On Some Uses of Simulation in Econometrics

Adrian Pagan
Economics Program
Research School of Social Sciences
Australian National University

Abstract Simulation methods are now widely used in econometrics. The range of uses covers the estimation of parameters of models as well as the use of models. In this paper we discuss how simulation methods can be used to investigate some issues that have proven extremely difficult to handle analytically. Specifically, we consider the questions of how to measure the business cycle characteristics associated with macroeconomic models and the estimation of the parameters of a popular latent variable model.

1. INTRODUCTION

Over a decade ago the late Chris Higgins reflected on what he had learned during a distinguished career in Australian economic modelling for policy decisions. He summarized his thoughts with the maxim 'simulate early and simulate often'. This is a remarkably prescient observation. Today simulation has firmly established itself within the set of activities engaged in by modellers and in econometric work more generally. It is almost impossible to write about econometrics today without acknowledging the importance of simulation methods. Yet in this continuity there has been change. When Chris spoke about the need to simulate, he was referring to the use of simulation methods as a way of learning about a model and the interactions within it. Indeed simulation was the basic ingredient in the response dissection technique of Helliwell and Higgins (1976), wherein one attempted to discover which parts of the system were responsible for particular outcomes by temporarily removing them from the system. What Chris did in this work was to use simulation methods to tackle questions regarding large scale and mildly non-linear systems that could not be readily answered analytically.

Most of us got our first exposure to simulation methods in one of two ways. The first was through the application just described, while the second was when we wished to know something more about the properties of estimators and tests that econometricians were either wont to dream up or to import from the statistics literature. Those of us who took this latter path became heavily involved in stochastic simulations, learning how to do it efficiently with the aid of techniques such as anti-thetic and control variables. In fact the two streams were not mutually exclusive. Everyone recognised the desirability of stochastically simulating econometric models, but few had the resources to do so. Moreover, it didn't seem likely that it would lead to any radically different insights, the reason being that the stochastic simulations simply tried to account for the fact that equations were not exact when drawing conclusions from experiments conducted with the models. Fundamentally, the uncertainty coming from these errors did not enter into the modelling process in any substantial way. Around 1980 this began to change. Lowered computational costs, the advent of powerful matrix languages such as GAUSS and MATLAB, and the development of stochastic programming methods meant that one could explicitly allow for the influence of such uncertainty upon decisions. Expectations about events now became a powerful factor in determining the dynamics of such models. The idea that unpredictable stochastic elements, 'shocks', are the driving forces of economic systems came to be treated as the standard view, a fact that is evident today in almost any academic and much policy discussion. The ability to stochastically simulate models became an important element in this transition.

Just as the possibility of performing stochastic simulations meant that the nature of economic modelling would be changed, so too did this possibility open up a much broader class of models that econometricians might be able to study. In particular, it become possible to handle models featuring latent variables. Earlier work with the latter used the Kalman filter but, given the strong assumptions employed in its derivation, it was desirable to produce methods that were more general. Under the heading of 'indirect estimation' such methods have now become quite popular. Since latent variables models appear in many guises—factors in the Arbitrage Pricing Model in capital asset pricing; the instantaneous rate in term structure models; and total productivity shocks in real business cycle models of the macro-economy—such developments are very important.

From the above sketch it is clear that simulation is now a vital part of quantitative analysis. But it really is a thumbnail sketch and totally abstracts from many other uses of simulation methods—examples being bootstrap methods for generating sampling distributions and the numerical evaluation of multivariate integrals, the latter having revolutionised Bayesian and much micro econometrics. It also ignores what are more mundane uses of the methods. I suspect that the standard approach to many problems today would be to first simulate what is to be studied in order to get some idea of the issues that are likely to arise when one is attempting to explain outcomes. Thirty years ago the analagous step in this sequence was to get a mathematical solution to the problem and to then engage in sensitivity analysis. Whilst the latter course of action is still probably the most informative, simulation has democratized the process of formulating and using models. One can learn a lot about them just from simulation and no longer does one have to be a superb mathematician to acquire such knowledge. What distinguishes workers now is their ability to think constructively about the design of simulation experiments, not their mathematical prowess.

Faced with all of the above ways in which simulation gets used in econometrics I have decided to recount two examples from my own research in which simulation has been the main technique for analysing an issue. The first of these topics relates to what it is we need to generate a business cycle while the second looks at the estimation of some popular latent variable models in econometrics.

Table 1: AR(2) fitted to GDP*

Country	b_1	b_2	Period
Australia	.889	.077	1960Q1–1996Q4
Canada	1.207	-.218	1960Q1–1966Q4
Denmark	.792	.077	1977Q1–1996Q4
France	1.079	-.142	1970Q1–1996Q4
Japan	1.130	-.145	1970Q1–1996Q4
The Netherlands	.610	.294	1977Q1–1996Q4
New Zealand	.444	.303	1982Q1–1996Q4
Switzerland	1.411	-.460	1967Q3–1996Q4
United Kingdom	.955	-.017	1960Q1–1996Q4
United States	1.258	-.299	1960Q1–1996Q4

* The model fitted is $\log GDP_t = a + b_1 \log GDP_{t-1} + b_2 \log GDP_{t-2} + ct$

Data Source: OECD Data base in DX.

2. GENERATING A BUSINESS CYCLE

Economies experience business cycles. For at least a century models have been built in an attempt to produce an explanation of this phenomenon. Many of the 'models' are too vague to be described mathematically, but those that have been made precise fall into two categories. In the first category are those which produce a description of output as a second order autoregressive process:

$$y_t = b_1 y_{t-1} + b_2 y_{t-2} \qquad (1)$$

where the roots of the polynomial in the lag operator L, $(1-b_1L-b_2L^2)$, are taken to be complex. Because this would produce a periodic cycle it is generally augmented with a shock u_t, so that cycles would not all be the same. Following the latter strategy leads to cycles being described as occurring when there is a peak in the spectral density of y_t. The biggest difficulty with this formulation is that there is almost no empirical evidence to support it. Letting y_t be the log of Gross Domestic Product (GDP), after regressing out a linear trend, Table 1 shows estimates of b_1 and b_2 from ten OECD countries. For no country are there any complex roots.

One response to this observation has been to dispense with the linearity of (1) and the resulting non-linear models for y_t have had a long history in business cycles research e.g, Goodwin (1967). Today there are many papers appearing with such an orientation. Some of the proponents of this solution are very dismissive of research into business cycles that feature linear models e.g. Blatt (1983) and Keen (1995), as they claim that linear models are incapable of generating a realistic cycle. A different response would be that it is not the model that needs to be changed but rather the definition of a cycle. Equating the existence of a business cycle with complex roots in (1), or the possibility of chaotic solutions in non-linear models, ignores the fact that this is not the way in which business cycles are actually identified and described. What we know about business cycles stems from an analysis of *patterns within data* and, in particular, from the identification of local peaks and troughs in a series purporting to measure activity. Ideally this is done visually, but any observer will almost certainly filter out some variation that is not regarded as sufficiently 'long-lived' or of insufficient magnitude to be treated as a recession or expansion. Replication of turning points would therefore be very difficult. Attempts to make the dating process replicable have therefore focused upon the formulation of some rules to give discipline to the task. Of these the best known set of rules would be those of the National Bureau of Economic Research in the USA, as codified in the computer program of Bry and Boschan (1971).[1] Table 2 presents these rules.

Table 2: Bry–Boschan Procedure for Programmed Determination of Turning Points

I Determination of extremes and substitution of values.

II Determination of cycles in 12-month moving average (extremes replaced)
- A. Identification of points higher (or lower) than 5 months on either side.
- B. Enforcement of alternation of turns by selecting highest of multiple peaks (or lowest of multiple troughs).

III Determination of corresponding turns in Spencer curve (extremes replaced)
- A. Identification of highest (or lowest) value within ±5 months of selected turn in 12-month moving average.
- B. Enforcement of minimum cycle duration of 15 months by eliminating lower peaks and higher troughs of shorter cycles.

IV Determination of corresponding turns in short-term moving average of 3 to 6 months, depending on MCD (months of cyclical dominance)
- A. Identification of highest (or lowest) value within ±5 months of selected turn in Spencer curve.

V Determination of turning points in unsmoothed series
- A. Identification of highest (or lowest) value within ±4 months of selected turn in short-term moving average.
- B. Elimination of turns within 6 months of beginning and end of series.
- C. Elimination of peaks (or troughs) at both ends of series which are lower (or higher) than values closer to end.
- D. Elimination of cycles whose duration is less than 15 months.
- E. Elimination of phases whose duration is less than 5 months.

VI Statement of final turning points.

Adopting the definition of a cycle implicit in the rules of Table 2 we are left with the question of whether a linear process like (1), without complex roots, is capable of generating a business cycle such as seen in countries

[1] This program is widely used in the actual dating of business cycles, although, in practice, the program output is the input into a complex process. As discussed in Bry and Boschan, it does quite well in predicting the results of this more complex process.

around the world. Columns (i)–(ii) of Table 3 present characteristics of the cycle in Australia and the United Kingdom. Clearly the cycles are of different length and it is noticeable that expansions are much longer than contractions, an 'asymmetry' feature that some argue is the justification for non-linear models e.g., Blatt (1983). Our objective is to study what processes for output could produce the observed characteristics. To simplify the presentation we will assume a special form of (1):

$$\Delta y_t = a + u_t, \qquad (2)$$

where y_t = logGDP-at, 'a' represents the trend growth rate in GDP, u_t is a shock and $\Delta y_t = y_t - y_{t-1}$. The shock potentially has some serial correlation i.e.,

$$u_t = \rho u_{t-1} + e_t, \qquad (3)$$

where e_t has no serial correlation with standard deviation of σ. Some of the countries in Table 1 can be represented very well with (2) and (3). Prominent among these are Australia ($\rho \equiv 0$), the United Kingdom ($\rho \equiv 0$) and the US ($\rho \cong .4$). We will focus upon these countries to illustrate the arguments.

Table 3: Mean Business Cycle Characteristics (months)

Process	*(i)*	*(ii)*	*(iii)*	*(iv)*	*(v)*	*(vi)*
Duration of contraction	15	14	12	14	21	22
Duration of expansion	65	45	72	47	21	22
Cycle length	80	59	84	61	42	44

Notes: (i) Australian data.
(ii) UK Data.
(iii) (2)–(3) for Australia, ρ=0
(iv) (2)–(3) for UK, ρ=0
(iv) (2)–(3) for UK, ρ=0, a=0
(v) (2)–(3) for UK, ρ=.4, a=0

Now it is very difficult to see if (2) and (3) could generate a cycle of the type evident in columns (i) and (ii) of Table 3. The principal difficulty comes from the fact that the rules of the Bry–Boschan program are almost impossible to analyse analytically. It is here that simulation can be a great help. For given values of 'a', ρ, and a way of generating u_t, it is possible to pass the simulated y_t through the Bry–Boschan program and to then check if (2)–(3) can produce a realistic cycle. Table 3, columns (iii) and (iv), give the average durations of the cycle and its phases using 200 simulated series from (2)–(3). In doing these two simulations the values of a and σ are estimated from quarterly data on GDP for Australia and the United Kingdom separately—see Pagan (1997a, 1997b) for details (note that ρ=0). Column (iii) is for Australia and column (iv) for the United Kingdom. The differences between the two simulations are that both a and σ are higher for Australia, as is the ratio a/σ.

Table 3 shows that the simple linear model of (2) and (3) is capable of reproducing the broad details of the cycle, specifically the duration of the cycle and the asymmetry of expansions and contractions. To understand the elements of the explanation, column (v) performs the same simulation as in column (iii) but now 'a' is set to zero. Obviously it is the trend growth in output that produces the asymmetry in cycles and it is a key element in determining the duration of the cycle. What then is the role of a non-zero value of ρ? A simulation is run in column (vi) that is identical to that in (v) except that it sets ρ=.4 when generating data. It is clear that the presence of small amounts of serial correlation in u_t has little effect upon the nature of cycles.

Clearly Table 3 is a challenge to those who maintain that linear models cannot generate a realistic cycle and the results raise the question of exactly what it is that a non-linear model of the cycle would provide that cannot be done with a linear model. Models such as Goodwin's don't provide an explanation of the magnitude of a and are silent about σ as well, since they are deterministic. To answer the question one needs to be able to simulate some of these non-linear models and to compute cycle characteristics. That leads to two complications. First, some stochastic elements must be introduced and there is no obvious place for inserting them. Second, and more importantly, it is hard to find any examples in the literature that are claimed to be representative of actual economies. Mostly, their proponents are content to point to the fact that the models can generate a cycle under certain calibrations of their model and there is no attempt to show that the processes for output generated by these non-linear models would look like the results in Table 1.

In fact, one would expect that many properly calibrated non-linear models, such as Goodwin's, will produce realistic cycles, simply because they imply a model such as (2) with u_t a non-linear function of Δy_{t-1} and the non-linearity is not needed to produce realistic cycles. These predator-prey models feature a mechanism whereby the prey (workers) earn income (wages) that depends on the aggregate amount of employment i.e. in conventional economic terms a Phillips curve is embedded within them. Changes in wage rates impact upon the rate of profit and, through a non-linear structure, the rate of investment. Trend growth, 'a', is assumed known and the near unit root predicted to be in output stems from a slow response of wages to unemployment. As already seen in Table 3, a non-linearity in u_t is unimportant for the *existence* of a cycle and for replicating its main features. Where non-linearity may have a role is in explaining some characteristics of the cycle not dealt with in Table 3. Specifically, there is the issue of *duration dependence* i.e. the contention that the probability of exiting from a state such as recession depends on how long has been spent in it. Evidence on such a characteristic is not easy to get, since there are so few cycles, and it is really only US data that is sufficiently developed to produce a reliable analysis. Consequently, substantial work needs to be done on this topic. Despite this, the important lesson to be learned from the simulations of Table 3 is that proponents of non-linear cycle models need to be much more precise about what characteristics of the cycle they seek to explain and why linear models cannot capture these effects. Early analysis, for example Blatt (1983), adopted this stance, but the characteristics focussed on, principally asymmetry, *can* be produced by linear models.

3. ESTIMATING SOME LATENT VARIABLE MODEL

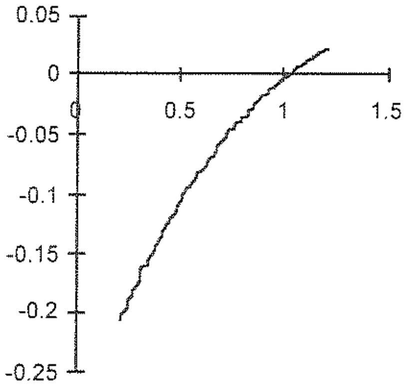
There has been an increasing trend in econometrics towards the use of latent variables as an essential part of models. In many of these cases it is hard to engage in maximum likelihood estimation, since the likelihood involves only those random variables for which there are realizations, and this makes it necessary to integrate out any latent variables. Sometimes the order of the integration which needs to be performed is very high. In these instances. Simulation based methods of estimation appeal. The variant adopted here is that termed 'indirect estimation' in Gourieroux, Monfort and Renault (1993), although it has been implemented as proposed in Gallant and Tauchen (1997). Martin and Pagan (forthcoming) contains details of the main application presented in this section.

The method distinguishes between a *statistical or auxiliary* model which is fitted to the data and a *model of interest* whose parameters it is desired to estimate. In the method of indirect estimation the parameters of interest are estimated *indirectly* from the auxiliary model rather than *directly*, as would occur with maximum likelihood. The underlying theory that justifies the approach is that of pseudo-maximum likelihood. If π are the parameters of the auxiliary model and θ are those of interest then it is well known that the scores with respect to π (d_π) i.e. the first derivative of the assumed log likelihood of the auxiliary model, have the property that $E_\theta\,[d_\pi(\pi^*)]=0$, where E_θ is the expectation taken with respect to the data generating process of the model of interest and π^* is the pseudo maximum likelihood estimator. The latter is a function of θ_0, the true value of θ. These facts suggest that one estimate θ by finding a value for θ which sets $E_\theta[d_\pi(\pi^*)]=0$. If π^* was known and E_θ could be found analytically, we could therefore estimate θ by solving the set of equations $E_\theta[d_\pi(\pi^*]=0$ for θ. Once this principle is accepted it only remains to replace π^* by some observable value and to find a method for evaluating E_θ. The former task is accomplished by using the MLE of π from the auxiliary model in place of π^*, as this is known to be a consistent estimator, while the second is resolved through simulation: for any given value of θ one can simulate artificial data from the model of interest thereby enabling one to estimate $E_\theta[d_\pi(\pi^*)]$ as a sample average.

A simple example can be used to illustrate the idea. Suppose the model of interest is a Probit model, with a single variable x_i influencing the probability of a decision i.e. $pr\{y_i=1\} = \Phi(x_i\theta)$, where Φ is the cumulative standard normal density. Regressing y_i upon x_i will not consistently estimate θ, but it can be used as the auxiliary model i.e. one can act as if $y_i = x_i\pi+e_i$, where e_i is taken to be normally and independently distributed with zero mean and unit variance, and then find the value of θ which sets the expected value of the scores for π, $n^{-1}\sum_{i=1}^{n} x_i\,(y_i-x_i\pi^*)$, to zero. This expectation is $n^{-1}\sum_{i=1}^{n} x_i\,[\Phi(x_i\theta)-x_i\pi^*]$. Replacing π^* by the OLS estimate of π found by regressing y_i on x_i, gives an equation $\sum_{i=1}^{n} x_i\,[\Phi(x_i\hat{\theta})-x_i\hat{\pi}] = 0$, which yields the indirect estimator of θ.

I have generated 3000 observations from a probit model with $\theta = 1$. Figure 1 plots $n^{-1}\sum_{i=1}^{n} x_i\,[\Phi(x_i\hat{\theta})-x_i\hat{\pi}]$ for a range of values of θ; $\hat{\theta}$ will be where the pseudo-score is zero. In fact it produces an estimate of θ that is virtually identical to its true value.

Fig 1 Pseudo-Score for Indirect Estimation of a Probit Model



A second example concerns volatility in asset price markets Such markets are very volatile at certain times and quiescent at others. The volatility in prices is an important phenomenon as it makes investments in assets risky. Consequently, the provision of appropriate models of the volatility has become a major focus of econometric research in the late 1980s and early 1990s. Central to this research has been the search for some simple ways of describing the volatility, followed by the development of methods to estimate the resulting models. An attractive description is to conceive of the market as being in one of two states, termed high or low volatility. Specifying a set of transition probabilities to describe how one moves from one state to the other would then constitute a simple model of volatility. Such hidden state models have been of interest to many disciplines for some time. The state space form used extensively in control engineering is a classic example of such a model, although in that structure the state is viewed as being a continuous random variable which is driven by shocks whose moments are taken to be known or to depend upon observable quantities. The simple two-state formulation just described can be written in a state space format but the state space is discrete and shocks have moments that depend on unobservable quantities, so the estimation of such models is non-standard. Under specific distributional assumptions the model just described was estimated with maximum likelihood techniques by Hamilton (1989) and, since then, has had many applications to economic data sets such as economic activity, interest rates, exchange rates etc. Pagan and Schwert (1990) is an early application. That paper used the model to describe volatility in US stock returns over the period 1834 to 1925, and we concentrate upon that same data set here

One has to select an auxiliary model for this data set. Perhaps the best known statistical models for asset returns are those in the Autoregressive Conditional Heteroskedasticity (ARCH) class, introduced by Engle (1982). These model conditional volatility as a function of past returns. A version that has been very popular in the analysis of stock returns has been Nelson's (1990) exponential generalized ARCH (EGARCH) process,

$\log \sigma_t = \pi_0+\pi_1 \log \sigma_{t-1}+\pi_3 z_{t-1}$, where $z_t = \left[|\varepsilon_t| - \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \right] + \pi_4\ \varepsilon_t$, while ε_t is an identical and independently distributed random variable with mean zero and density equal to that of Student's t with π_5 degrees of freedom. Many programs exist to fit the EGARCH model to a set of data on returns. It therefore appeals as a good auxiliary model for the estimation of the parameters of the latent variable model. To formally describe the latter, let v_t be volatility at time t and let z_t take the value of unity when volatility is high and zero when volatility is low.[2] Then $v_t = \theta_0+\theta_1 z_t$ and the model is completed by describing how z_t moves from one state to another. Table 4 details these transition probabilities.

Table 4: Transition Probabilities for z_t

t-1\t	0	1
0	θ_3	$1-\theta_3$
1	$1-\theta_2$	θ_2

Table 5: Parameter estimates of Hamilton's model Stock returns data, monthly, 1834 to 1925

Parameter	Indirect Estimates	MLE (Hamilton)
θ_0	6.1173×10^{-4}	6.1022×10^{-4}
θ_1	19.9976×10^{-4}	18.4889×10^{-4}
θ_2	0.9017	0.9008
θ_3	0.9610	0.9614

Table 5 gives indirect estimates of θj (j=1, ..., 4), where simulation was used to evaluate the expected value of the pseudo-scores. It is noticeable that the MLE using the distributional assumptions in Hamilton (1989) and the indirect estimates are quite close so that any analysis using the indirect estimates seems likely to be applicable to the MLE as well.

Why would one to indirect estimation? Apart from the fact that maximum likelihood may be infeasible in many latent variable models, it is the case that one can learn a lot about the nature of a latent variable model by studying it through the 'lens' of the auxiliary model, simply because the latter summarizes characteristics of the data in a way that may be more familiar to us. To illustrate this consider the auxiliary EGARCH model. This model is known to capture a characteristic of volatility sometimes referred to as the 'leverage effect', i.e. volatility is higher in a bear market than a bull market. Consequently, it is interesting to ask whether the two state latent variable model being estimated above is capable of reproducing this characteristic. Again, simulation methods can be used to provide an answer to this. Using the parameter estimates of Table 5, 3000 realizations were fitted to an EGARCH process and, in all cases, it was noticeable that the EGARCH parameter which captures an asymmetric response of the conditional variance to news (π_4) is close to zero and *positive* (around .02). In the EGARCH model estimated from data it is -.11. Consequently, Hamilton's model does not replicate the leverage effect that exists in the data, whereby volatility is larger when returns are negative than positive, and this would lead us to reject it as a good description of the data.

[2] To aviod confusion it needs to be appreciated that v_t is meant to be the actual volatility whereas σ_t is a statistical description.

REFERENCES

Blatt, J.M., *Dynamic Economic Systems*, M.E. Sharpe, New York, 1983.

Bry, G. and C. Boschan, *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, NBER, New York, 1971.

Engle, R.F., Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica*, 50, 987–1008, 1982.

Gallant, A.R. and G. Tauchen, Which moments to match?, *Econometric Theory*, 12, 657-681, 1996.

Goodwin, R., A Growth Cycle, in C.H. Feinstein, ed., *Socialism, Capitalism and Economic Growth*, Cambridge, Cambridge University Press, 165–70, 1967.

Gourieroux, C., A. Monfort and E. Renault, Indirect inference, *Journal of Applied Econometrics*, 8, S85—S118, 1993.

Hamilton, J.D, A new approach to the economic Analysis of nonstationary time series and the business cycle, *Econometrica*, 57, 357–384, 1989.

Helliwell, J. and C.I. Higgins, Macroeconomic adjustment processes, *European Economic Review*, 7, 221–238, 1976.

Keen, S., Finance and economic breakdown: modelling Minsky's financial instability hypothesis, *Journal of Post Keynesian Economics*, 17, 608–635, 1995.

Martin, V. and A.R. Pagan, Simulation Based Estimation of Some Factor Models in Econometrics (with V. Martin), in T. Schuermann and R.S. Mariano, eds, *Simulation Based Estimation in Econometrics*, Cambridge University Press, 1997.

Nelson, D.B., Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, 347–370, 1990.

Pagan, A.R. and G.W. Schwert, Alternative models for conditional stock volatility, *Journal of Econometrics*, 45 (1990) 267—290, 1990.

Pagan, A.R., Towards an understanding of some business cycle characteristics, *Australian Economic Review*, 30, 1-15, 1997a.

Pagan, A.R., Policy, Theory and the Cycle, *Oxford Review of Economic Policy* (forthcoming), 1997b.