What is the price of maximum likelihood?

Søren Johansen*
Department of Applied Mathematics and Statistics
University of Copenhagen

18 October 2004

Abstract

This paper is based upon the paper by Peter Ireland 'A method for taking models to the data' Ireland (2004*a*). The model analysed by Ireland is a DSGE model and the method suggested is to add some autocorrelated error terms to the identities derived from theory, and estimate the statistical model by maximum likelihood methods using the Kalman filter. This method is undoubtedly a useful idea to make the economic relations more elastic, see Haavelmo (1943), but the application of the Gaussian likelihood has its price. We discuss, using some simple examples, how it is easy to make mistakes when applying Gaussian maximum likelihood without checking that the model chosen fits the data at hand. Many different methods can be derived by Gaussian maximum likelihood, and many give different results for inference when applied to a given data set. Thus one has to check the basic assumption of likelihood theory, namely that the density of the data is in the statistical model chosen. We suggest a number ways in which the statistical model embedding the DSGE model could be checked against the data. We conclude by a simulation experiment which investigates the consequence of insisting that a root of 0.9987 is in fact a highly persistent stationary root. We demonstrate that one cannot conduct inference on the steady state values and conclude by showing how a cointegration analysis gives a useful formulation of the problem.

*I would like to thank Massimo Franchi, Christian Groth, Henrik Hansen, Mikael Juselius and the members of the EFS-EMM network, as well as the audiences in Rome and Helsinki for useful comments.

1 Introduction

When taking an economic model to the data in order to conduct inference 'we need a stochastic formulation to make simplified relations elastic enough for applications', to quote Haavelmo (1943). Thus the relations derived from theory should be supplemented by some extra error terms to describe the stochastic variation of the measurements. We thereby extend the economic model to a statistical model, where we have embedded the economic model (or some of its consequences) as relations in the statistical model. In this way one can obtain estimates of relevant parameters and their uncertainty, and hence conduct inference on the economic relations.

The papers by Ireland (2001, 2004*a*) suggest such a method for taking a DSGE model to the data, a method which we describe briefly as follows:

The DSGE model delivers relations and first order conditions involving forward looking expectations and, assuming the variables are trend stationary, the relations are expanded around their steady state values. This gives identities and stochastic difference equations driven by the shocks of the theory model. The identities are then made 'more elastic' by adding an autocorrelated error term. The statistical model thus formulated is analyzed by Gaussian maximum likelihood methods.

This can clearly be a useful way of taking an economic model to the data, as it allows one to conduct inference on coefficients, provided one has successfully described the stochastic variation of the data.

The analyses of both the economic DSGE model and the embedding statistical model require mathematics.

The first order conditions for the optimization problem and the conditions for stability of the economic model, for instance, obviously have to be correctly derived by proper use of mathematics. Nobody wants a good guess of a first order condition. It is not enough just to *assume* that a first order condition has a certain form, or *assume* that the stability assumption is satisfied. That has to be checked, and that is of course done, see Ireland (2004*a*, 2004*b*).

On the other hand, when it comes to conducting inference applying the asymptotic theory of mathematical statistics, the mathematical assumptions behind these results are dealt with in a much less stringent fashion. It seems as if it is enough to *assume* stationarity of a process, and it seems as if it is enough to *assume* that a chosen model can be used for conducting inference for a given data set. Asymptotic results are used without mentioning that there could be finite sample problems, it is just *assumed* that there are enough observations.

There are, therefore, two aspects of the proposed methodology which need careful attention and the present note will focus on these:

- Which statistical model should one choose to embed the economic relations?
- Given a well chosen model, can we rely on the asymptotic results found in the statistical literature for the analysis of the data at hand?

In section 2 we describe the method proposed by Ireland in connection with the example analysed and briefly summarize the findings. We then discuss in section 3 the method of maximum likelihood and the basic assumption behind likelihood methods. We give some suggestions for how the statistical model formulated by Ireland can be checked before reliable inference can be conducted. Finally we have in section 4 some comments on the role of unit roots. We show by a simple simulation that if we insist that a root very close to one is a highly persistent stationary root, we may need many more observations than what is usually available for conducting inference on steady state values. Even so, some combinations of the steady state values allow valid inference and this can be formulated as a cointegration analysis assuming the root is one. Section 5 concludes.

2 The model, method and findings of Ireland

The DSGE model considered by Ireland (2001, 2004*a*, 2004*b*) has a representative agent producing output Y_t with capital K_t and labor H_t, measured as hours worked, according to

$$Y_t = A_t K_t^{\theta} (\eta^t H_t)^{1-\theta}. \quad (1)$$

The coefficient $\eta > 1$ measures the 'gross rate of labor-augmenting technological progress'. The agent has preferences over consumption C_t and hours worked H_t and wants to maximize

$$E \left[\sum_{t=0}^{\infty} \beta^t (\ln C_t - \gamma H_t) \right], \quad (2)$$

with respect to $\{C_t, H_t\}_{t=0}^{\infty}$ subject to a number of constraints. The technology shocks A_t follow a stationary $AR(1)$ process

$$\ln(A_t) = (1-\rho)\ln A + \rho \ln(A_{t-1}) + \varepsilon_t. \tag{3}$$

The variables Y_t, K_t, C_t and investment I_t obey the identities

$$Y_t = C_t + I_t \tag{4}$$

and

$$K_{t+1} = (1-\delta)K_t + I_t. \tag{5}$$

The first order conditions become

$$\gamma C_t H_t = (1-\theta)Y_t, \tag{6}$$

and

$$1 = \beta E_t\left[\frac{C_t}{C_{t+1}}(\frac{\theta Y_{t+1}}{K_{t+1}} + 1 - \delta)\right]. \tag{7}$$

The assumption of trend stationarity of the variables is added, in order to linearize the solution of these equations around their steady state values. We define the means or steady states y, c, i, k, a, h of Y_t/η^t, C_t/η^t, I_t/η^t, K_t/η^t, A_t, H_t respectively. For the observations Y_t, C_t, and H_t, expressed as deviations of their logarithm from their trend, the three variables

$$f_t = \begin{pmatrix} \log Y_t - t\log\eta - \log y \\ \log C_t - t\log\eta - \log c \\ \log H_t - \log h \end{pmatrix}$$

are introduced.

The two variables A_t and K_t are considered unobserved and define

$$s_t = \begin{pmatrix} \log K_t - t\log\eta - \log k \\ \log A_t - \log a \end{pmatrix}.$$

A linearization of the model equations around the steady state of the variables gives equations of the form

$$\begin{aligned} s_t &= \mathcal{A}s_{t-1} + \mathcal{B}\varepsilon_t, \\ f_t &= \mathcal{C}s_t. \end{aligned}$$

Here the matrices $\mathcal{A}, \mathcal{B}$, and $\mathcal{C}$ and the steady state values are computable functions of the parameters of the model: $\beta, \gamma, \theta, \eta, \delta, \gamma, A, \rho, \sigma$.

Because $\mathcal{C}$ is 3×2 and $f_t = \mathcal{C}s_t$, there must be a vector ξ so that $\xi'\mathcal{C} = 0$, and hence $\xi' f_t = 0$. No such identity holds in the data, so the method suggested for taking the model to the data is to consider the enlarged system

$$\begin{aligned} s_t &= \mathcal{A}s_{t-1} + \mathcal{B}\varepsilon_t, \\ f_t &= \mathcal{C}s_t + u_t, \\ u_t &= \mathcal{D}u_{t-1} + \xi_t, \end{aligned}$$

where u_t is an unobserved $AR(1)$ process with innovations ξ_t, i.i.d. $N_3(0, V)$. Thus the statistical model for the five variables, Y_t, K_t, C_t, H_t, and A_t are driven by four errors, $\varepsilon_t, \xi_{1t}, \xi_{2t}$, and ξ_{3t}. It is also assumed that ε_t is i.i.d. $N(0, \sigma^2)$, and independent of ξ_t.

This model can be written in state space form with state vector $x_t = (s_t', u_t')'$, state equation

$$\begin{pmatrix} s_t \\ u_t \end{pmatrix} = \begin{pmatrix} \mathcal{A} & 0 \\ 0 & \mathcal{D} \end{pmatrix} \begin{pmatrix} s_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} \mathcal{B}\varepsilon_t \\ \xi_t \end{pmatrix},$$

and observation equation

$$f_t = \mathcal{C}s_t + u_t = (\mathcal{C}, I_3) \begin{pmatrix} s_t \\ u_t \end{pmatrix}.$$

In this form the Gaussian likelihood function can be calculated using the Kalman filter and from these values the estimates of the parameters can be found by optimizing the likelihood function. The standard deviations are then found from the second derivative of the likelihood function.

The economic model is supplemented with the assumption that $\beta = 0.99$ and $\delta = 0.025$, as these coefficients were difficult to determine from the data.

2.1 The data

The data for consumption, investment, output and population are taken from the Federal Reserve Bank of St. Louis' FRED database; data for hours worked are from the Bureau of Labor Statistics' Establishment Survey. All data are quarterly from 1948:1 to 2002:2. The data can be found, together with supplementary material on the solution of the equations, on the home page http://www2.bc.edu/~irelandp.

There are observations of the variables
N_t = Civilian, noninstitutional population, age 16 and over.
C_t = Real Personal Consumption Expenditures in chained 1996 dollars/N_t.
I_t = Real Gross Private Domestic Investment in chained 1996 dollars/N_t.
H_t = Hours of wage and salary workers on private, non-farm payrolls/N_t.

2.2 The findings

The main positive finding is that the model seems to predict well out of sample as compared to some competitors, a VAR(1) and a VAR(2) model and a model with diagonal D.

As model misspecification check it was found that by comparing the smoothed values of ε_t and ξ_t, the estimated correlations were rather small (-0.0634, 0.0133, 0.0010). It was also found by Wald tests that the parameters of the economic model before and after 1980 have hardly been the same for the whole period.

The estimated values of ρ (0.9987) and the largest eigenvalue of D (0.9399) are very close to one, so that some doubt is thrown on the stationarity assumptions.

3 The method of maximum likelihood

There is hardly any reason for giving the details of this method which is standard in econometrics, but let me briefly summarize the method as follows:

We want to analyse data $x_1, \ldots, x_T$ and define a statistical model by the densities

$$p(x_1, \ldots, x_T, \theta),\ \theta \in \Theta,$$

where Θ is the parameter space. A submodel, or hypothesis on the parameter θ, is formulated as $\theta \in \Theta_0 \subset \Theta$. The likelihood function $L(\theta)$ is the density as a function of θ defined on the set Θ.

Applying the method of maximum likelihood gives, under certain stationarity and regularity conditions, the following benefits:

- The estimates are calculated by optimizing the likelihood function $L(\theta)$ over the parameter set Θ, and they are consistent and asymptotically Gaussian.

- The observed information is $I(\hat{\theta}) = -d^2 \log L(\theta)/d^2\theta|_{\theta=\hat{\theta}}$, and an estimate of the asymptotic variance can be found from $I(\hat{\theta})^{-1}$.
- The likelihood ratio test of $\theta \in \Theta_0$ is $-2\log(\max_{\theta\in\Theta_0} L(\theta)/\max_{\theta\in\Theta} L(\theta))$ and the asymptotic distribution is χ^2.

Thus, once we can apply this methodology, we have general theorems from statistics and probability, which ensure that we have answers to a number of relevant questions in inference, without having to do more than finding a program to optimize the likelihood function and print out $\hat{\theta}, I(\hat{\theta})^{-1}$ and $L(\hat{\theta})$, for the various parameter spaces corresponding to the hypotheses we want to test.

This looks like a free lunch, but of course the method has its price. We have to choose a reasonable model on which to base our likelihood analysis, and it is in the art of model choice that the cost involved in likelihood based inference is found.

In order to make this point very clear, consider a univariate time series generated from the following DGP

$$x_t = 0.9x_{t-1} + 1.0 + \varepsilon_t,\ t = 1, \ldots, 100$$

where ε_t are i.i.d. $N(0,1)$. Note that $E(x_t) = 1/(1-0.9) = 10$, and $Var(x_t) = 1/(1-0.9^2) \sim 5.26$. A sample of the data is shown in Figure 1.

Consider now two different methods for conducting inference in the form of asymptotic confidence intervals on the mean of the process x_t. Both methods are based on Gaussian maximum likelihood, but inference based upon the two model is very different.

Method 1: We model the data as

$$x_t = \mu + \varepsilon_t,$$

with ε_t i.i.d. $N(0,\sigma^2)$, so that $E(x_t) = \mu$. We want to make inference on $E(x_t)$ and find that the maximum likelihood estimator is the average $\hat{\mu} = \bar{x} = T^{-1}\sum_{t=1}^{T} x_t$, which is distributed as $N(\mu, \sigma^2/T)$, and $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(x_t - \bar{x})^2$ so that an asymptotic 95% confidence interval is given by

$$\hat{\mu} \pm \frac{1.96}{\sqrt{T}}\hat{\sigma} = 9.14 \pm 0.45.$$

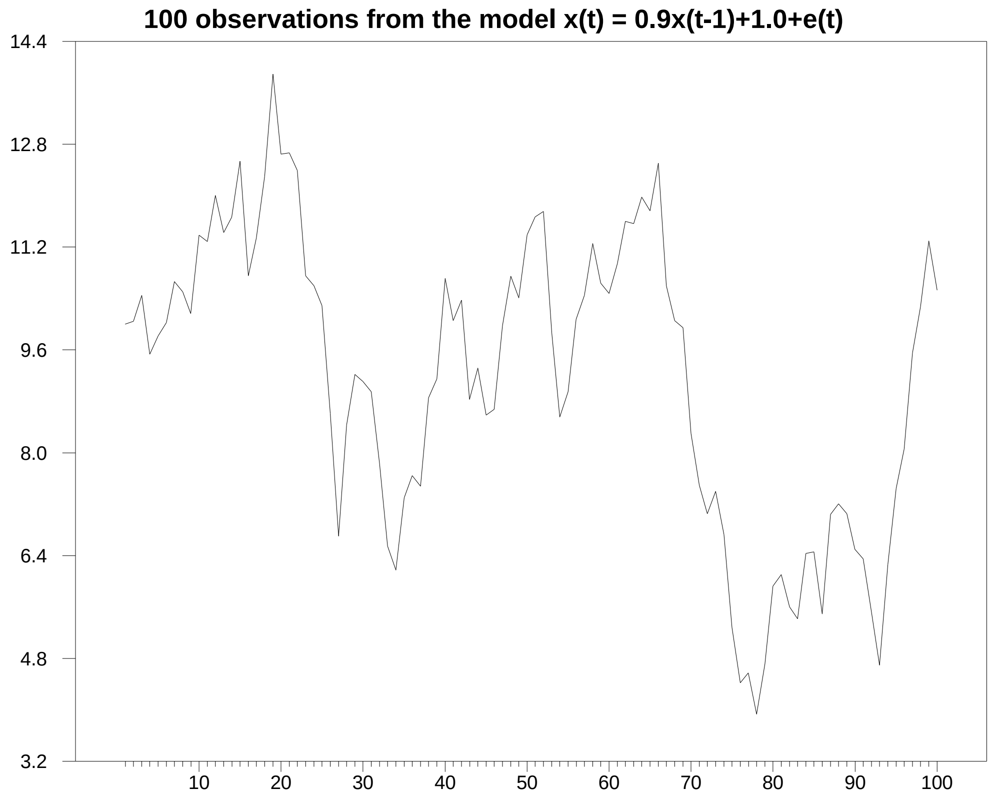

Figure 1: A simulation of 100 observations from a univariate AR(1) process $x_t = 0.9x_{t-1} + 1.0 + \varepsilon_t$, where ε_t i.i.d. $N(0, 1)$.

Method 2: We could also model the data as a stationary time series

$$x_t = \rho x_{t-1} + \mu(1-\rho) + \varepsilon_t,$$

where again $E(x_t) = \mu$, if $|\rho| < 1$. We find the estimate

$$\hat{\mu} = \bar{x} - \hat{\rho}T^{-1}(x_0 - x_T)/(1-\hat{\rho}),$$

and $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^T(x_t - \hat{\rho}x_{t-1} - \hat{\mu}(1-\hat{\rho}))^2$. For $|\rho| < 1$, $\hat{\mu}$ has the same asymptotic distribution as the average $\bar{x}$, but now with a larger variance $\sigma^2 T^{-1}(1-\rho)^{-2}$, which gives a wider asymptotic 95% confidence set for μ

$$\hat{\mu} \pm \frac{1.96}{\sqrt{T}}\frac{\hat{\sigma}}{1-\hat{\rho}} = 9.20 \pm 23.21.$$

In order to conduct Gaussian likelihood inference we have to choose between these two model and possibly many more. If we choose the first model we get a confidence interval which is much smaller than if we choose the second model. Thus inference is seriously different depending on the model we choose, even though in both cases we are interested in the same quantity, namely the mean of the process.

3.1 What is the price of maximum likelihood?

The two methods outlined above are Gaussian maximum likelihood methods, but they are evidently different, as they are based upon different models and therefore different likelihoods and the confidence interval for the second method is much wider than the interval given by the first method when $\rho = 0.9$. Even though both methods give almost the same estimator, the estimate of the variance of the estimator and hence the confidence intervals are different.

Evidently using maximum likelihood does not in itself guarantee useful inference, as different methods can be derived this way depending on the model chosen. So one has to ask the question of which method can be applied to the data at hand.

The price paid for applying the methods of maximum likelihood is that the assumptions behind the results have to be satisfied. The assumption that we focus upon here is the fundamental one which is sometimes forgotten:

Assumption : *In order to apply the maximum likelihood methodology to some data* $x_1, \ldots, x_T$ *we need to assume that the density* $f(x_1, \ldots, x_T)$ *of* *the data is a member of the statistical family to which we apply the methodology. This means that there is some parameters value θ_0, the true value, for which the density $p(x_1, \ldots, x_T, \theta_0)$ is the same as the density of the data $f(x_1, \ldots, x_T)$.*

Obviously the density of the data in Figure 1 corresponds to the parameter values $\rho = 0.9, \mu = 2, \sigma = 1$ in the AR(1) model, and hence the basic assumption is satisfied for Method 2, but there are no parameters (μ, σ^2) from the first model that correspond to the data density. Thus applying Method 2, we can use all the results of maximum likelihood, but with Method 1 we cannot rely on the standard error of the estimate or the confidence interval when applied to the data in 1. In the process of checking the Assumption above we should be able to show that the model chosen for the first maximum likelihood analysis does not satisfy the assumption that for some values of the parameters the model captures the density generating the data. More precisely, the residuals $\hat{\varepsilon}_t = x_t - \bar{x}$ based upon Method 1 will surely show sign of autocorrelation, so that Model 1 cannot be used for inference. In fact we find an estimate of the correlation between $\hat{\varepsilon}_t$ and $\hat{\varepsilon}_{t-1}$ to be 0.92, which is clearly different from zero, which is the value assumed by the first model, could be used for inference.

Thus we are warned that the mathematical results of the likelihood based methods are no longer valid, and we shall have to derive the properties of the estimator found as maximum likelihood estimator, from first principles. This is clearly impossible without first establishing how to describe the density of the data. The properties of the estimators are not automatically given by the theory behind the maximum likelihood estimator if the fundamental assumption is not satisfied.

Thus there is nothing wrong with the two models analyzed, they are both standard models. What is wrong, in view of the data we want to analyse, is that only the second model describes the data, the first does not. Hence application of the asymptotic results of maximum likelihood of the second method is valid and application of the asymptotic results for the first is not, even though both are simple applications of Gaussian maximum likelihood.

3.2 What to do?

In practice there is obviously a serious problem with the method of maximum likelihood, because we can of course never know with certainty if we have the right model, and the basic Assumption is difficult, or even impossible, to check. But we can sometimes show that *it is not satisfied*, and this is a very useful piece of information because it implies that *we cannot use the model considered for conducting inference.*

What we have to do in this case is of course to find a new and better statistical model by specifying a different and more flexible family of densities, which has a better chance of capturing the density of the data. Only then can we conduct inference on the validity of the economic model that is of interest. More constructively we should ask the question

"Which statistical model describes the data."

If we have a classical economic model, like the one analyzed by Ireland, it looks like a waste of time to spend too much effort on the above question, which is clearly a statistical question. But that depends on the purpose of the analysis.

If the idea is to test hypotheses on the parameters of the economic model, then we need to establish a sound statistical basis for doing so, that is, we need to embed the economic model in a statistical model that describes the data. This is then used as a platform for making inference on coefficients and relations. In case we find that a model chosen does not describe the data, the explanation can be that the economic model is inadequate, that the error structure we have modelled is not flexible enough, and finally of course that the data quality is not good enough, but the model cannot be used for inference.

If the purpose of the analysis, however, is not to make inference on coefficients or model equations, but to keep an economic model that we believe firmly in with the purpose of making predictions, say, then perhaps it is enough to pick some likelihood method to get some estimates. But then one cannot count on consistency of the estimators, let alone their standard errors derived from the likelihood theory. We can only believe that such results are valid, if we have carefully checked the model for misspecification.

If we are in the first situation, that we want to learn from the data about our model, we have to check the many assumptions behind the likelihood methods. Thus, when we have suggested a statistical model, we should always ask the question

"How can we prove that this model is incorrect?"

Having tried in many different ways to reject the model and not been able to do so, we can be more confident in applying the general theory of likelihood. For instance, with the above time series, we would try to determine the lag length, either by information criteria or by testing significance of autocorrelations and partial autocorrelations, we would try to check the distribution of residuals using histograms, test for ARCH effects, constant parameters etc. If the model satisfies these criteria reasonably well, we can apply the theory of likelihood based inference with some confidence, still running the risk, of course, that someone else will analyze the data, and show that the model we have chosen is incorrect, and hence that the conclusions we have reached may need to be modified.

In this methodology there is the implicit hope, that if a model passes all the misspecification tests reasonably well, then the conclusions, which build on likelihood methods, will also hold reasonably well.

3.3 Suggestions for checking the model analyzed by Ireland

In the paper by Ireland the statistical model is used for inference without seriously checking the validity of the model for the analysis of the given data. This means that we can have no confidence what-so-ever that the standard errors are valid, see the stylized examples above under Methods 1 and 2. We give here a list of possible ways of checking for model misspecification, some of which are related to the economic model and some to the embedding statistical model.

1. By first fitting a model with a general 5×5 matrix in the state equation, one can test that is has the form

$$\begin{pmatrix} \mathcal{A} & 0 \\ 0 & D \end{pmatrix},$$

suggested by the theory. Then one can test whether the specific parametric form given for $\mathcal{A}$, see the Ireland (2004*b*), is a good description of what goes on. One could also test for the diagonality of D, as has been assumed by some authors.

2. One can test the zeros and singularity of the error variance matrix, by comparing a model with a general 5×5 matrix Σ with the model where the

matrix has the form

$$\begin{pmatrix} \sigma^2 \mathcal{B}\mathcal{B}' & 0 \\ 0 & V \end{pmatrix}.$$

Thus the idea that capital and technology is driven by the same shock could be checked against the data.

3. The lag length is chosen to one. One could test whether this lag length is the right one. A lag length of two for u_t, say, would involve a state variable of dimension eight of the form $(s_t', u_t', u_{t-1}')'$, so that the same methodology could be applied.

4. One of the basic assumptions in the economic model, and also in the statistical model, is that of parameter constancy for the full period of 50 years. One should therefore perform a recursive analysis to check for parameter stability, or a subsample analysis. This was done in the paper by Ireland (2004*a*), where it was shown, under the assumption that the model is well chosen for each subperiod, that preferences before and after 1980 were not the same. This should have consequences for the economic model.

5. A main assumption behind the derivations of the equations which are then made more elastic is the assumption of trend stationarity of the logarithmic variables. In view of the very large root found: $\hat{\rho} = 0.9987$, this assumption appears not to be very useful, see the next section

4 Do unit roots matter?

It was found empirically that $\hat{\rho} = 0.9987$, that the maximal eigenvalue of $\mathcal{D}$ was 0.9399 and finally $1 - \delta$ was set to 0.975. Still it is maintained that the processes are stationary (highly persistent) around a trend in order to define the steady state values. The assumption of stationarity implies that we have standard asymptotic χ^2 inference for the estimated coefficients. It is, however, to be expected that with such large roots the asymptotic results offer poor approximations to the finite sample distribution, which is needed for making reliable inference. This can be illustrated by a simple simulation experiment, where we choose a bivariate $VAR(1)$ model with a constant term included, so that the process, under stationarity assumptions, has a mean. We want to show that inference on this mean is problematic in the presence of large roots.

4.1 Simulations

The equations determining the bivariate process x_t are

$$\Delta x_t = \Pi(x_{t-1} - \mu) + \varepsilon_t, t = 1, \ldots, T \tag{8}$$

where ε_t are i.i.d. $N_2(0, \Omega)$. We use the notation

$$\begin{aligned}
\bar{x}_- &= T^{-1}\sum_{t=1}^{T} x_{t-1}, \overline{\Delta x} = T^{-1}\sum_{t=1}^{T}\Delta x_t = T^{-1}(x_T - x_0) \\
S_{11} &= T^{-1}\sum_{t=1}^{T}(x_{t-1} - \bar{x}_-)(x_{t-1} - \bar{x}_-)' \\
S_{10} &= T^{-1}\sum_{t=1}^{T}(x_{t-1} - \bar{x}_-)(\Delta x_t - \overline{\Delta x})' \\
S_{00} &= T^{-1}\sum_{t=1}^{T}(\Delta x_t - \overline{\Delta x})(\Delta x_t - \overline{\Delta x})'
\end{aligned}$$

and find the maximum likelihood estimators

$$\hat{\Pi} = S_{01}S_{11}^{-1}, \tag{9}$$

$$\hat{\mu} = \bar{x}_- - \hat{\Pi}^{-1}\overline{\Delta x} \tag{10}$$

$$\hat{\Omega} = S_{00} - S_{01}S_{11}^{-1}S_{10}$$

Under the assumption of stationarity, the asymptotic distribution of $\hat{\mu}$ is given by

$$T^{1/2}(\hat{\mu} - \mu) \xrightarrow{w} N_2(0, \Pi^{-1}\Omega\Pi'^{-1}).$$

The problem we want to discuss is how inference can be distorted when there is a root close to one. The roots are determined from $\det((1-z)I_2 - \Pi z) = 0$, so that $z^{-1} - 1$ is an eigenvalue of Π. We see immediately, that as the root approaches one, the matrix Π becomes near singular with the result that the variance blows up and there is absolutely no information about the mean of the two processes. Not only does the variance blow up, the finite sample distribution deviates seriously from the asymptotic Gaussian distribution, so that inference of the type

$$\hat{\mu}_1 \pm 1.96\sqrt{asVar(\hat{\mu}_1)}$$

Rejection probability in % for test of $\mu_1 = \mu_2 = 0$			
$\delta \backslash T$	50	100	500
-0.5	8.3	6.3	5.5
-0.1	14.3	10.5	6.2
-0.01	21.4	19.0	13.2
-0.001	21.3	20.9	20.1

Table 1: In the DGP (11) we simulate the rejection probability of the likelihood ratio test statistic for the hypothesis $\mu_1 = \mu_2 = 0$, for various values of δ and T using the quantile $\chi^2(2) = 5.99$. The results are based on 10,000 simulations.

becomes highly misleading and tests on μ_1 become unreliable. In order to illustrate this we choose as simulation experiment a DGP of the form

$$\begin{aligned} \Delta x_{1t} &= -0.5x_{1t-1} + (0.5+\delta)x_{2t-1} + \varepsilon_{1t} \\ \Delta x_{2t} &= 0.5x_{1t-1} - 0.5x_{2t-1} + \varepsilon_{2t} \end{aligned} \tag{11}$$

that is, we specify Π as given, take $\Omega = I_2$, and choose $\mu = 0$. We do 10,000 simulations for $\delta = -0.5, -0.1, -0.01, -0.001$, corresponding to the largest root being 0.5, 0.954, 0.995 and 0.9995. We choose $T = 50, 100$, and 500. As a statistical method we choose Gaussian maximum likelihood inference based on model (8), which in this case happens to be the correct one because the basic Assumption is satisfied. The estimators are calculated by (9) and (10).

We first simulate the distribution of the estimator of the first component of $\hat{\mu}$. Figure 2 shows some QQ plots against the Gaussian distribution scaled by the asymptotic standard deviation under the assumption of stationarity for $T = 50$. Note that for $\delta = -0.5$, the asymptotic distribution describes the finite sample distribution quite well, but for smaller δ the distribution is very distorted, and inference, assuming it is Gaussian, is highly misleading.

In Table 1 we calculate the rejection probabilities of the likelihood ratio test that $\mu_1 = \mu_2 = 0$, using the asymptotic critical value of $\chi^2_{0.95}(2) = 5.99$. We see that even 500 observations are not enough for getting anywhere near a 5% test when δ is -0.01 or -0.001. Thus one can insist that a root of 0.995 is a highly persistent stationary root, but the price paid is that one cannot make inference on the steady state values using usual sample sizes.

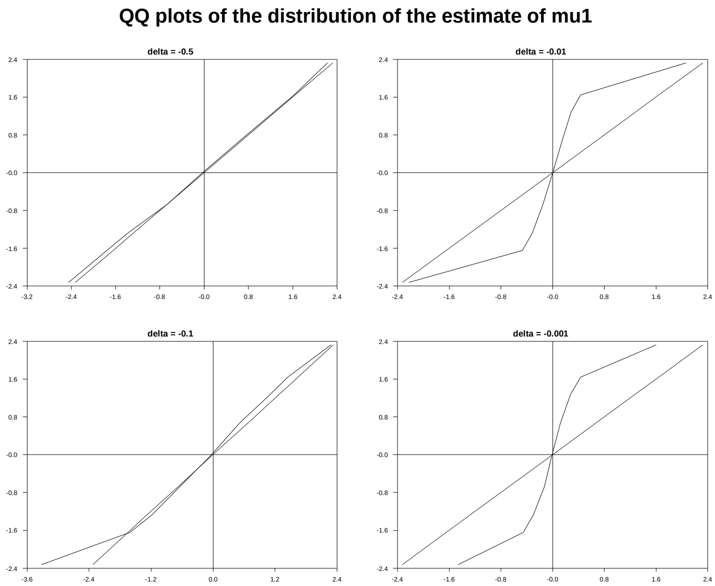

Figure 2: In the DGP (11) we simulate the distribution of the estimator of μ_1 for various values of δ and $T = 50$. The Gaussian distribution is scaled with the asymptotic standard error under the assumption of stationarity.

4.2 Imposing a unit root

The limiting covariance $\lim_{T\to\infty} Var(T^{-1/2}\sum_{t=1}^{T} x_t) = \Pi^{-1}\Omega\Pi'^{-1}$, the so-called long-run covariance, is singular in the limit of $\delta \to 0$, corresponding to the fact that for a process with a root of 0.9987, there is hardly any information in the data on the mean of the process. But the singularity on the other hand hides a very simple result, namely that for the DGP chosen the variance of $\hat{\mu}_1 - \hat{\mu}_2$ is relatively small and inference can be reliably conducted for $\mu_1 - \mu_2$. Thus by investigating the singularity of the long run variance we can get information about the linear combination of μ_1 and μ_2 for which we can make reliable inference. The estimated long-run variance is

$$\hat{\Pi}^{-1}\hat{\Omega}\hat{\Pi}'^{-1} = S_{11}S_{01}^{-1}(S_{00} - S_{01}S_{11}^{-1}S_{10})S_{10}^{-1}S_{11}.$$

If we solve the eigenvalue problem

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0, \tag{12}$$

for eigenvalues $\hat{\Lambda} = diag(\hat{\lambda}_1, \hat{\lambda}_2)$ with $1 \geq \hat{\lambda}_1 \geq \hat{\lambda}_2 > 0$, and eigenvectors $\hat{V} = (\hat{v}_1, \hat{v}_2)$ so that

$$\hat{V}'S_{10}S_{00}^{-1}S_{01}\hat{V} = \hat{\Lambda},\ \hat{V}'S_{11}\hat{V} = I_2$$

then

$$\hat{V}'\hat{\Pi}^{-1}\hat{\Omega}\hat{\Pi}'^{-1}\hat{V} = \hat{\Lambda}^{-1} - I_2,$$

which shows that the singularity of the long-run variance corresponds to a near zero eigenvalue. The test for a zero eigenvalue in (12) is exactly the test for cointegrating rank one.

If we find that a unit root is acceptable by the data, then x_t is nonstationary, Δx_t is stationary, that $\Pi = \alpha\beta'$ and $\beta' x_t$ is stationary. In this model the mean of the process x_t is not well defined, but the stationary process $\beta' x_t$ has a mean $\beta'\mu$, for which we can test the hypothesis $\beta'\mu = 0$. The asymptotic distribution is $\chi^2(1)$, provided the data is actually a cointegrated process. Now suppose the root is as indicated above, so that the processes are stationary. What happens if we never-the-less impose the unit root and perform the test in a cointegrated model. We evidently find that the basic Assumption is not contradicted by the analysis, and we find the results in Table 2, which show that although we have generated the data by a stationary (highly persistent) process the analysis by a simple cointegration analysis

Rejection probability in % for test of $\beta'\mu = 0$			
$\delta \backslash T$	50	100	500
-0.5	7.9	6.3	5.4
-0.1	7.3	6.0	4.9
-0.01	7.1	5.7	5.5
-0.001	7.4	6.0	5.0

Table 2: In the DGP (11) we simulate the rejection probability of the likelihood ratio test statistic for the hypothesis $\beta'\mu = 0$ in the cointegrated model $\Delta x_t = \alpha\beta'(x_{t-1} - \mu) + \varepsilon_t$ for various values of δ and T using the quantile $\chi^2(1) = 3.84$. The results are based on 10,000 simulations.

will give reliable inference on the parameter $\beta'\mu$, which is the only aspect of $\beta'\mu$ that is identified in the cointegrated model.

Thus cointegration can be seen as a way of investigating which aspects of the mean can be reliably discussed in the case of a large root.

Finally we address briefly the question of what happens with the economic model if we leave the assumption of trend stationarity. Going through the derivations one can see that $Y_t = I_t + C_t$ is consistent with $y_t - c_t$ and $y_t - i_t$ being stationary. The relation $K_{t+1} = (1 - \delta)K_t + I_t$ is consistent with Δk_t and $k_t - i_t$ being stationary and finally $\gamma C_t H_t = (1 - \theta)Y_t$ is consistent with h_t being stationary. Finally the production function can be taken as the definition of a_t and implies that a_t is non-stationary and cointegrates with detrended k_t :

$$a_t - (1 - \theta)(k_t - t \log \eta) = y_t - k_t - (1 - \theta)h_t.$$

Thus the root 0.9987 is replaced by 1, but all the important relations between the variables have an interpretation in terms of cointegrating relations, see Ireland (2001). Note that imposing the unit root $\rho = 1$, also imposes a number of known cointegrating relations, so that a test for $\rho = 1$ is simultaneously a test that these cointegrating relations are in accordance with the data.

5 Conclusion

We have not questioned the usefulness of Gaussian maximum likelihood inference as suggested by Ireland, but we claim that the mathematical assumptions behind the asymptotic theory of the Gaussian likelihood methods requires at least as much thought and careful analysis as the mathematics behind the derivation of the first order conditions of the economic model. Only when it has been convincingly demonstrated that the assumptions behind the Gaussian likelihood method are not too far from satisfied, can we rely on the inference conducted.

Of special interest is the near unit roots found in the data. If one insists that such roots are stationary, one may need a lot of data to conduct inference, using asymptotic methods. By imposing the unit roots we get a cointegration model that fits equally well, and the economic relations have to be reinterpreted as cointegrating relations. Imposing the unit roots can be seen as a way of finding those relations between the steady state value on which we can conduct inference. We find that the statistical assumption of stationarity around a linear trend, which is part of the economic model formulation, is replaced by the assumption of stationarity around a stochastic and linear trend.

6 References

Haavelmo, T. (1943). Statistical implications of a system of simultaneous equations, *Econometrica*, 11, 1-12.

Ireland, P. (2001), Technology shocks and the business cycle: An empirical investigation, *Journal of Economic Dynamics and Control 25, 703-719.*

Ireland, P. (2004*a*), A method for taking models to the data, *Journal of Economic Dynamics and Control 28, 1205-1226.*

Ireland, P. (2004*b*), Supplementary material: http://www2.bc.edu/~irelandp.