-true value-the value toward which the quasi-maximum likelihood estimator converges-and the true value of the parameter being estimated. This approach requires knowledge of the family of distributions to which the distribution generating the data belongs.

The relationship between the true value of the parameter and the pseudo-true parameter is defined by the binding function. Ideally, we would like to have the expression for the inverse of the binding function, but that's not usually possible. The method proposed by Gallant and Tauchen (1996) is called the Efficient Method of Moments (EMM) and is based on the pseudo-score vector. They remark that the true expectation of the vector of pseudo score should be equal to zero for the quasi maximum likelihood estimator. They therefore propose replacing this expectation with an average over a sample of values of the endogenous variable obtained by simulation under the true distribution that generated the data, for a given value of the parameter vector characterizing that distribution. The estimator of the vector of parameters of interest is the value of the parameters in the true distribution that minimizes the empirical mean of the vector of pseudo-scores when calculated using the quasi-maximum likelihood estimator. The value of the quasi-maximum likelihood estimator contains the information derived from the observed sample. The EMM estimator is also an MSM estimator, since it is constructed based on conditions regarding the empirical moments (the expected value of the pseudo-score vector is

zero). This method is not merely a variant of the MSM estimator, as it opens the door to an entire family of estimators. These estimators share the characteristic that they can be used when samples can be generated from the distribution under study, yet they allow for the use of an estimation method on the observed sample that is as easy to implement as ordinary least squares. This class of estimators was subsequently referred to as indirect inference estimators (Gouriéroux, Monfort & Renault, 1993).

The model used for the estimation is the auxiliary model. This model is chosen such that estimating its parameter vector is simpler than estimating the parameter vector of the initial model. The data available in the observed sample are derived from the initial model for the true value of its parameter vector. The common feature of estimators based on indirect inference is that they simulate samples from the initial model for a fixed value of the parameter vector associated with that model. It is generally assumed that the dimension of the parameter vector associated with the auxiliary model is at least equal to that of the parameter vector in the original model. This condition is necessary to ensure that the optimization program from which the estimator by indirect inference is derived has a unique solution. The various estimators based on indirect inference differ in terms of the calibration step, which specifies how the parameter vector of the initial model is incorporated into the objective function. In other words, the type of calibration used is determined by the objective function being optimized. There are currently three categories of calibration in this field. If calibration is performed using the mean of the pseudo-score vectors, we can say that we have score-type calibration. If calibration is based on comparing the quasi-maximum likelihood estimators derived from the auxiliary model, we can say that we have Wald-type calibration. Finally, if the calibration is based on a comparison of the log-pseudo-likelihoods, it can be said that we have likelihood ratio-type calibration. Likelihood-ratio calibration was proposed by Bruins, Duffy, Keane and Smith (2018).

Interestingly, Gouriéroux and Monfort (1995a) show that if the number of parameters in the initial model and in the auxiliary model is the same and if the sample size is sufficiently large, then the estimators obtained by indirect inference from the first two types of calibration-Score and Wald-are identical. These same authors specify the conditions under which the estimators obtained by indirect inference from these two types of calibration are convergent and asymptotically normally distributed for a fixed number of draws (see also, Gouriéroux et al., 1993).

In this context, the binding function relates the values of the parameters that constitute the solution to the asymptotic optimization problem to the values of the parameters in the initial model. In general, this function is not known analytically. In general, the formulation of the asymptotic variance-covariance matrix of the estimator via indirect inference depends on the first derivatives of the binding function, which raises the problem of estimating the asymptotic standard deviations. Gouriéroux and Monfort (1995a) nevertheless provide an expression for this matrix that does not require knowledge of these derivatives.

When implementing an estimator based on indirect inference, a difficulty arises, however, when at least one of the endogenous variables is discrete, as is the case, for example, in a discrete choice model. In these models, the objective function in the

calibration step is not continuous with respect to the model parameters. To illustrate the problem, consider a model with a dichotomous endogenous variable. It takes the value 1 if the latent variable is positive and 0 otherwise. This latent variable is the sum of two terms: the first is an error term distributed according to a normal distribution, and the second is the conditional expectation of the latent variable. This conditional expectation depends linearly on the model parameters. The draws of the error term corresponding to the latent variable are fixed. However, when the values of the parameters are varied, the resulting latent value can shift from the negative to the positive range and vice versa. Then, the draws of the endogenous variables may vary discontinuously depending on the model parameters. This also applies to the objective function during the calibration step.

The main advantage of the indirect inference method is that it is relatively easy to implement. However, for models involving discrete choices, the objective function is not continuous everywhere with respect to the parameters being estimated. This difficulty can nevertheless be overcome by choosing one of the three techniques that have been proposed to address it.[5]

The first technique involves smoothing the objective function with respect to the model parameters. To do this, a smooth function of the latent utility (or latent variable) is selected as the endogenous variable used in the auxiliary model (see Bruins et al., 2018). To do this, when the model of choice involves more than two alternatives, the technique relies on a multivariate mean-zero cumulative distribution function (CDF) with a smoothing parameter. The method of replacing discrete endogenous variables in the auxiliary model with smooth functions of the parameters is called generalized indirect inference (Bruins et al., 2018). They consider three alternative calibration methods for estimating the parameters. They show that, under assumptions of regularity, the resulting estimator asymptotically follows a normal distribution for a fixed number of draws. As the smoothing parameter approaches zero, the objective function approaches that corresponding to the indirect inference approach without smoothing. If the smoothing parameter is too small, the optimization algorithm will have difficulty finding the value of the estimator. To determine the value of the estimator, Bruins et al. (2018) propose an iterative method that gradually reduces the value of the smoothing parameter and increases the number of draws.

Frazier, Oka and Zhu (2019) consider a auxiliary model that imply moment conditions. They consider two alternative calibration methods. The objective function is a function of the error term draws of the latent variables in the discrete choice model. Frazier et al. (2019) propose replacing these draws with new draws using a change in variable technique, such that the derivatives of the objective function exist and are continuous in the parameters. This change in variables requires weighting the empirical moment conditions derived from the auxiliary model. Frazier et al. (2019) show that, under conditions of regularity and for a fixed number of draws,

[5] Let us remark that for the calibration based on the score function, as the indirect inference estimator is related to the MSM estimator, the regularity assumptions proposed by Pakes and Pollard (1989) may be considered and, in such a context, the asymptotic properties do not require the continuity of the objective function.

the estimator based on indirect inference is consistent and has an asymptotic normal distribution.

The third technique involves the use of an importance function. To illustrate the method, let's consider the example of a conditional Probit model. Of course, the parameters of a Probit model can be easily estimated by maximizing the likelihood function, but here we are using this simple example to illustrate its implementation in the context of indirect inference. The observed endogenous variable results from the dichotomization of a latent variable. Sauer and Taber (2021) propose generating samples of the latent variable for a fixed set of model parameters, once and for all. Since each realization of the latent variable is fixed, the corresponding value of the endogenous variable is also fixed and no longer depends on the parameters in the objective function. Next, the indirect inference estimator is applied by substituting all values of the latent variables for each sample with these initially obtained values. Since the latent variables are obtained for a specific value of the parameters, each contribution to the objective function in the auxiliary model must be weighted. The weight function is the ratio of the likelihood function for the estimated parameter vector to the likelihood function for the initial parameter value used to generate the latent variable draws. Sauer and Taber (2021) demonstrate the convergence of the resulting estimator and its asymptotic normality. This technique is fairly intuitive in principle and relatively easy to implement.

For certain auxiliary models, such as a GARCH model, constraints in the form of inequalities on the auxiliary model's parameters must be imposed to ensure that the auxiliary model's estimates behave well. Such a set of restrictions imposed on the parameters of the auxiliary model could affect the statistical properties of the indirect inference estimator. Frazier and Renault (2020) propose a new method of indirect inference that accounts for and corrects for this characteristic of the indirect inference approach, and they provide the asymptotic properties of this new simulation based estimator.

Indirect inference emerged as an estimation method later in the history of econometrics (Gallant & Tauchen, 1996). Implementation difficulties arose when—as is often the case—at least one of the endogenous variables is discrete. These difficulties have now been overcome thanks to various extensions (see Sauer & Taber, 2021). This technique is valued for its ease of implementation and has great potential for application. This technique has aged well in the sense that it has now reached maturity.

8.7 Simulation for Dynamic Structural Models

The econometric literature now contains many examples of estimation problems involving dynamic discrete-choice models. In these structural models, which describe the behavior of economic agents faced with decisions that have future consequences, one is required to evaluate high-dimensional integrals. This dimensionality problem stems not only from the number of possible choices but also to the dimension of the

state space. For a recent review, readers may refer to Aguirregabiria and Mira (2010)[6]. For an overview of several application examples, one can refer to M. P. Keane and Wolpin (2009). In this context, it may be advantageous to use a simulation-based estimation method.

For example, M. P. Keane and Wolpin (1994) present an approximation method for estimating, for each period, each expected maximum of the alternative-specific value function. In this discrete choice dynamic problem, the value functions specific to the alternatives satisfy Bellman's equation, Bellman (1957). The method is based on Monte Carlo integration. This method involves replacing the expected values with empirical averages obtained from a finite number of draws of the model's error term vector obtained from its distribution.

M. P. Keane and Wolpin (1994) use Monte Carlo integration to deal with large scale discrete decision problems (DDP). Rust (1997), considers a broad class of random algorithms, that include the one proposed by M. P. Keane and Wolpin (1994). The Rust's class of algorithms succeed to break the curse of dimensionality. Finally, to sum up, Rust (1997) proposes stochastic algorithms for approximating the solution to markovian decision problems (MDPs), which are sequential decision problems (see also H. Chang, Fu, Hu & Marcus, 2006).

Blevins (2016) proposes a class of methods devoted to the estimation of dynamic structural models with serially correlated latent state variables. Blevins (2016), as an application of the method, considers an extension to the seminal work of Rust (1987). Rust (1987) models the bus engine optimal replacement decision. In the problem considered by Blevins (2016), there are two state variables, the mileage and the latent state of the engine. The latter cannot be observed by the econometrician and may be serially correlated. In the modeling framework used by Blevins (2016), the likelihood function is an integral of the density of the sequence of endogenous realizations conditional on the latent variables, with respect to the density of the distribution of the latent variables. To factorize the likelihood function, he assumes that the joint distribution of the endogenous and latent variables at time t, conditional on the past realizations of these two variables, depends only on the most recent realization of this pair at $t-1$ (this is a first-order Markov process). Finally, the likelihood function is written as a product of integrals, one for each period, which consists of integrating the density of the endogenous variable for that period, conditional on the value of the lagged endogenous variable and the latent variable for the current period, with respect to the density of the current latent variable, given the set of realizations of the endogenous variable from the initial date up to the previous date. Particle filters, or sequential Monte Carlo methods, can be used to draw according to the density of the latent variable for the current period, given the past realizations of the endogenous variable. By approximating the integrals in this way through simulation, we obtain a maximum filtered likelihood estimation (MFLE).

The estimation of dynamic structural models relies on various simulation-based estimation techniques. This field of econometrics is rapidly evolving, as incorporating

[6] See also Rust (1994).

more structure helps to highlight the complexity and richness of the relationships among economic agents.

8.8 Nonparametric Simulated Maximum Likelihood

This simulation-based estimation method is particularly well-suited to a model in which we can simulate draws from the dependent variable but for which we do not have a closed-form expression for the conditional density of the realizations of that variable.

The nonparametric simulated maximum likelihood estimation method was proposed by Fermanian and Salanié (2004). The central idea is to approximate each contribution to the likelihood function using a kernel estimator based on the difference between the observed value and the simulated value of the dependent variable. Simulated values of the dependent variable are obtained for each value of the parameter vector. For each individual contribution, the kernel estimator of the conditional density depends on a smoothing parameter, the bandwidth. The higher this value is, the more smoothing is applied. This parameter should approach 0 as the number of draws used for each individual realization approaches infinity. As small errors can be amplified for small value of the contribution to the likelihood, Fermanian and Salanié (2004) propose a way to trim the smallest values of estimated individual contributions. Under regularity conditions, they show that the proposed parameter estimator is consistent and asymptotically normally distributed as both the number of observations and the number of draws tend to infinity.

Kristensen and Shin (2012) extend the method introduced by Fermanian and Salanié (2004) to dynamic models, providing theoretical results for this specific case. The approach can include non stationary and time-inhomogeneous processes. Here, once again, it is assumed that the conditional density function associated with each realization does not have a closed-form expression and that this density will be simulated. These simulations will be obtained from the observed data by using draws of the endogenous variable and a kernel density estimator. Generally, kernel density estimators suffer from the curse of dimensionality. Notably, Kristensen and Shin (2012) show that the proposed density estimator does not suffer from this 'curse of dimensionality' problem. The variance of the estimator has the usual standard parametric rate of $1/n$, where n is the sample size. To study the asymptotic properties of the estimator they consider that the sample comes from a stationary and ergodic process and some additional regularity conditions. Among these conditions of regularity, some describe how the bandwidth and trimming sequence should behave as the number of realizations and draws approaches infinity. They conclude that the estimator is convergent and asymptotically normally distributed and efficient. Using Monte Carlo experiment, they show that the NPML estimator performs well across a wide range of bandwidth choices.

Tincani (2021) uses Chilean administrative data to estimate a structural model designed to evaluate teacher policies in a context where two school sectors (private

and public) coexist. The model accounts for parental choice among different types of schools, individuals' self-selection into teaching (teacher supply), and across different sectors. The model describes teachers' salaries in private schools. Tincani (2021) considers two steps of estimation of the parameters. The second step in the estimation process involves estimating the cost parameters for the school sector that uses vouchers based on variations in wage rates across different markets. As part of this second step, the estimation is performed using the nonparametric simulated maximum likelihood (NPSML) method.

This estimation method is relatively recent. The use of nonparametric techniques is a key feature of this method, which may have hindered its implementation. And yet, the results presented by Kristensen and Shin (2012) show that, in this context, these estimators do not suffer from the 'curse of dimensionality'. One might therefore conclude that this method has not 'aged so well', but this could simply be because the method is still in the early stages of its 'life cycle'.

8.9 Adversarial Estimation

Kaji, Manresa and Pouilot (2023) propose a simulation based estimation method, the adversarial estimation. This method can be used when data can be simulated from the model under study, and is useful when the likelihood function is not available. This is the case for a semiparametric model or when the conditional density function does not have a closed-form expression. This estimation technique may be preferred over the method of simulated moments (MSM) when the model includes many moment conditions. In fact, in such cases and by analogy, the GMM estimator may exhibit poor properties on a small sample (Altonji & Segal, 1996). The idea is to find the value of the vector of parameters of the structural model such that the simulated data cannot be distinguished from the observed data given the discriminator used. However, the method also involves selecting the discriminator from within a family. Therefore, the adversarial estimator is a solution to a minimax problem. A discriminator is an classification algorithm that separates observed data from simulated data. One select the discriminator that yields the best accuracy (max) and the parameter vector value that makes the observed and simulated data indistinguishable (min).

The asymptotic properties of the adversarial estimator depend of the choice of the family to which the discriminator belongs. If the discriminator is logistic the asymptotic variance of the adversarial estimator is the same as the one of optimally-weighted MSM estimator. Kaji et al. (2023) propose an estimation algorithm in the case that the family of discriminators used is a class of neural networks. In this context, they propose to use the Adam algorithm to train the discriminator (Kingma & Ba, 2015). Under some regularity conditions, if the number of draws grows faster than the number of observations, if the model is correctly specified and the discriminator used is the Oracle discriminator, then the adversarial estimator is convergent, asymptotically normally distributed, and efficient. In other words,

if the family of discriminators used is sufficiently rich, then the estimator will be asymptotically efficient.

Adversarial estimation was developed very recently (see Kaji et al., 2023). Kaji and Manresa (2026) uses this method to study the reasons why older adults save. This method can use a neural network as a discriminator. The method is therefore linked to recent advances in algorithms, which are also bringing about profound changes in many fields. Therefore, the method is likely to aged well.

8.10 Conclusion

Simulation based estimation techniques provide ways to estimate model parameters more easily. This may be the case when the likelihood function involves multiple integrals, or when the parameters are estimated using moment conditions involving multiple integrals, or because there are many moment conditions and the sample size is relatively small, or because the model is based on assumptions about individual behavior such that we do not have the expression for the conditional density of the endogenous variable.

The availability of these simulation-based estimation techniques opens up new possibilities early on, during the modeling phase. Indeed, to ensure realism, it is then possible to incorporate in the model error terms with complex structures, possibly multiple endogenous variables, possibly incorporating some dynamic, possibly with several unobserved heterogeneity terms, possibly measurement errors. Altonji, Smith and Vidangos (2013) use indirect inference to estimate a joint model of earnings, employment, job changes, wage rates, and work hours over a career using panel data. This model was designed to incorporate the characteristics already highlighted in the literature-such as the effects of experience and tenure-while also adding elements to the model that allow for the exploration of new questions, such as the effects of the quality of the employer-employee relationship.

References

Aguirregabiria, V. & Mira, P. (2010). Dynamic Discrete Choice Structural Models: A Survey. *Journal of Econometrics*, *156*, 38–67.

Altonji, J. & Segal, L. (1996). Small-Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business & Economic Statistics*, *14*(3), 353–366.

Altonji, J., Smith, A. & Vidangos, I. (2013). Modeling Earnings Dynamics. *Econometrica*, *81*(4), 1395–1454.

Amemiya, T. (1984). Tobit Models: a Survey. *Journal of Econometrics*, *24*(1-2), 3–61.

Arcidiacono, P. & Jones, J. (2003). Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm. *Econometrica*, *71*(3), 933–946.

Arellano, M., Blundell, R. & Bonhomme, S. (2017). Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework. *Econometrica*, *85*(3), 693–734.

Arellano, M., Blundell, R., Bonhomme, S. & Ligh, J. (2024). Heterogeneity of Consumption Responses to Income Shocks in the Presence of Nonlinear Persistence. *Journal of Econometrics*, *240*(2), 1–45.

Arellano, M. & Bonhomme, S. (2016). Nonlinear Panel Data Estimation via Quantile Regressions. *Econometrics Journal*, *19*, C61–C94.

Bellman, R. (1957). *Dynamic Programming*. Princeton, NJ: Princeton University Press.

Blevins, J. (2016). Sequential Monte Carlo Methods for Estimating Dynamic Microeconomic Models. *Journal of Applied Econometrics*, *31*(5), 773–804.

Bonhomme, S. (2006). *Standard Errors Estimation in Mixture of Partial Likelihood Models.* University of Chicago. https://sites.google.com/site/stephanebonhommeresearch/.

Boyd, D., Lankford, H. & Wyckoof, J. (2013). Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Disentangling the Preferences of Teachers and Employers. *Journal of Labor Economics*, *31*(1), 83–117.

Broniatowski, M., Celeux, G. & Diebolt, J. (1983). Reconnaissance de Mélanges de Densités par un Algorithme d'Apprentissage Probabiliste. In *Data analysis and informatics.* North Holland, Amsterdam.

Bruins, M., Duffy, J., Keane, M. & Smith, A. (2018). Generalized Indirect Inference for Discrete Choice Models. *Journal of Econometrics*, *205*(1), 177–203.

Brunet, C., Kamionka, T. & Lacroix, G. (2025). Homeownership, Labour Market Transitions and Earnings. *Applied Economic*, *57*(14), 1614–1636.

Cameron, S. & Heckman, J. (1998). Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males. *Journal of Political Economy*, *106*(2), 262–333.

Cappellari, L. & Jenkins, S. (2003). Multivariate Probit Regression Using Simulated Maximum Likelihood. *The Stata Journa*, *3*(3), 278–294.

Celeux, G. & J., D. (1985). The SEM Algorithm: a Probabilistic Teacher Algorithm Derived from EM Algorithm for the Mixture Problem. *Computational Statistics Quarterly*, *2*, 73–82.

Chang, H., Fu, M., Hu, J. & Marcus, S. (2006). *Simulation-Based Algorithms for Markov Decision Processes*. Springer.

Chang, S.-K. (2011). Simulation Estimation of Two-Tiered Dynamic Panel Tobit Models with an Application to the Labor Supply of Married Women. *Journal of Applied Econometrics*, *26*(5), 854–871.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38.

Devroye, L. (1986). *Non uniform random variate generation*. Springer Science + Business Media.

Duffie, D. & Singleton, K. (1993). Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, *61*(4), 929–952.

Eckstein, Z. & Wolpin, K. (1999). Why Youths Drop Out of High School: The Impact of Preferences, Opportunities and Abilities. *Econometrica*, *67*(6), 1295–1339.

Fermanian, J. & Salanié, B. (2004). A Nonparametric Simulated Maximum Likelihood Estimation Method. *Econometric Theory*, *20*(4), 701–734.

Fougère, D. & Kamionka, T. (1992). Un Modèle Markovien du Marché du Travail. *Annales d'Économie et de Statistique*, *27*, 149–188.

Fougère, D. & Kamionka, T. (2003). Bayesian Inference for the Mover-Stayer Model in Continuous Time with an Application to Labour Market Transition Data. *Journal of Applied Econometrics*, *18*(6), 697–723.

Frazier, D., Oka, T. & Zhu, D. (2019). Indirect Inference with a Non-Smooth Criterion Function. *Journal of Econometrics*, *212*(2), 623–645.

Frazier, D. & Renault, E. (2020). Indirect Inference with(out) Constraints. *Quantitative Economics*, *11*(1), 113–159.

Gallant, A. & Tauchen, G. (1996). Which Moments to Match? *Econometric Theory*, *12*(4), 657–681.

Gates, R. (2006). A Mata Geweke-Hajivassiliou-Keane Multivariate Normal Simulator. *The Stata Journal*, *6*(2), 190–213.

Geweke, J. (1989a). Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica*, *57*(6), 1317–1339.

Geweke, J. (1989b). *Efficient Simulation from the Multivariate Normal Distribution Subject to Linear Inequality Constraints and the Evaluation of Constraint Probabilities.* Discussion paper (Duke University, Durham, NC).

Gilbert, L., Kamionka, T. & Lacroix, G. (2011). The Impact of Government-Sponsored Training Programs on the Labor Market Transitions of Disadvantaged Men. In *Advances in econometrics - theory and applications, dr. Miroslav Verbic (ed.).* InTech.

Gouriéroux, C. & Monfort, A. (1991). Simulation Based Inference in Models with Heterogeneity. *Annales d'Economie et de Statistique*, *20/21*, 69–107.

Gouriéroux, C. & Monfort, A. (1993). Simulation-Based Inference A Survey with Special Reference to Panel Data Models. *Journal of Econometrics*, *59*(1-2), 5–33.

Gouriéroux, C. & Monfort, A. (1995a). *Simulation Based Econometric Methods*. Core Lecture Series.

Gouriéroux, C. & Monfort, A. (1995b). *Statistics and Econometric Models, vol. 1 and 2*. Cambridge University Press.

Gouriéroux, C., Monfort, A. & Renault, E. (1993). Indirect Inference. *Journal of Applied Econometrics*, *8*, S85–S118.

Gouriéroux, C., Monfort, A. & Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Theory. *Econometrica*, *52*(3), 681–700.

Haan, P. & Uhlendorff, A. (2006). Estimation of Multinomial Logit Models with Unobserved Heterogeneity Using Maximum Simulated Likelihood. *The Stata Journal*, *6*(2), 229–245.

Hajivassilou, V. & McFadden, . (1990). *The Method of Simulated Scores for the Estimation of LDV Models: with an Application to the Problem of External*

Debt Crises. Cowles foundation discussion paper No. 967. Yale University, New Haven, CT.

Heckman, J. & Singer, B. (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica*, *52*(2), 271–320.

Hole, A. (2007). Fitting Mixed Logit Models by Using Maximum Simulated Likelihood. *The Stata Journal*, *7*(3), 388–401.

Kaji, T. & Manresa, E. (2026). Why Do the Elderly Save? Using Health Shocks to Uncover Bequest Motives. *The Japanese Economic Review*, *77*, 355–378.

Kaji, T., Manresa, E. & Pouilot, G. (2023). An Adversarial Approach to Structural Estimation. *Econometrica*, *91*(6), 2041–2063.

Kamionka, T. (1998). Simulated Maximum Likelihood Estimation in Transition Models. *The Econometrics Journal*, *1*, 129–153.

Kamionka, T. (2022). Sporting Activity, Employment Status and Wage. *Revue d'économie politique*, *132*(1), 49–78.

Kamionka, T. (2025). *Does Neighborhood Matter? The Impact of Proximity to (Dis)amenities on Home Price.* Hal, hal-05332717, https://polytechnique.hal.science/hal-05332717/.

Keane, M. (1994). A Computationally Practical Simulation Estimator for Panel Data. *Econometrica*, *62*(1), 95–116.

Keane, M. & Sauer, R. (2009). Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior. *Econometrica*, *77*(3), 975–991.

Keane, M. & Sauer, R. (2010). A Computationally Practical Simulation Estimation Algorithm for Dynamic Panel Data Models with Endogenous State Variables. *International Economic Review*, *51*(4), 925–958.

Keane, M. & Wolpin, K. (1997). The Career Decisions of Young Men. *Journal of Political Economy*, *105*(3), 473–522.

Keane, M. P. & Wolpin, K. I. (1994). The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence. *The Review of Economics and Statistics*, *76*(4), 648–672.

Keane, M. P. & Wolpin, K. I. (2009). Empirical Applications of Discrete Choice Dynamic Programming Models. *Review of Economic Dynamics*, *12*(1), 1–22.

Kennan, J. & Walker, J. (2011). The effect of expected income on individual migration decisions. *Econometrica*, *79*(1), 211–251.

Kiefer, N. (1980). A Note on Switching Regressions and Logistic Discrimination. *Econometrica*, *48*(4), 1065–1068.

Kingma, D. & Ba, J. (2015). *Adam: A Method for Stochastic Optimization.* In International Conference on Learning Representations (ICLR).

Kloek, T. & van Dijk, H. (1978). Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo. *Econometrica*, *46*(1), 1–19.

Kristensen, D. & Shin, Y. (2012). Estimation of Dynamic Models with Nonparametric Simulated Maximum Likelihood. *Journal of Econometrics*, *167*(1), 76–94.

Laplace, P.-S. (1886). *Œuvres Complètes de Laplace: Théorie Analytique des Probabilités, volume 7, gauthier-villars.*

Laroque, G. & Salanié, B. (1993). Simulation-Based Estimation of Models with Lagged Latent Variables. *Journal of Applied Econometrics*, *8*(S), S119–S133.

Leclerc, G.-L. (1777). *Essai d'Arithmétique Morale.* Supplément à l'histoire naturelle", Volume 4.

Lee, B.-S. & Ingram, B. (1991). Simulation Estimation of Times-Series Models. *Journal of Econometrics*, *47*(2-3), 197–205.

Lee, L.-F. (1992). On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models. *Econometric Theory*, *8*, 518–552.

Lee, L.-F. (1997). Simulated Maximum Likelihood Estimation of Dynamic Discrete Choice Statistical Models Some Monte Carlo Results. *Journal of Econometrics*, *82*, 1–35.

Lerman, S. & Manski, C. (1981). On the Use of Simulated Frequencies to Approximate Choice Probabilities. In *C. Manski and D. McFadden, ed., Structural analysis of discrete data with econometric applications.* MIT Press, Cambridge.

Marschak, J. (1960). Binary Choice Constraints on Random Utility Indications. In *Arrow K. (ed) Stanford symposium on mathematical methods in the social sciences* (pp. 312–329). Stanford University Press.

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

McFadden, D. & Train, K. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*, *15*(5), 447–470.

Meng, X.-L. & Rubin, D. (1993). Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. *Biometrika*, *80*(2), 267–278.

Metropolis, N. (1987). The Beginning of the Monte Carlo Method. In *Los alamos science special issue* (pp. 125–130).

Michener, R. (1984). Permanent Income in General Equilibrium. *Journal of Monetary Economics*, *13*(3), 297–305.

Mroz, T. (1999). Discrete Factor Approximations in Simultaneous Equation Models: Estimating the Impact of a Dummy Endogenous Variable on a Continuous Outcome. *Journal of Econometrics*, *92*(2), 233–274.

Newey, K. & West, K. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708.

Nielsen, S. (2000a). On Simulated EM Algorithms. *Journal of Econometrics*, *96*, 267–292.

Nielsen, S. (2000b). The Stochastic EM Algorithm: Estimation and Asymptotic Results. *Bernoulli*, *6*(3), 457–489.

Pakes, A. (1986). Patents as Options: Some Estimates of the Value of Holding European Patent Stock. *Econometrica*, *54*(4), 755–784.

Pakes, A. & Pollard, D. (1989). Simulation and the Asymptotics of Oprimization Estimators. *Econometrica*, *57*(5), 1027–1057.

Plum, A. (2016). bireprob: An estimator for Bivariate Random-Effects Probit Models. *The Stata Journal*, *16*(1), 96–111.

Provencher, B., Baerenklau, K. & Bishop, R. (2002). A Finite Mixture Logit Model of Recreational Angling with Serially Correlated Random Utility. *American Journal of Agricultural Economics*, *84*(4), 1066–1075.

Revelt, D. & Train, K. (1998). Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. *The Review of Economics and Statistics*, *80*(4), 647–657.

Robert, C. & Casella, G. (2001). A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Statistical Science*, *26*(1), 102–115.

Robert, C. & Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer New York.

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, *55*(5), 999–1033.

Rust, J. (1994). Structural Estimation of Markov Decision Processes. In *Handbook of Econometrics* (Vol. 4, pp. 3082–3139). Amsterdam: North Holland.

Rust, J. (1997). Using Randomization to Break the Curse of Dimentionality. *Econometrica*, *65*(3), 487–516.

Ruud, P. (1991). Extensions of Estimation Methods Using the EM Algorithm. *Journal of Econometrics*, *49*(3), 305–341.

Sarrias, M. (2016). Discrete Choice Models with Random Parameters in R: The Rchoice Package. *Journal of Statistical Software*, *74*(10), 1–30.

Sauer, R. & Taber, C. (2021). Understanding Women's Wage Growth Using Indirect Inference with Importance Sampling. *Journal of Applied Econometrics*, *36*(1), 453–473.

Stewart, M. (2006). Maximum Simulated Likelihood Estimation of Random-Effects Dynamic Probit Models with Autocorrelated Errors. *The Stata Journal*, *6*(2), 256–272.

Tincani, M. (2021). Teacher Labor Markets, School Vouchers, and Student Cognitive Achievement: Evidence from Chile. *Quantitative Economics*, *12*(1), 173–216.

Train, K. (2008). EM Algorithms for Nonparametric Estimation of Mixing Distributions. *Journal of Choice Modelling*, *1*(1), 40–69.

Train, K. (2016). Mixed logit with a Flexible Mixing Distribution. *Journal of Choice Modelling*, *19*, 40–53.

Train, K. & Winston, C. (2007). Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers. *International Economic Review*, *48*(4), 1469–1496.

Wu, C. (1983). On the Convergence Properties of the EM Algorithm. *Annals of Statistics*, *11*(1), 95–103.

Xun, Z. & Lubrano, M. (2016). Simulation Estimation of Two-tiered Dynamic Panel Tobit Models with an Application to the labour Supply of Married Women: A Comment. *Journal of Applied Econometrics*, *31*(4), 756–761.