Rochester Center for
Economic Research

Four Econometric Fashions and the Kalman Filter Alternative

Bomhoff, Eduard J.

Working Paper No. 324
June 1992

University of
Rochester

Four Econometric Fashions and the Kalman Filter Alternative

Eduard J. Bomhoff

Rochester Center for Economic Research
Working Paper No. 324

Four econometric fashions and the Kalman Filter alternative

Eduard J. Bomhoff
Erasmus University
P.O. Box 1738
3000 DR Rotterdam
The Netherlands
tel.: 010-4081408
fax: 010-4527357

1. Introduction[1]

Causality tests, vector autoregression, unit root tests, cointegration: four of the prominent econometric advances of the past 15 years. These techniques have been thoroughly, some would say over-actively applied to virtually all macroeconomic and financial questions. All four are derivatives of ordinary least squares (ols) with the consequence that they share a characteristic of the ols technique that is always mentioned in the first chapter of textbooks on regression, but often neglected in practice: the natural context for ols and its applications is that of stationary time series, and severe problems of interpretation arise when ols is used for nonstationary data.

If time series are stationary, their first and second moments are well defined and there is no conceptual problem in computing unconditional means, variances and covariances based on observations over some sample period. By contrast, nonstationarity implies that the "mean" of a sample may become a function of its length whilst the mean of the true series - the population mean - may not be defined. Also, although sample variances and covariances can of course be computed, they cannot possibly provide information about true unconditional second moments, since these are undefined. It follows that in the case of nonstationary time series extrapolations or forecasts should only be made in a conditional manner: one can produce forecasts of X_{t+1} conditional on the realized process up to observation X_t, but such forecasts are no more than conditional extrapolations instead of the unconditional forecasts correctly produced by ols techniques in the context of stationary time series.

In this paper, based on materials from my (1993) book, I use Monte Carlo techniques to illustrate some properties of causality tests, vector autoregressions, unit root tests and cointegration techniques for stationary, (almost) stationary and clearly nonstationary time series and contrast the findings with calculations performed by a multivariate Kalman

[1] Camiel de Koning contributed substantially to the development of the Kalman filter software. Erzo Luttmer and Patrick Groenendijk very competently organized and executed the calculations. I am very grateful to them and also to René den Hertog for all their help.

filter. Because the Kalman filter methodology processes observations sequentially, either forward, backward or both, it produces conditional distributions for means and variances and therefore lends itself in a much more natural manner to analysis of nonstationary series. I shall also discuss the common response within the ols context to nonstationarity - whether certain or suspected - which is to difference the data. The results will show that the Kalman filter does not require the user to make a definite decision regarding the need for differencing the data and, if so, once or twice, but instead offers automatic processing capacity for a wide class of nonstationary time series.

In addition to the analysis of near-stationarity and nonstationarity, I consider a second issue in time series analysis with important consequences for the interpretation of ols-based statistical techniques. Traditional time series analysis as exemplified by the Box and Jenkins (1970) book has considered any single time series as being driven by a simple, serially uncorrelated white noise process. Autoregressive and/or moving average coefficients in the time series model describe any desired pattern for the persistence of these simple innovations. Roots of the autoregressive part of the model that lie on the unit circle represent the desired degree of nonstationarity. By contrast, the Kalman filter methodology would look for a simpler structure in modelling a univariate time series, but make it subject to different types of shocks. The Kalman filter models in this paper will consider three types of disturbances that effect a time series continuously:

(1) temporary shocks to the level of the series;

(2) permanent shocks to the level of the series;

(3) permanent changes in the stochastic rate of growth of the series.

Formally, the two different descriptions, either a single type of innovation but a complicated autoregressive and/or moving average structure, or a simpler model driven by a variety of different shocks, are equivalent. The Kalman filter representation has important technical advantages over ARIMA modelling in the sense that parameters will always lie within the admissible region, but on the other hand Kalman filter estimation of the underlying hyperparameters - the variances of the different types of shocks that hit the system - is more complicated than the estimation of a simple autoregressive model. More important than these technical considerations, however, are the differences in interpretation that become crucial when the model is used for simulations. To take a very simple example, assume that a time series can be modelled in the ARIMA methodology as follows:

$$(1) \quad \Delta y_t = (1-0.5L)\, a_t$$

where Δ denotes the difference operator, L denotes the lag operator and a_t is normally distributed and serially independent, with mean zero and variance one.

The corresponding state-space model would be:

$$(2A) \quad y_t = c_t + \upsilon_t$$

$$(2B) \quad c_t = c_{t-1} + \omega_{1t}$$

with c_t an unobservable state variable. The errors υ_t and ω_{1t} are uncorrelated with mean zero and variances σ^2_υ and $\sigma^2_{\omega 1}$ respectively, and are serially independent. The state variable is not correlated with these errors. The state-space model that corresponds exactly to equation (1) requires values of 0.5 for σ^2_υ and 0.25 for $\sigma^2_{\omega 1}$.[2]

The natural context for a simulation exercise would be to assume that no shocks have taken place for some time and then to compute the consequences of a single innovation of size one at time $t=1$, that is, $a_1 = 1$. In the Box-Jenkins methodology that produces the following results:

---

[2] The variances of the disturbances in the state space model are computed by equalizing the complete autocovariance function for the first difference of y, both for the Box-Jenkins model and in in the state space formulation.

time	-2	-1	0	1	2	3	4	...
Δy_t	0	0	0	1	-0.5	0	0	...
y_t	0	0	0	1	0.5	0.5	0.5	...

The user of a Kalman filter model would need no reminding that each simulation exercise requires an answer to the question whether the shock or shocks that affect the system should be of a temporary or permanent nature or whether they should combine the different types of fundamental disturbances in the same proportions as observed during some historical period. Hence, the Kalman filter methodology is potentially much closer to addressing the concerns of the so-called "Lucas critique". Assume that one wants to perform a simulation exercise that corresponds exactly to studying the consequences of a single innovation in the Box-Jenkins model. Now an apparent contradiction appears. On one hand the assumed value of the moving average parameter in the Box-Jenkins model of equation (1) equals 0.5, implying that the innovation should be regarded as being fifty percent temporary, fifty percent permanent. On the other hand, performing the same exercise with the state-space model requires us to apply a mixture of the two basic disturbances in which the temporary component carries much greater weight than the permanent shock.[3] How can this be?

We shall see that solution of the apparent paradox requires us to be much more precise when specifying the thought experiment to which the simulation exercise is supposed to provide the answers. Specifically, we shall find that the apparently natural context for simulations, which is to assume that no significant shocks have taken place immediately before the single innovation under study, is logically incorrect given the estimated parameters of both the Box-Jenkins time series model and the state-space model. Also, analysis of the state-space model will lead to the conclusion that the size of the innovation in the thought experiment becomes important when trying to compute the outcomes of the simulation.

---

[3] Conversely, if in the state space model the variances for the temporary and permanent components are equal, indicating that temporary and permanent shocks are equally important, the parameter in the corresponding (0,1,1) Box-Jenkins model would be equal to $1.5 - 0.5\sqrt{5} = 0.38$, instead of 0.5 as one might naively expect.

The subsequent sections of the paper will contain the description and results of a variety of Monte Carlo experiments. In section 2 I describe a first set of experiments in which I use one random walk type variable to predict a related random walk. By way of illustration, I show a single representative realization before tabulating the results of the complete set of replications for which every experiment is repeated 100 times. The alternative Kalman filter methodology is described in section 3. Section 4 then contains outcomes for the simulated data when both ols and the Kalman filter model are applied. In section 5 I describe a second series of experiments based on data that exhibit medium term cyclical fluctuations. Once again, in each experiment two series are paired and I investigate whether the "X" series can be useful in predicting the "Y" series and vice versa. The description again provides a selected representative illustration. Tables for the outcomes of ols techniques, for each series of 100 replications are in section 6 of the paper together with Kalman filter computations for the same set of data. Section 7 investigates how useful the four econometric fashions that have been so influential over the past 15 years prove to be when applied to these artificial data. Finally, section 8 summarizes the paper and draws some conclusions.

2. Experiments with artificial random walks

Ordinary least squares (ols) and its more sophisticated derivatives continue to be the dominant statistical methodology in econometrics. To work well and generate results that can be unambiguously interpreted, ols requires stationary data. Hence, the great majority of simulation studies that have tried to investigate what happens when regression models are applied in practice, have used stationary time series or series that become unambiguously stationary after taking first differences, or series that were stationary around a deterministic trend. In this paper, by contrast, all experiments are based on nonstationary series, or on series where the analyst is uncertain about the stationarity issue. Differences between X and Y will be stationary or will contain a random walk component as well as a small stochastic trend.

The first set of experiments works with constructed "random walks with stochastic trends" i.e. series that do not have a fixed mean but are likely to reach ever larger distances from the origin. I have used a series length of 100 data points, equivalent to 25 years of quarterly observations or about 8 years of monthly data.[4] Here follows the law of motion for the first series, to be called Y.

$$(3A) \quad y_t = c_t + \upsilon_t$$

$$(3B) \quad c_t = c_{t-1} + tr_{t-1} + \omega_{1t} + \omega_{2t}$$

$$(3C) \quad tr_t = tr_{t-1} + \omega_{2t},$$

where the errors υ_t, ω_{1t} and ω_{2t} are mutually uncorrelated with mean zero and variances σ^2_{υ}, $\sigma^2_{\omega 1}$ and $\sigma^2_{\omega 2}$ respectively.

With the specific values of 1 for σ^2_{υ}, 1 for $\sigma^2_{\omega 1}$ and 0.1 for $\sigma^2_{\omega 2}$, equation (3) can be written in state-space notation as:[5]

$$(4A) \quad y_t = [1 \quad 0] \begin{pmatrix} c_t \\ tr_t \end{pmatrix} + \upsilon_t \quad , \quad var(\upsilon_t) = 1$$

$$(4B) \quad \begin{pmatrix} c_t \\ tr_t \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} c_{t-1} \\ tr_{t-1} \end{pmatrix} + \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \omega_{1t} \\ \omega_{2t} \end{pmatrix} \quad , \quad var \begin{pmatrix} \omega_{1t} \\ \omega_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0.1 \end{pmatrix}$$

I use two different specifications for the second series, to be called X. In a first set of experiments, each realization of Y is paired with a realization of X that is constructed as follows:

[4] 120 Observations are generated, so that the first 20 can be discarded in order to avoid transient effects of the initialization of the autoregressive model components.

[5] The next section has more on state space models.

1. subtract the temporary component from Y. This leaves a pure random walk with stochastic trend in which all increments in the series are 100% percent permanent;

2. add a high-frequency autocorrelated disturbance to this stripped-down Y series, as follows:

(5) $\upsilon_t = \theta_1\upsilon_{t-1} + \theta_2\upsilon_{t-2} + \varepsilon_t$

Combining equation (4) and (5), with $\theta_1 = 0$, $\theta_2 = -0.75$ and $var(\varepsilon_t) = 10$, the state-space model for X_t becomes:[6]

$$(6A) \quad X_t = [1\ 0\ 1\ 0] \begin{pmatrix} c_t \\ tr_t \\ \upsilon_t \\ \upsilon_{t-1} \end{pmatrix}$$

$$(6B) \quad \begin{pmatrix} c_t \\ tr_t \\ \upsilon_t \\ \upsilon_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -0.75 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} c_{t-1} \\ tr_{t-1} \\ \upsilon_{t-1} \\ \upsilon_{t-2} \end{pmatrix} + \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \omega_{1t} \\ \omega_{2t} \\ \varepsilon_t \\ 0 \end{pmatrix} , \quad var \begin{pmatrix} \omega_{1t} \\ \omega_{2t} \\ \varepsilon_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$

In this first set of experiments, each pair of an Y and a X have the same underlying dynamics, as represented by c, tr, ω_1 and ω_2. But, in the case of Y there is a simple serially uncorrelated noise added before Y is observed by the econometrician; in the case of X there is no such "output or observation noise", but the underlying process of X is contaminated by the high-frequency "vibration". Because the amplitude of the added high-frequency disturbance in X is large, X loses some of its usefulness in making forecasts of Y. In short:

{X} - {serially correlated stationary disturbance} + {observation noise} = {Y}.

[6] Having the current residual as a state variable results in an error-free observation equation. Maybeck (1979) shows how the usual Kalman filter algorithm can be adjusted for this case.

I use the following context to perform statistical analysis of the relationship between Y and X:

1. observations of X are available for the past, the present and the future;

2. the econometrician is not certain that all deviations between Y and X are temporary, even though we know that such is the case.

Our interest now is in investigating standard statistical techniques for relating Y to X and later to compare the results to those obtained with the multivariate Kalman filter.

For a second series of experiments, X remains as before, but is paired with an Y that is more loosely connected to X, because differences between Y and X are no longer temporary and stationary as before, but increase slowly over time as follows:

$$(7A) \qquad y_t = c_t + d_t + \upsilon_t$$

$$(7B) \qquad d_t = d_{t-1} + tr_{help,t-1} + \omega_{3t} + \omega_{4t}$$

$$(7C) \qquad tr_{help,t} = tr_{help,t-1} + \omega_{4t}$$

with $var(\omega_3) = 0.05$, $var(\omega_4) = 0.005$ and c_t as in equation (3B) above.

In this case, we can summarize the relationship between X and Y as follows:

{X} - {serially correlated stationary disturbance} + {random walk with stochastic trend} + {observation noise} = {Y}.

Figure 1 shows a typical realization. For this second series of statistical experiments I make the same two assumptions as before:

1. X is available to be used in predicting Y;

2. the analyst is not certain about the longer-term connections between Y and X and has to derive from his statistical analysis whether differences between Y and X are well behaved and stationary, or whether Y and X may diverge ever further as time goes on.

As before, I shall investigate how standard econometric techniques and the Kalman filter alternative compare in terms of being useful for tests of the hypothesis that the long-run elasticity between X and Y is equal to 1.0.

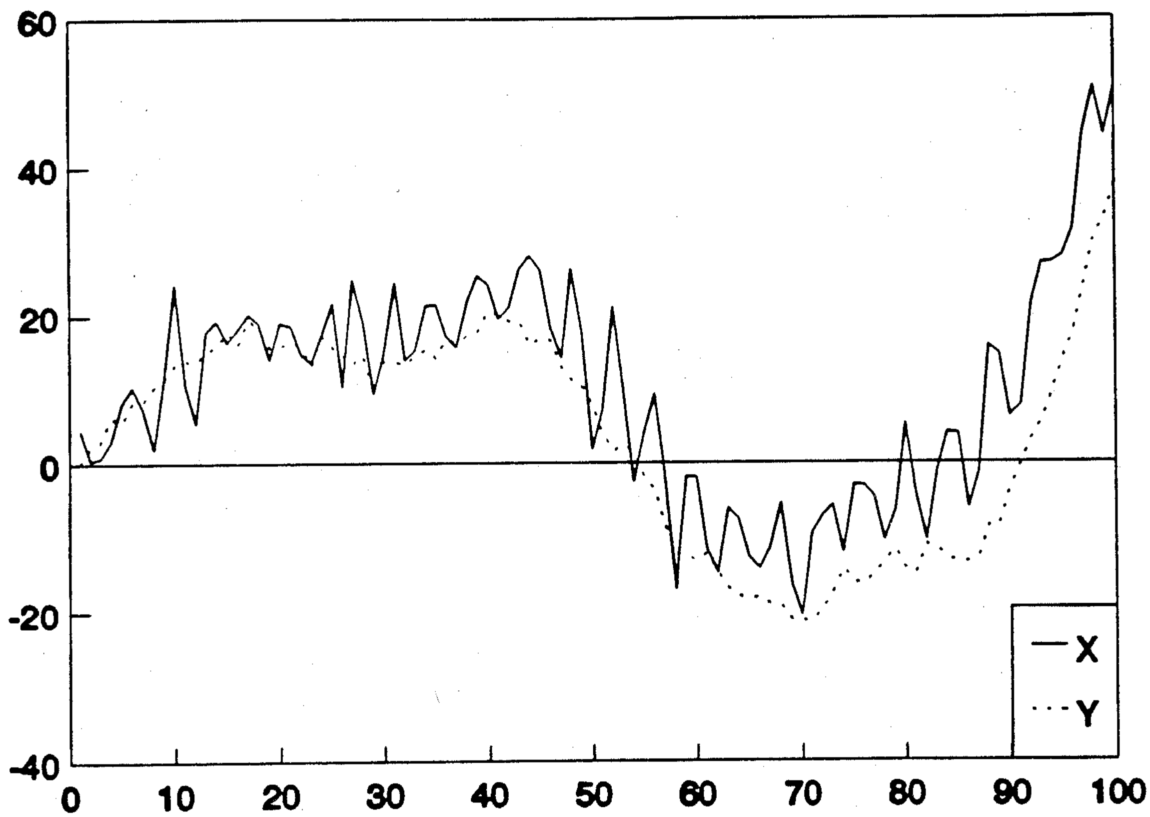

Figure 1: Two non-cointegrated nonstationary series. The ols results from a levels regression of Y on X were: DW = 0.861, R^2 = 0.778, b = 0.871 with S.E. = 0.047, LBQ(29) = 492.71, ADF(1) = -5.16, ADF(4) = -1.096 and ADF(8) = -1.073. For abbreviations see table 1 below.

3. A multivariate Kalman filter

One way to embed any least squares equation in a richer dynamic model is to change to the state-space formulation. The state vector is composed of the regression intercept, all regression coefficients α plus as many current and lagged residuals as are required to represent any significant serial correlation in the error term. The general specification of the multivariate Kalman filter in the case of no serial correlation in the residuals and a single regression coefficient α is as follows:

$$(8A)\qquad y_t = \begin{pmatrix} 1 & 0 & X_t \end{pmatrix} \begin{pmatrix} c \\ tr \\ \alpha \end{pmatrix}_t + \upsilon_t$$

$$var(\upsilon) = R$$

$$(8B)\qquad \begin{pmatrix} c \\ tr \\ \alpha \end{pmatrix}_t = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c \\ tr \\ \alpha \end{pmatrix}_{t-1} + \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix}_t$$

$$var \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \begin{pmatrix} Q_1 & 0 & 0 \\ 0 & Q_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Equation (8A) is the observation equation. It states that the level of y equals the sum of a shift parameter, c_t, the product of the regression coefficient and the explanatory variable, and a residual term υ_t.

The Kalman filter methodology adds equation (8B), the so-called state update equation. It shows how the three state variables change from period to period. The right hand side of this equation has two parts. In the first part, the shift parameter c_{t-1} is adjusted upwards by the amount tr_{t-1}, which represents the stochastic trend. In the second part, the shift parameter c_t is subject to permanent shocks ω_{1t} and ω_{2t} before being included in the observation equation. The trend is subject to a stochastic shock, ω_{2t}, whilst the regression coefficient, α, is not subject to change over time.

The Kalman filter allows for nonstationarity in y_t. In fact, Kalman filter theory arose because of the inadequacy of the Wiener-Kolmogorov theory for coping with applications in which nonstationarity of the signal and/or the noise was essential to the most natural description the model. Since it operates in the time domain, the Kalman filter can provide densities of the state variables conditional on the history of the system. This is the essential advantage of Kalman filtering over frequency domain methods when nonstationarity is a relevant feature of the data.

The user of a Kalman filter is asked to provide estimates of the variances Q_1, Q_2, and R.[7] One also needs to postulate whether the filter starts estimation with a diffuse prior or with some notion about the range of the parameters and/or the state variables. The Kalman filter then processes the data "on line" and produces estimates of the state variables - here: the shift parameter, the trend and the regression coefficient - and their variance-covariance matrix, P_t. The variances, Q_1, Q_2, and R may be chosen in such a way that the specification becomes equivalent to either a regression equation between Y and X in levels or in first or second differences. The Kalman filter specification includes all three possibilities as special cases. Other statistical techniques for comparing levels and first or second difference specifications suffer from the disadvantage that the two competing hypotheses are non-nested.[8]

A normal ("forward") Kalman filter produces an estimate of the state variables at time T+1 (in our case: the shift parameter, the trend and the regression coefficient α, together a three-element vector V) based on all the data from time t=1 up to and including time T=t. The forward filter revises its estimate of the state variables for period T after it has observed the dependent variable (in our case: the level of Y) for that period. Denote the forward forecast for period T+1 based on $\{V_1, V_2, ..., V_T; X_1, X_2, ..., X_T, X_{T+1}\}$ as $V_f(T+1|T)$. $V_f(T+1|T+1)$ will refer to the revised estimate of V_{T+1} after the current level of Y has been observed. In each iteration, a backward filter is also implemented. It generates a backward "forecast" for time T using all the data from period T+1 through to the final period, j.

---

[7] In the next section, the EM algorithm will allow for continuous updates of these variance estimates. One also has to insert an arbitrary non-zero value for the variance of the third element in the vector ω. This is purely for computational reasons (see Maybeck, 1979) and has no effects on the estimates.

[8] See Nelson and Plosser (1982) for discussion of traditional econometric tests of the levels versus first differences specification. Such tests have low power. See also Christiano and Eichenbaum (1990) for similar conclusions regarding the univariate analysis of US gnp.

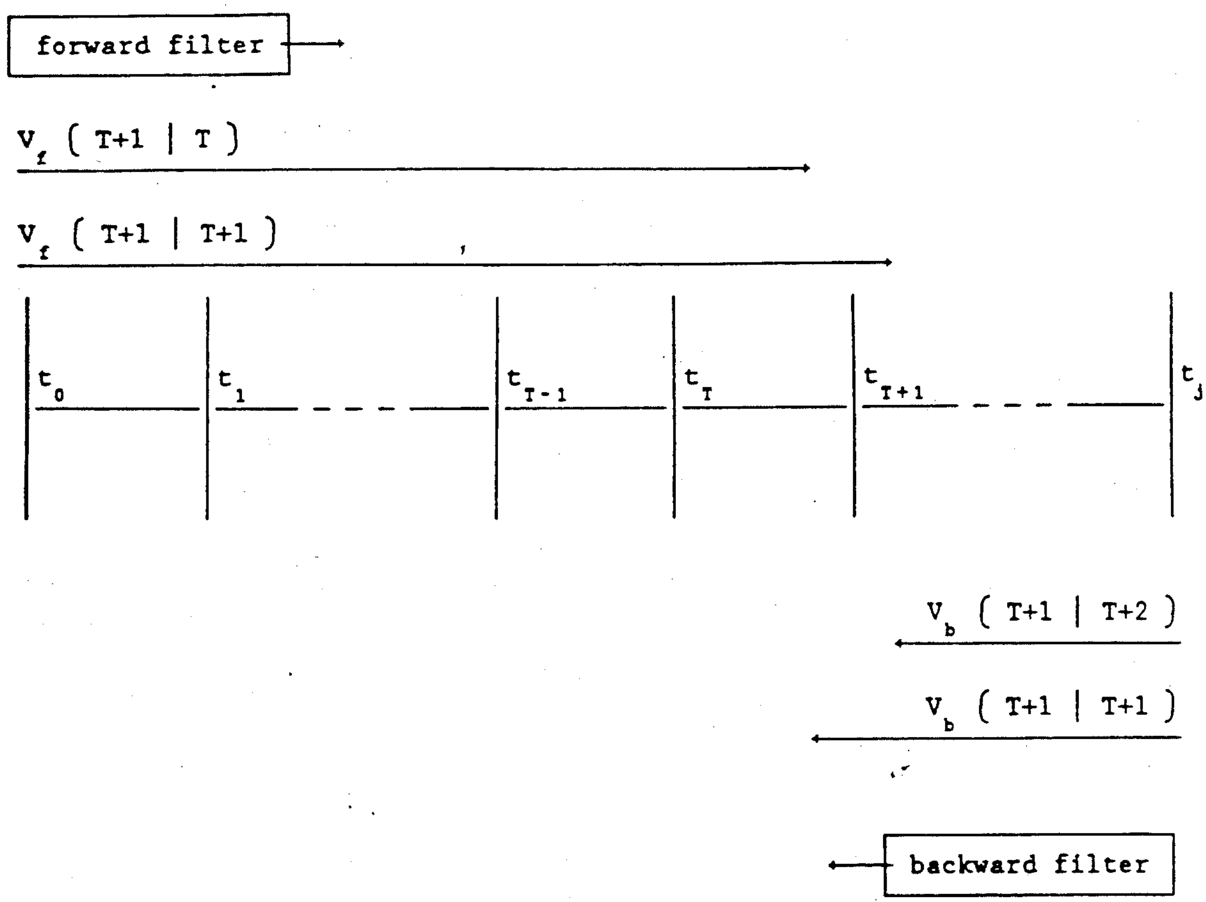

Figure 2: Combining forward and backward filters.

For the backward filter, $V_b(T|T+1)$ will represent the backward forecast based on $\{V_j, V_{j-1}, \ldots, V_{T+2}, V_{T+1}; X_j, X_{j-1}, \ldots, X_{T+2}, X_{T+1}, X_T\}$ where j represents the final period in the sample and $V_b(T|T)$ stands for the revised backward estimate after V_T has been processed by the backward filter. Figure 2 shows the relative position of the four estimates for the state variables at time $T=t$.

A smoothed estimate of the state at time $T=t$ can be formed by combining $V_f(T|T)$ and $V_b(T|T+1)$. The optimal weights are directly proportional to the amount of information in each of the estimates:

$$(9) \qquad V(T|t_j) = P(T|t_j)\{P_f^{-1}(T|T)\,V(T|T) + P_b^{-1}(T|T+1)\,V(T|T+1)\}$$

In equation (9) the subscripts f and b refer to respectively the forward and backward filters. $V(T|t_j)$ represents the smoothed estimate based on all the data and $P(T|t_j)$ is its covariance matrix. The inverse matrix $P^{-1}(T|t_j)$ equals the sum of $P_f^{-1}(T|T)$ and $P_b^{-1}(T|T+1)$. Obviously, one would obtain the same results if the smoother combined the estimates $V_f(T+1|T)$ from the forward filter and $V_b(T+1|T+1)$ from the backward filter.

In order to generate a covariance matrix for the smoothed estimates of the states that is immediately suitable for hypothesis testing, one will usually want to initialize both filters with an uninformative prior distribution for this covariance matrix. Such uniformative priors imply that the elements of the covariance matrix of the states are infinitely large before the first observation on the exogenous variable is processed. Often, therefore, this matrix is initialized for computational reasons with large numbers on the principal diagonal and zero for all off-diagonal elements. This procedure involves a numerical approximation. I have followed an exact alternative procedure, advocated by Maybeck (1979, 1982) which involves the so-called inverse covariance formulation. Whenever the matrix P is ill-conditioned, one computes instead conditional estimates of its inverse. In the implementation of the forward filter, the change to the "normal" forward specification is made as soon as the matrix P becomes numerically invertible; the backward filter remains in the inverse-covariance formulation for the complete period of estimation. Hence, equation (9) is not actually used but replaced by a smoother that combines either one ordinary filter and one inverse-covariance filter or two inverse-covariance formulations (see Maybeck, 1979, ch.5 and Maybeck, 1982, ch.8).

With this initialization, the smoothing algorithm will exactly reproduce the ols variance matrix of the parameters (and the ols residuals) in the special, restricted cases in which the dynamics of Y and its relationship to X can be represented by an ols regression for either the levels or the first or second differences. This depends on the computed or imposed values for the three unknown variances, Q_1, Q_2, and R. The remaining issue is how to discover these optimal estimates.[9] I use the Expectation Maximization

[9] The literature discusses a number of analytical procedures (see Ljung and Söderström, 1983, Maybeck, 1982). However, these authors tend to concentrate on the physical sciences where the data series are much longer than in economic applications and many model parameters are often known

(EM) algorithm, to compute maximum likelihood estimates. The EM algorithm is described by Dempster *et al.* (1977) and Watson and Engle (1983), and adapted to our case by Shumway and Stoffer (1982).[10][11][12]

4. Results for pairs of nonstationary time series

The first set of 100 pairs of Y and X consists of cointegrated series that only differ because Y has a temporary component and X a high-frequency autoregressive part. The second set of 100 pairs combines similar X series with an Y series that differs from X because of a third, unobserved, random walk with stochastic trend. Figure 1 above is a representative realisation of the second set of 100 experiments. In both cases, the long-term elasticity between X and Y equals 1.0. I shall investigate whether ols techniques are useful for tests of this elasticity and whether the unitary elasticity is delivered by the Kalman filter.

The first 100 pairs of X and Y are tightly connected and cointegrated, but the serially correlated disturbances in X would be expected to produce poor values for the Durbin-Watson statistic in the regressions in terms of levels, even though levels would be correct in terms of the longer-term dynamics of the series. It will be difficult for the analyst to decide whether these warning signals of positive serial correlation in the residuals of a regression in terms of levels signify an omitted variable, pointing to the need for differencing once or twice (not in fact the case

---

with great precision. See also Pagan (1980) for an analysis within the generalized least squares context.

[10] Separately, a parameter grid search was executed for a few models, combined with a Newton-Raphson algorithm for a more precise determination of the parameters. The grid search confirmed the outcomes of the EM method.

[11] Steyn (1987) and Nelson and Kim (1988) discuss other ways to compute the variances that are needed for the state update equation in the Kalman filter. Both papers caution against estimating hyperparameters on the basis of a single run through the data.

[12] See Nelson (1988) for evidence from his univariate research of U.S. gnp that optimization with respect to the unknown variances of the different shocks to the level and the shocks to the trend of a nonstationary time series may be a delicate matter. This is a topic for additional research.

here, but the correct diagnosis in the next series of experiments when a random walk with trend is placed "between" X and Y), or whether the serial autocorrelation should be dealt with using a model for the residuals or a smoother for the X series.

This prior is borne out by the results in table 1. Note the excellent estimates of the elasticity with a mean of 0.96. 73 Percent of the sample estimates are close enough to 1.0 that the null hypothesis of a unitary elasticity would not be rejected at the 0.05 percent level. If, however, the analyst decided to proceed with first-differencing the data because of the unacceptably high values for the Ljung-Box statistics, outcomes are much less useful. For the equations in first difference form, the null hypothesis of a unitary elasticity would be rejected in 100 percent of all cases.[13] According to the Ljung-Box statistics, one might prefer the equations in first difference form but that would mean not getting a handle on the elasticity between X and Y.

Table 1: Cointegrated random walks with trends

	ols levels				ols first differences			
	mean	stdev.	min.	max.	mean	stdev.	min.	max.
DW	1.90	0.18	1.13	2.20	1.95	0.37	1.12	2.68
R^2	0.95	0.083	0.51	1.00	0.035	0.031	0.00	0.12
b	0.96	0.079	0.55	1.00	0.050	0.033	-0.044	0.16
(S.E.)	0.016	0.012	0.00	0.054	0.030	0.0047	0.022	0.044
LBQ	174.3	66.4	47.3	361.9	90.1	72.2	24.10	327.6
ADF(1)	-9.49	0.80	-10.97	-6.21	-9.82	1.78	-13.97	-6.22
ADF(4)	-5.05	1.35	-7.26	-0.82	-2.85	0.87	-5.15	-0.92
ADF(8)	-2.91	1.04	-5.04	0.070	-1.75	0.76	-3.78	-0.0054

Note: DW = Durbin-Watson statistic, R^2 = coefficient of determination, b = estimated value of coefficient of explanatory variable, (S.E.) = estimated standard error of b, LBQ = Ljung-Box Q-statistic with 29 (in levels) and 28 (first differences) degrees of freedom (critical values 42.56 and 41.34 respectively), ADF(x) = Augmented Dickey-Fuller statistic with lag length x indicated in parentheses (critical value -2.89).

Turning now to the other 100 pairs of X and Y in which a random walk with stochastic trend is placed between X and Y, we obviously have series that no longer are cointegrated. Hence, the levels regressions would be inappropriate. Results in table 2 show - as could be expected - a clear

[13] The equations in second differences are even worse.

deterioration over those in the previous table, with much more uncertainty about b.

Table 2: Non-cointegrated random walks with trends

	ols levels				ols first differences			
	mean	stdev.	min.	max.	mean	stdev.	min.	max.
DW	1.14	0.58	0.0074	2.08	1.93	0.39	1.03	2.91
R^2	0.86	0.25	0.0005	1.00	0.035	0.031	0.00	0.13
b	0.92	0.39	-0.73	1.84	0.050	0.035	-0.042	0.17
(S.E.)	0.029	0.04	0.0036	0.24	0.031	0.0049	0.022	0.044
LBQ	291.5	228.0	65.0	1148.4	95.3	81.4	23.9	333.8
ADF(1)	-6.17	2.58	-10.27	2.33	-9.75	1.96	-16.39	-5.73
ADF(4)	-2.00	1.48	-6.05	1.78	-2.87	0.91	-5.12	-1.03
ADF(8)	-1.37	0.93	-2.96	1.34	-1.77	0.83	-3.66	0.56

Note: For abbreviations see table 1, LBQ = Ljung-Box Q-statistic with 29 (in levels) and 28 (first differences) degrees of freedom.

Visual inspection of the X series clearly shows the high-frequency cyclical movements in the explanatory variable. Hence, the analyst might try the somewhat old-fashioned technique of using a moving average of X rather than the current value of X only. Largely because of the prominence of the four econometric fashions that are the subject of this paper, older techniques involving moving averages, such as Almon lags, are much less prevalent nowadays than regressions in which large numbers of lagged values are included without restrictions on their coefficients. For the purpose of this illustration, I have used a simple 9-point moving average smoother, defined as follows:[14]

[14] This particular moving average, a linear smoother that is symmetric around the origin, works well in view of the periodicity of our autoregressive process. I have performed a large number of experiments with a more limited 5-point moving average that produced similar results, but obviously of somewhat lower quality than those in tables 3 and 4 below. Also, I have performed experiments in which a first order Almon lag was applied, so that the data determined whether the weights in the linear filter were equal or declined linearly over time. In all cases the estimated value of the slope parameter in the Almon lag was undistinguishable from zero, so that the results reported below are representative also of the linear Almon lag technique.

t:	-4	-3	-2	-1	0	1	2	3	4
weight:	1/16	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/16

Use of the linear filter assumes that the analyst made the correct inference from the autoregressive part in X as well as from the Ljung-Box statistics in tables 1 and 2, and used this or a similar smoother on the X series in order to eliminate movements in X that are not useful in forecasting and understanding Y. Table 3 deals with the 100 regressions in which each pair of X and Y are cointegrated; table 4 with the pairs of regressions in which a random walk with stochastic trend is added to the underlying level of X to produce the permanent level of Y.

Table 3: Cointegrated random walks with trends - smoothed series

	ols levels				ols first differences			
	mean	stdev.	min.	max.	mean	stdev.	min.	max.
DW	1.42	0.17	1.00	1.80	2.64	0.15	2.29	3.04
R^2	0.99	0.015	0.91	1.00	0.25	0.14	0.037	0.61
b	1.01	0.014	0.98	1.08	0.84	0.12	0.50	1.05
(S.E.)	0.0062	0.0062	0.00	0.036	0.17	0.045	0.082	0.28
LBQ	39.1	14.1	15.6	95.6	42.7	11.9	18.4	84.4
ADF(1)	-7.08	0.68	-8.62	-5.36	-13.36	1.21	-16.81	-11.06
ADF(4)	-4.41	0.64	-6.32	-3.26	-6.75	0.89	-9.09	-4.81
ADF(8)	-3.31	0.59	-4.61	-1.84	-4.08	0.86	-6.46	-2.37

Note: For abbreviations see table 1, LBQ = Ljung-Box Q-statistic with 27 (in levels) and 27 (first differences) degrees of freedom (critical value 40.11).

In table 3, the elasticity is well-determined in case of a regression in terms of levels and fairly close for regressions in first difference form. The smoother has eliminated much of the irrelevant high-frequency correlated noise in X and since the series are cointegrated and very similar, both types of regressions work well. The Durbin-Watson statistic would not provide clear guidance whether to work with levels or first differences, being on average 1.42 for the 100 levels regressions and 2.64 for the regressions in first differences.

The elasticity is also reasonably well-determined in the regressions for levels and first differences in table 4 which deals with series that are not cointegrated. The null hypothesis of a unitary elasticity is rejected in 98 percent of all cases. However, in this set of experiments, correct knowledge of the dynamics of the unobservable random walk with trend that has been placed "between" X and Y would dictate the taking of second differences. Ols regressions for second differences of Y and X - not shown in table 4 - produce extremely poor estimates for the regression coefficient b, which has a mean of -0.0036 over the 100 experiments with a standard deviation of 0.47.

Results for the Kalman filter models are in tables 5 through 8. Tables 5 and 6 relate to the original sets of data for Y and X; for tables 7 and 8 I used the 9-period symmetric smoother described above to the X series before applying the Kalman filter. We see immediately that results are very poor for the Kalman filter models using the original, unsmoothed X data. However, the Kalman filter is capable of producing worthwile estimates for the long-run elasticity between X and Y when the 9-point smoother is applied to the X series. In the case of the cointegrated series, the null hypothesis of a unitary elasticity is rejected in 6 out of 100 cases; with the random walk with stochastic trend between X and Y, the hypothesis is rejected in 15 percent of all cases. It is interesting how the Kalman filter estimates the elasticity with almost the same degree of exactitude in both cases: putting a random walk with stochastic trend between X and Y does not make it that much harder for the algorithm to estimate the long-term relationship between X and Y. In the ols experiments described in tables 3 and 4 there was more deterioration of the estimates in the experiments in which an unobserved random walk with trend was added to X.

Table 4: Non-cointegrated random walks with trends - smoothed series

	ols levels				ols first differences			
	mean	stdev.	min.	max.	mean	stdev.	min.	max.
DW	0.37	0.33	0.0076	1.52	2.58	0.16	2.13	2.97
R^2	0.89	0.23	0.0029	1.00	0.24	0.15	0.026	0.64
b	0.96	0.41	-0.82	2.00	0.83	0.17	0.33	1.37
(S.E.)	0.026	0.044	0.0017	0.25	0.17	0.047	0.082	0.29
LBQ	443.1	252.938	24.1	1108.5	40.2	11.8	16.3	74.6
ADF(1)	-2.75	1.72	-7.39	2.08	-12.95	1.21	-16.69	-10.13
ADF(4)	-1.56	1.18	-4.43	2.30	-6.07	0.90	-8.55	-4.02
ADF(8)	-1.50	0.95	-4.20	1.28	-3.41	0.75	-5.19	-1.96

Note: For abbreviations see table 1, LBQ = Ljung-Box Q-statistic with 27 (in levels) and 27 (first differences) degrees of freedom.

Table 5: Kalman filter for cointegrated series - no smoothing

	mean	stdev.	min.	max.
DW(r)	2.68	0.17	2.29	3.10
b	0.032	0.029	-0.047	0.11
(S.E.)	0.026	0.0039	0.019	0.036
LBQ(r)	50.8	15.1	24.0	101.9

Note: DW(r) = Durbin-Watson statistic of fit errors, b = estimated value of coefficient of explanatory variable, (S.E.) = estimated standard error of b, LBQ(r) = Ljung-Box Q-statistic of fit errors with 29 degrees of freedom.

Table 6: Kalman filter for non-cointegrated series - no smoothing

	mean	stdev.	min.	max.
DW(r)	2.69	0.17	2.25	3.13
b	0.032	0.029	-0.047	0.12
(S.E.)	0.026	0.0040	0.020	0.037
LBQ(r)	51.8	15.6	25.2	102.7

Note: for abbreviations see tables 1 and 5, LBQ(r) = Ljung-Box Q-statistic of fit errors with 29 degrees of freedom.

Table 7: Kalman filter for cointegrated series with smoothed X

	mean	stdev.	min.	max.
DW(r)	1.99	0.29	1.52	2.67
b	1.00	0.071	0.63	1.10
(S.E.)	0.060	0.039	0.017	0.21
LBQ(r)	35.9	11.9	18.0	74.7

Note: for abbreviations see tables 1 and 5, LBQ(r) = Ljung-Box Q-statistic of fit errors with 27 degrees of freedom.

Table 8: Kalman filter for non-cointegrated series with smoothed X

	mean	stdev.	min.	max.
DW(r)	2.27	0.26	1.66	2.91
b	0.92	0.14	0.48	1.22
(S.E.)	0.11	0.041	0.027	0.26
LBQ(r)	38.2	12.4	17.0	87.9

Note: for abbreviations see tables 1 and 5, LBQ(r) = Ljung-Box Q-statistic of fit errors with 27 degrees of freedom.

How do results for the Kalman filter compare to those of ols when smoothed X are used? In the case of the cointegrated series, ols produces a more precise estimate of the elasticity than the Kalman filter. However, in the case of non-cointegrated random walks described in table 4, results for the

Kalman filter are considerably better than the ols outcomes for levels and first differences. The enormous discrepancy in table 4 between the average estimated standard error of b and the sample estimate of the same standard deviation is one indication that ols is not appropriate in the case of nonstationary and non-cointegrated series. Results in table 4 for the first-differenced series are poorer than those of the Kalman filter. Results for second differences - as briefly mentioned above - are useless in the case of ols. The Kalman filter encompasses all three cases and produces results that are less precise in the case of cointegrated series, but still useful when differencing the data once or twice would be in order. The major advantage of the Kalman filter remains that it does not require a choice between levels, first differences and second differences.

5. Artificial experiments with cyclical data

The levels (or natural logarithms of levels) of many macroeconomic time series exhibit nonstationarity of the random-walk type, but other series, particularly those that are studied in terms of percentage changes over the previous quarter or year, show a combination of medium term cyclical movements and temporary shocks. Important examples are growth rates of output and employment and percentage changes in wages and prices. I have tried to construct artificial time series that mimic this type of time series behaviour. First, I have computed the average quarterly rate of inflation in the so-called G-7 countries (United States, Japan, West-Germany, France, United Kingdom, Italy and Canada) over the period 1965-1989. In each country, inflation peaked around the time of the first oil crisis of 1973-1974, exhibited a temporary low at some later stage during the 1970s, a second peak in the early 1980s in connection with the second oil price crisis and a further low in the mid-1980s. I have fitted a cubic spline function to the averages of these four local extreme values. The resulting pattern in the form of a camel's back, exhibited in figure 3 forms the basis for the experiments of this section.

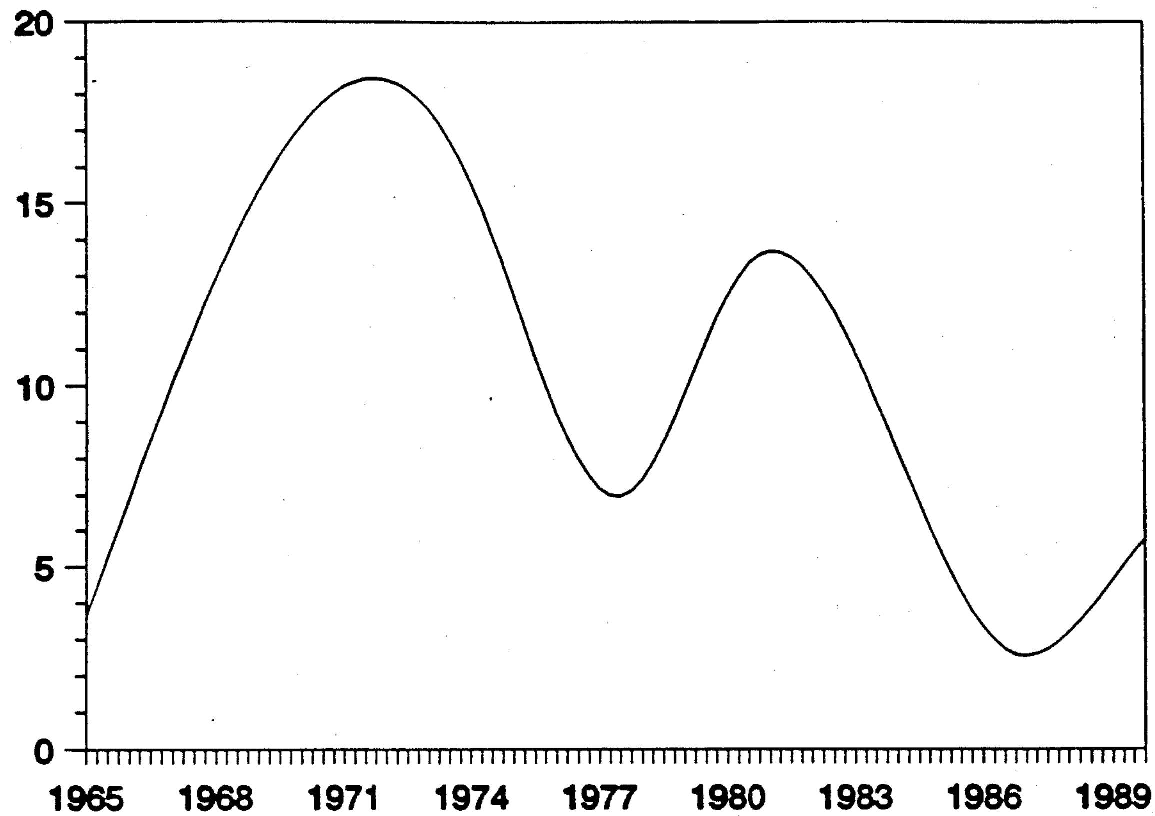

Figure 3: Fitted cubic spline function producing a camel's back pattern.

Adding a second order autoregressive model to this basic pattern produces realizations of the X process. This is similar to the construction of the X series in section 2, but in order to obtain greater similarity between the artificial series and the real-life examples involving rates of price and wage change, I have used a Student-t density function with 5 degrees of freedom for the error process. This process has fatter tails than the normal distribution.

As before, there are two sets of Monte Carlo experiments. First, I construct 100 replications of the X process and pair each one of these with an Y process that consists of the same camel's back pattern with added serially uncorrelated noise which has a t(5) distribution. In the second series of 100 experiments, the X series remains as before, but now an ARIMA(0,2,2) process is added to the camel's back to produce a wedge between the permanent level of X and the permanent level of Y before the observation errors are added to produce the Y series. The specification of the (0,2,2) process is as follows:

(10A) $$d_t = d_{t-1} + tr_{help,t-1} + \omega_{3t} + \omega_{4t}$$

(10B) $$tr_{help,t} = tr_{help,t-1} + \omega_{4t}$$

with $var(\omega_3) = 0.0025$, $var(\omega_4) = 0.00005$, and further properties as in equations (7B) and (7C) above.

As before, in the first set of 100 experiments, the two series are cointegrated, whereas in the second Monte Carlo study the discrepancies between the permanent levels of X and Y are nonstationary and may grow without bound as time goes on.

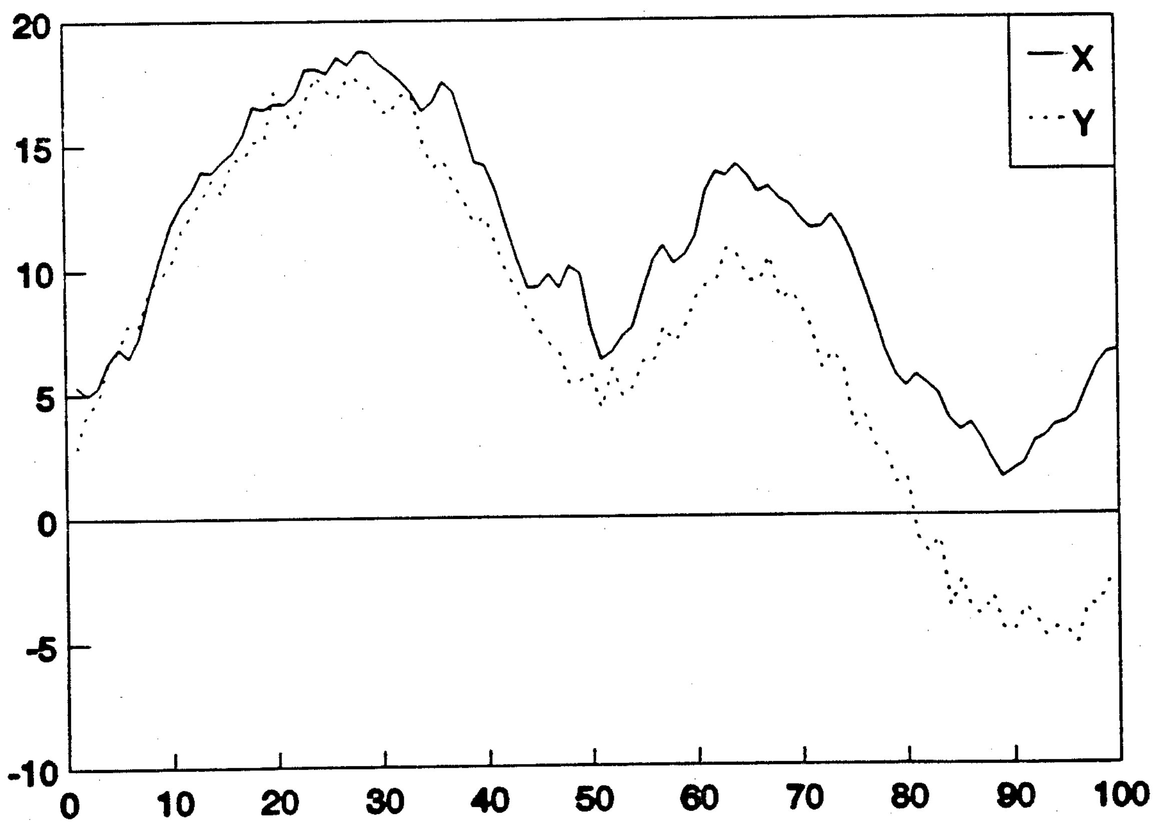

Figure 4: Two non-cointegrated cyclical series. The ols results from a levels regression of Y on X were: DW = 0.257, R^2 = 0.902, b = 1.307 with S.E. = 0.044, LBQ(29) = 255.683, ADF(1) = -12.35, ADF(4) = -10.318, ADF(8) = -6.115. For abbreviations, see table 1 above.

Figure 4 shows a typical example of the second set of experiments. Since the connection between X and Y is - as before -

> X - {autocorrelated temporary shocks} + {random walk with stochastic trend} + {observation noise} = Y,

it follows that the long-term elasticity between X and Y equals 1.0 also in this case. As before, I shall investigate whether ols-techniques are useful for tests of this elasticity, both in the case of cointegrated and that of non-cointegrated series, and whether this unitary elasticity is delivered by the Kalman filter.

6. Analysis of cyclical data

The first 100 pairs of X and Y are cointegrated series with an identical underlying structure in the shape of a camel's back. Additionally, Y has serially uncorrelated random noise, whereas to each X series a second-order autoregressive process of small amplitude has been added. As noted above, the noise series are non-normal and have a t-distribution with 5 degrees of freedom. I assume that the econometrician disposes of past and future X and is interested in producing longer-term forecasts of Y for which a proper estimate of the elasticity between X and Y is essential. One would expect ols in terms of the levels of X and Y to produce a sharp estimate of the unitary elasticity, but a poor value for the Durbin-Watson statistic and other indicators of serial correlation.

This prior is borne out by the results in table 9. Note the excellent estimates of the elasticity with a mean of 0.97. 62 Percent of the sample estimates are close enough to 1.0 that the null hypothesis of a unitary elasticity would not be rejected at the 0.05 percent level. If, however, the analyst decided to proceed with first-differencing the data because of the poor Durbin-Watson statistics and the often unacceptably high values for the Ljung-Box statistics, outcomes are much less useful. For the equations in first difference form, the null hypothesis of a unitary elasticity would be rejected in 100 percent of all cases. The equations in second differences are even worse. According to the Durbin-Watson statistics, one might prefer the equations in first difference form but that would mean not getting a handle on the elasticity between X and Y.

Table 9: Cointegrated cyclical variables

	ols levels				ols first differences			
	mean	stdev.	min.	max.	mean	stdev.	min.	max.
DW	1.08	0.17	0.74	1.56	2.43	0.18	1.96	2.93
R^2	0.96	0.012	0.91	0.97	0.12	0.051	0.035	0.28
b	0.97	0.028	0.88	1.03	0.36	0.091	0.13	0.60
(S.E.)	0.021	0.0027	0.017	0.029	0.10	0.013	0.063	0.14
LBQ	67.2	24.1	28.4	152.2	47.7	14.2	21.4	100.5
ADF(1)	-6.08	0.68	-7.89	-4.76	-12.41	1.21	-16.28	-9.71
ADF(4)	-5.24	0.74	-7.24	-4.04	-3.84	0.54	-6.04	-2.85
ADF(8)	-3.49	0.56	-4.97	-1.66	-2.15	0.29	-2.89	-1.53

Note: For abbreviations see table 1, LBQ = Ljung-Box Q-statistic with 29 (in levels) and 28 (first differences) degrees of freedom.

Table 10: Non-cointegrated cyclical variables

	ols levels				ols first differences			
	mean	stdev.	min.	max.	mean	stdev.	min.	max.
DW	0.64	0.29	0.14	1.40	2.42	0.18	1.95	2.93
R^2	0.90	0.081	0.51	0.97	0.12	0.051	0.036	0.27
b	0.95	0.18	0.55	1.35	0.36	0.092	0.13	0.59
(S.E.)	0.030	0.0095	0.018	0.059	0.10	0.013	0.063	0.14
LBQ	167.0	110.4	35.34	486.14	47.3	14.2	18.7	104.8
ADF(1)	-4.48	1.11	-7.29	-2.32	-12.36	1.22	-16.28	-9.63
ADF(4)	-3.45	0.86	-5.74	-1.55	-3.83	0.54	-6.13	-2.83
ADF(8)	-2.43	0.67	-4.14	-0.47	-2.15	0.28	-2.88	-1.55

Note: For abbreviations see table 1, LBQ = Ljung-Box Q-statistic with 29 (in levels) and 28 (first differences) degrees of freedom.

Turning now to the other 100 pairs of X and Y, in which a random walk with stochastic trend is placed between X and Y, we obviously have series that are no longer cointegrated. Hence, the levels regressions would be inappropriate. Results in table 10 show - as could be expected - a clear deterioration over those in the previous table.

Table 11: Cointegrated cyclical variables with smoothed X

	ols levels				ols first differences			
	mean	stdev.	min.	max.	mean	stdev.	min.	max.
DW	1.39	0.27	0.66	2.18	2.88	0.17	2.50	3.33
R^2	0.98	0.0045	0.97	0.99	0.28	0.043	0.19	0.37
b	1.02	0.030	0.97	1.09	0.98	0.051	0.85	1.10
(S.E.)	0.016	0.0018	0.012	0.02	0.17	0.018	0.14	0.22
LBQ	56.6	31.7	15.97	227.4	57.7	22.7	28.3	148.8
ADF(1)	-6.98	1.076	-10.68	-4.30	-15.51	1.72	-21.54	-12.14
ADF(4)	-3.51	0.60	-4.73	-1.79	-6.50	0.74	-8.95	-4.71
ADF(8)	-3.23	0.68	-4.94	-1.49	-4.23	0.55	-5.98	-2.45

Note: For abbreviations see table 1, LBQ = Ljung-Box Q-statistic with 27 (in levels) and 27 (first differences) degrees of freedom.

Table 12: Non-cointegrated cyclical variables with smoothed X

	ols levels				ols first differences			
	mean	stdev.	min.	max.	mean	stdev.	min.	max.
DW	0.73	0.37	0.15	1.85	2.87	0.17	2.49	3.32
R^2	0.93	0.074	0.55	0.98	0.27	0.043	0.18	0.38
b	1.00	0.20	0.53	1.45	0.98	0.053	0.84	1.10
(S.E.)	0.025	0.0087	0.014	0.05	0.17	0.019	0.14	0.22
LBQ	185.8	105.4	26.3	433.5	56.8	22.1	25.8	144.9
ADF(1)	-4.68	1.31	-9.00	-2.40	-15.40	1.70	-21.44	-12.06
ADF(4)	-2.63	0.64	-4.31	-1.19	-6.39	0.75	-9.00	-4.67
ADF(8)	-2.56	0.60	-4.17	-1.43	-4.11	0.56	-5.95	-2.49

Note: For abbreviations see table 1, LBQ = Ljung-Box Q-statistic with 27 (in levels) and 27 (first differences) degrees of freedom.

The next two tables give results for regressions in which the X series have been prefiltered with the symmetric 9-period filter. As before, I assume that the analyst made the correct inference from the serial correlation in the residuals and used a smoother on the X series in order to eliminate movements in X that are not useful in forecasting and understanding Y. Table 11 deals with the 100 regressions in which each pair of X and Y is cointegrated; table 12 with the regressions in which a random walk with (small) stochastic trend is added to the underlying level of X to produce the permanent level of Y.

In table 11, the elasticity is well-determined, both in case of a regression in terms of levels and for regressions in first difference form.

The smoother has eliminated much of the irrelevant high-frequency correlated noise in X and since the series are cointegrated and very similar, both types of regressions work well. The elasticity is also well-determined in the regressions for levels and first differences in table 12 which deals with series that are not cointegrated. It is interesting to note that the deterioration in the estimates using ols when non-cointegrated series are used is much less in this case than before when the series were of the "random walk with stochastic trend" type. This phenomenon is related to the size of the increments in the random walks and their trends in the two respective experiments, and it is also obvious that cyclical variables are more amenable to ols analysis than random walks with stochastic trends that do not exhibit any low frequency cyclical movements.

Table 13: Kalman filter for cyclical cointegrated series with smoothed X

	mean	stdev.	min.	max.
DW(r)	2.34	0.20	1.91	2.80
b	1.01	0.047	0.90	1.15
(S.E.)	0.057	0.018	0.022	0.11
LBQ(r)	34.8	12.7	17.3	94.5

Note: for abbreviations see tables 1 and 5, LBQ(r) = Ljung-Box Q-statistic of fit errors with 27 degrees of freedom.

Table 14: Kalman filter for cyclical non-cointegrated series with smoothed X

	mean	stdev.	min.	max.
DW(r)	2.38	0.19	1.72	2.83
b	1.01	0.047	0.88	1.13
(S.E.)	0.062	0.017	0.024	0.11
LBQ(r)	35.7	12.5	16.8	88.7

Note: for abbreviations see tables 1 and 5, LBQ(r) = Ljung-Box Q-statistic of fit errors with 27 degrees of freedom.

Results for the Kalman filter models are in tables 13 and 14. In both cases I have applied the 9-period symmetric smoother to the X series before running the Kalman filter. In the case of the cointegrated series, the null hypothesis of a unitary elasticity is rejected in 4 out of 100 cases; with the random-walk-with-stochastic-trend between X and Y, the null hypothesis is rejected in 5 percent of all cases. It is interesting how the Kalman filter model estimates the elasticity with almost the same degree of exactitude in both cases: putting a random walk with stochastic trend between X and Y does not make it harder for the algorithm to estimate the long-term relationship between X and Y. In the ols experiments there was some deterioration of the estimates in the experiments in which an unobserved random walk with trend was added to X (see tables 11 and 12).

Using the smoothed X series, results for the Kalman filter are - as before - similar to the best of the three sets of results with ols. The advantage of the Kalman filter remains that it does not require a choice between levels, first differences and second differences.

7. Four fashions in econometrics revisited

The general idea of a so called causality test is to regress a variable of interest on a number of its own lags as well as on observations from another variable. If, for instance, it appears that the predictions of a series Y based on past Y (notation: {Y-}) are inferior to forecasts based not only on past Y but also on past X (notation: {Y-,X-}), then one may say that X causes Y in the sense of Granger (1969). In another version of the test, named after Geweke, Meese and Dent (1982), one regresses X on past X as well as on past and future Y, tests whether {Y+} helps at the margin in predicting X, and if so concludes that X causes Y in the sense of Geweke *et al*.

The analyst has to decide how many lagged values to include in the specification, but the actual regressions are performed without restrictions on the coefficients of the included lagged endogenous or exogenous variables. Also, no attempt is made to interpret the pattern of the coefficients: the focus of attention is on comparing residual sums of squares or an F-test in order to judge whether inclusion of a group of explanatory variables does or does not help in predicting Y. If the conclusion is that X does indeed "cause" Y, this finding is not necessarily interpreted in a causal sense - after all people buy travel insurance before they go on a trip, but it makes no sense to state that the purchase of travel insurance "causes" the vacation. The finding that X "causes" Y means no more than that X helps at the margin in predicting Y which might signify either an errors in variables situation where the true process for the underlying level of Y is observed with error so that using values of X is valuable at the margin, or that movements in X antedate movements in Y. The causality tests are incapable of distinguishing between these two quite different interpretations.

Table 15: Causality tests

	Granger		Geweke	
	X causes (?) Y	Y causes (?) X	X causes (?) Y	Y causes (?) X
cointegrated random walks with trends	7%	100%	11%	100%
non-cointegrated random walks with trends	28%	97%	9%	92%
cointegrated cyclical series	49%	97%	51%	95%
non-cointegrated cyclical series	43%	91%	38%	78%

Note: the figures in the table indicate the percentage of cases in which an F-test reached significance at the 0.05 level.

Now consider the causality tests in the case of our pairs of artificial series Y and X. All four sets of experiments have contemporaneous movements in both series; there are no instances where X would change before Y or vice versa. Results are in table 15. Note that the hypothesis that Y does not cause X in the sense of Granger is rejected in 100 percent of all cases in the first 100 experiments (two cointegrated random walks with stochastic trends), in 97 percent of the second set of 100 experiments (two random walks that are not cointegrated), 97 percent in the third set (two cointegrated cyclical variables) and 91 percent in the fourth set of 100 experiments (two cyclical variables that are not cointegrated). The very high numbers signify that Y is almost always useful in forecasting X, which is understandable because X has high-frequency cyclical movements that obscure changes in the underlying level and underlying trend of X.

Tests for causality in the other direction reject the null hypothesis of no causality from X to Y in far fewer cases. Again, we know that X does not at all cause Y in the sense that movements in X antedate movements in Y, but that X may be useful at the margin in producing forecasts of Y. Note that the results for Granger's test and Geweke's test are quite close in most cases. Only in the case of the non-cointegrated random walks do the results for the test whether X causes Y differ considerably between Granger's methodology and Geweke's specification.

The results starkly exemplify the unavoidable ambiguity in the interpretation of such causality tests. X and Y are obviously related in our experiments, so that past values of either variable are useful both in predicting its own future and in predicting future values of the other variable. Since each variable X contains a high-frequency autocorrelated component, past values of Y are helpful in distinguishing whether the most recent movement in X has been predominantly temporary (part of the autoregressive component) or predominantly permanent (part of the underlying process, either the random walk with trend or the cyclical process). Y also has a temporary component, but it is not persistent and less important than the autoregressive temporary component in X. Hence, {X-} can be somewhat useful in predicting future values of Y if the regression alreay includes {Y-}, but the marginal contribution of {X-} to predictions of Y will be less than the marginal contribution of {Y-} in predicting X, for two reasons: past X is contaminated by temporary autocorrelated movements that are irrelevant for future Y, and Y itself only has uncorrelated observation noise in addition to its fundamental process, so that past Y is a higher-quality series for predicting future Y than past X.

In sum: causality tests only relate to questions about marginal predictive power and have no clear relationship to the question whether movements in one variable occur before another variable changes. In the experiments with our artificial series we note that different types of temporary disturbances can heavily influence the results of the causality tests, leading to the conclusion that one series (almost) always causes another series, even though movements in the permanent level and growth rate of the series occur simultaneously.

Vector autoregressions have become a prominent tool in the analysis of the dynamics of time series and especially in their interaction. Current levels (or logarithms of levels) of two or more time series are regressed simultaneously on a substantial number of past values, both of the series itself and of all the other time series in the analysis. In the present study of pairs of Y and X, we simultaneously regress levels of Y and X on {Y-} and {X-}. The regression coefficients are put in a matrix - in this specific case with two rows and sixteen columns because we use eight lagged values of both Y and X - and this matrix is manipulated to produce estimates of so-called "impulse response functions". These purport to show the effects of an "innovation" in either Y or X on each of the two series.

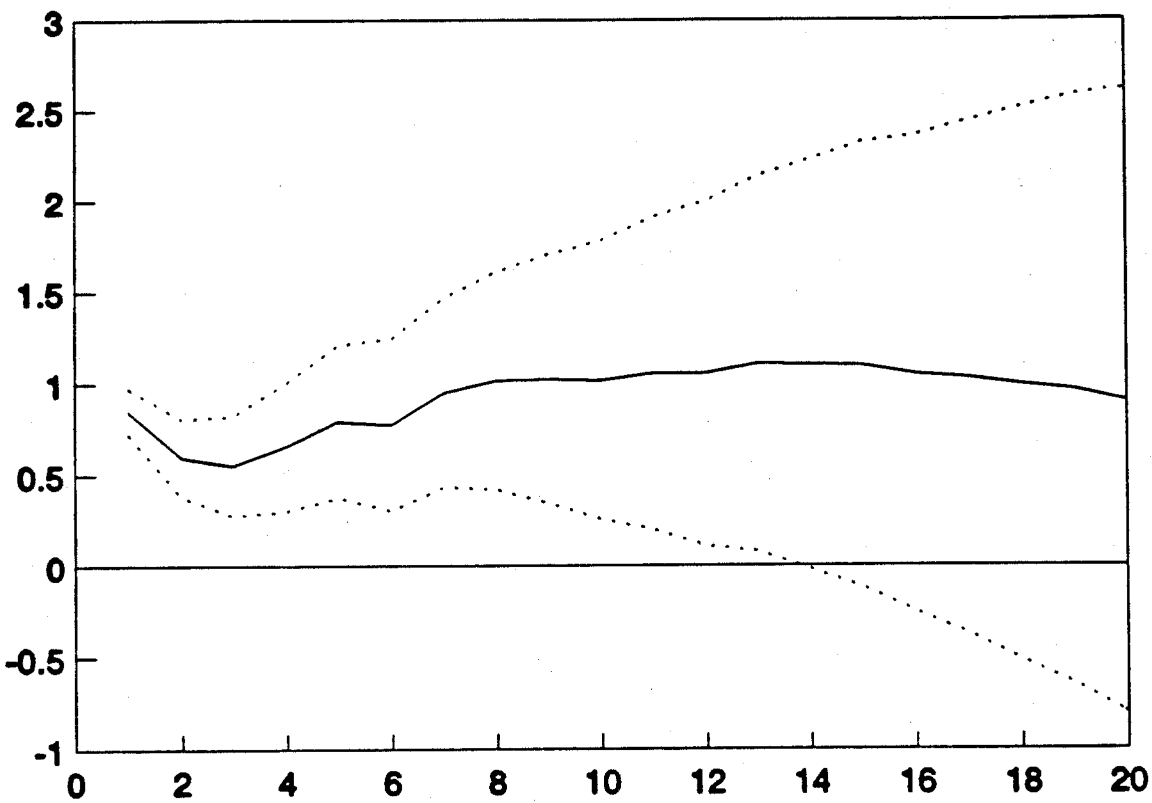

Figure 5: The response of Y to an innovation in Y: mean and two-standard deviation bands; computed using the Bayesian Monte Carlo integration technique.

Figure 5 shows a representative implementation of this technique for a pair of Y and X from the experiments underlying table 2 (non-cointegrated random walks with stochastic trends, one of which (X) has a high-frequency temporary component). The figure indicates the response of Y to a single innovation in Y that occurs at time $t=1$. The indicated standard errors for the estimated path have been computed using a Bayesian Monte Carlo Integration.[15] The next figure, figure 6, provides the same information

[15] See Kloek and Van Dijk (1978).

for the effects of a unit innovation in X on X itself.[16] Table 16 shows statistics for our experiments with cointegrated and non-cointegrated random walks with stochastic trends and temporary noise. The left-hand part of the table deals with the cointegrated pairs of Y and X; the right-hand part with the non-cointegrated series.

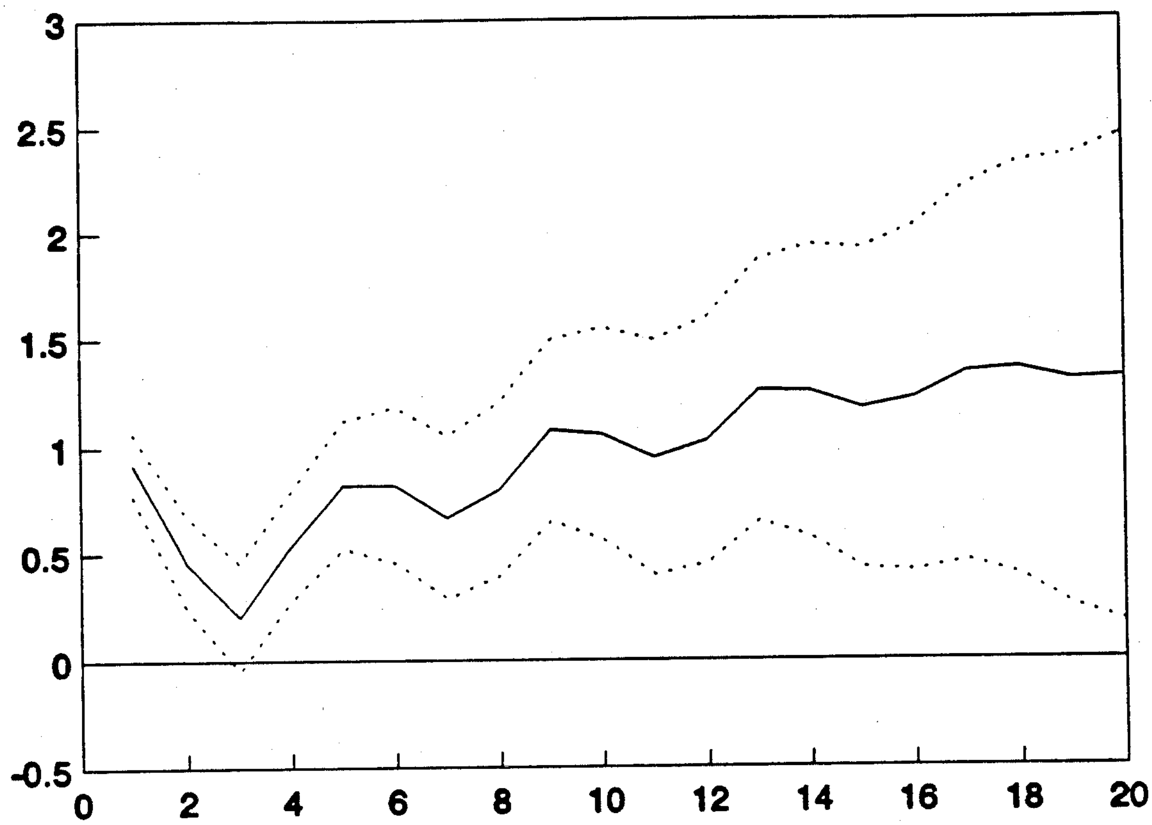

Figure 6: The response of X to a unit innovation in X: mean and two-standard deviation bands; computed using the Bayesian Monte Carlo integration technique.

The margins of uncertainty in figures 5 and 6 relate to uncertainty about the coefficients in the vector autoregression. For, the size of the innovation that takes place in period 1 is standardized across all experiments to unit size and not followed by any subsequent observations. A quite opposite way to consider the uncertainty surrounding this type of experiment would be to change to the state-space formulation. Even though state-space models are formally equivalent to ARIMA-models, the state-space

[16] It is also possible to use the vector autoregression to compute effects of innovations in X on Y and in Y on X, but these are not discussed here.

formulation emphasizes the continuous mixture of different types of shocks and also makes a careful distinction between the observed values of the process and the "underlying" level of the process without its temporary component. To bring out this different perspective on the uncertainty regarding simulation experiments, I have computed the expected path as well as the accompanying two-standard deviation bands for the state-space representations of Y and X in the asymptotic case that all hyperparameters of the stae-space model are known with certainty. This assumption corresponds, of course, to a ARIMA representation or Box-Jenkins model in which all model parameters are also known without error. Even in this extreme, asymptotic case, uncertainty remains regarding the path of the level of Y or X after a unit innovation during period 1, for two reasons:

- we observe a unit surprise, but do not know to what extent the size of the innovation is influenced by mistaken believes about the underlying level of the process as estimated after period 0 for period 1;

- the unit innovation is an unknown mixture of a temporary shock, a permanent disturbance to the level and a permanent disturbance to the rate of growth, and without knowing the allocation of the surprise over these three components, it is impossible to compute with certainty future values of the state variables.

Table 16: Vector autoregressions

	Cointegrated Y and X (as in table 1)				Non-cointegrated Y and X (as in table 2)			
	mean	min.	max.	stdev.	mean	min.	max.	stdev.
X on X								
expectation	0.79	-0.0019	1.74	0.36	0.48	-0.35	1.48	0.36
variance	0.22	0.035	3.28	0.34	0.12	0.0094	0.83	0.12
Y on Y								
expectation	1.23	0.080	2.77	0.63	1.04	-0.95	3.00	0.83
variance	0.63	0.074	2.71	0.48	0.49	0.028	1.82	0.41

Note: the figures in the table refer to the values of the variables 20 periods after the unit innovation has taken place. X on X means the response of X to a unit innovation in X, Y on Y indicates the response of Y to a shock in Y.

Figures 7, 8 and 9 show the theoretical response to a unit innovation for the following three cases:

1. the response of X to a unit innovation in X according to the state-space representation of the model for X as used in the first two sets of experiments;

2. the response of Y to a unit innovation in Y using the state-space representation of Y as used in the first set of experiments;

3. the response of Y to a unit innovation in Y, using the state-space model for variable *d* which was defined in equation (10A) as the differnce between Y and the permanent component of X.

The figures show the asymptotic results that are based on numerical solutions to the Riccati equations associated with the state-space models in equations (4) and (6) above.[17]

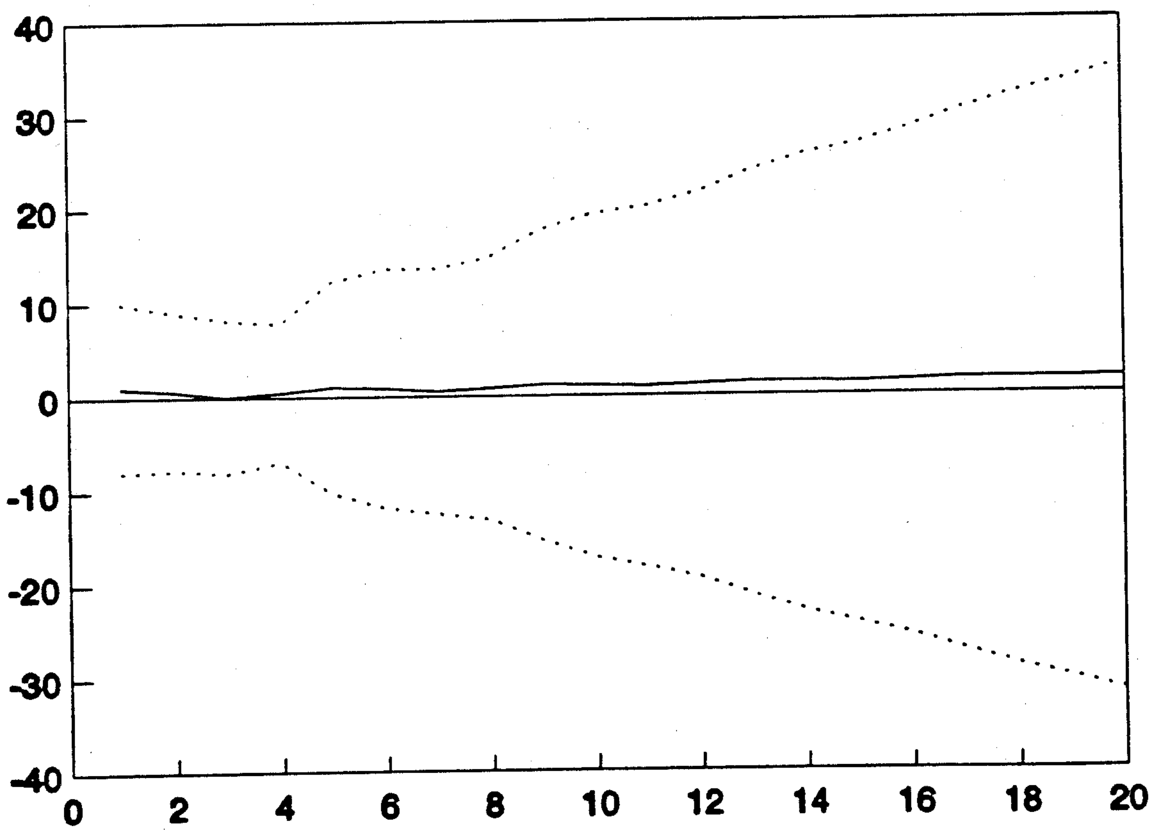

Figure 7: the theoretical response of X to a unit innovation in X: asymptotic path and two-standard deviation bands.

---

[17] The precise deriviations can be found in Stengel (1986) and Mendel (1987).

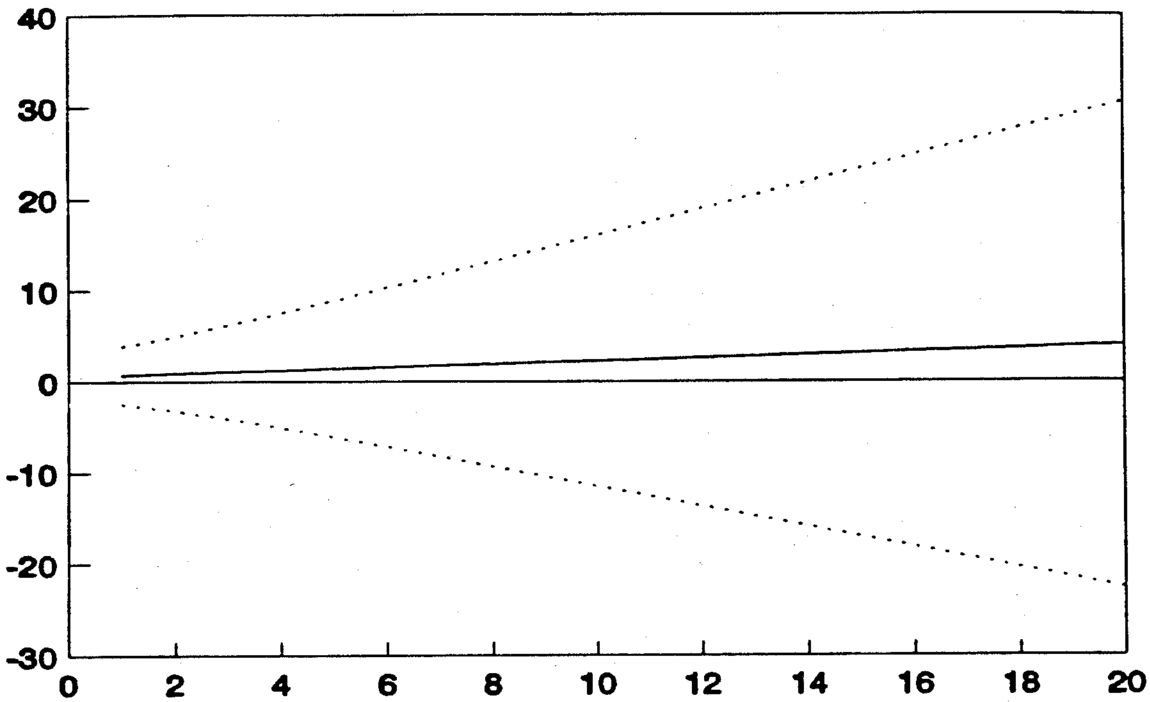

Figure 8: the theoretical response of Y to a unit innovation in Y: asymptotic path and two-standard deviation bands.

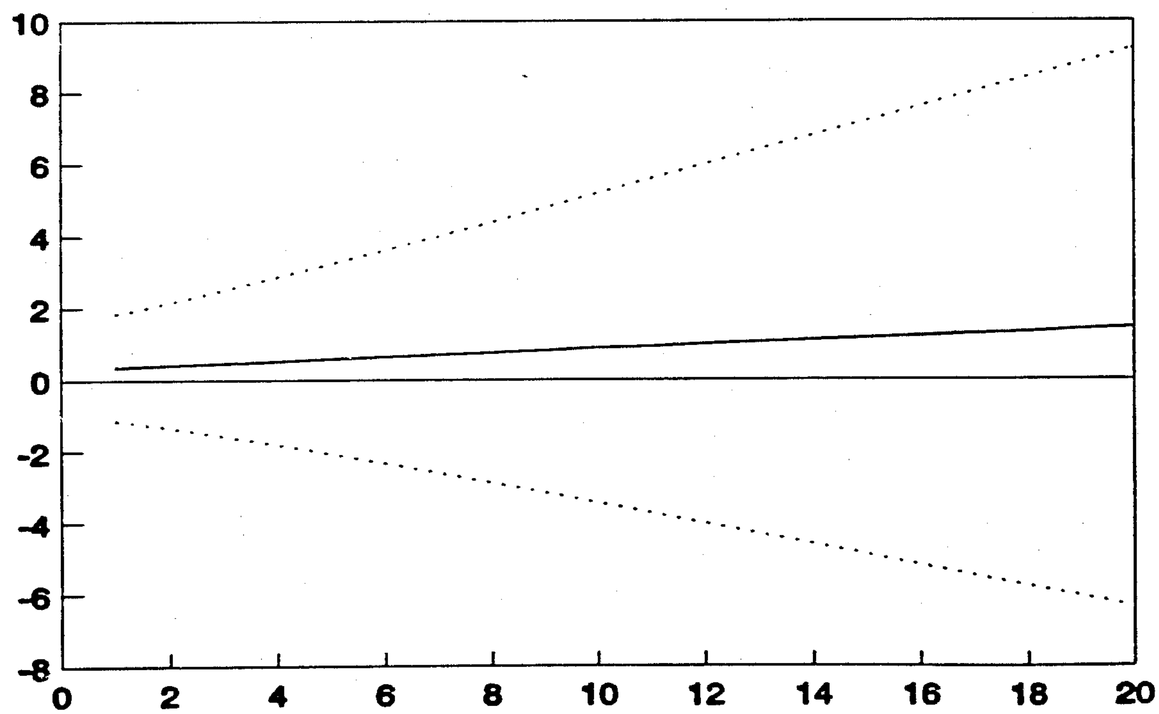

Figure 9: the theoretical response of Y to a unit innovation in Y: asymptotic path and two-standard deviation bands.

Analysis of the theoretical responses of a state-space model to a unit innovation shows that the effects of an innovation will depend on its size. The larger the innovation, the less important will it be that at the moment it occurs we have an unavoidable residual amount of uncertainty about the true state of the sustem, because in all periods prior to t=1 there also were shocks that hit the system. For instance, in the introduction, we looked at the simple case in which the Box-Jenkins parameter in an ARIMA(0,1,1) model equaled 0.5, corresponding to variances of respectively 0.5 and 0.25 for the temporary and permanent shocks in the equivalent state-space model. If now the innovation hitting such a system follows upon a long series of similar innovations we can use the Box-Jenkins representation or the Riccati equation to compute the expected future path after the innovation. However, if the true state of the system is known without error when the shock hits, or - equivalently - if the shock is much larger than those that occurred in the past, we should classify the surprise into temporary and permanent parts according to the sizes of the two variances in the state-space model, implying that in this case two thirds of the shock should be classified as temporary and only one third as permanent.

Here follows the same argument for the actual case of the Y series in our first set of experiments. Solving the Riccati equation in the asymptotic case of an infinitely long realization of Y results in the vector (0.712, 0.170). It follows that the innovation is classified in the following way: 71% is considered to be a permanent increase in the level, 17% is judged to be a permanent increase in the rate of growth, and the remaining 12% should be considered to be purely temporary. However, the three variances in the model of equation (4) above are 1 for the temporary component, 1 for the permanent shifts and 0.1 for the permanent changes in the stochastic trend, suggesting a quite different split of any surprise over the three components. In case a particularly large innovation had to be analyzed, the classification of the surprise into its three components would shift from the numbers indicated by the solution of the Riccati equation to the proportions as indicated by the three variance terms.

Unit root tests have become quite popular for testing whether a single time series should be differenced before further statistical analysis is undertaken. Table 17 shows the results of unit root tests applied to the four sets of series in this Monte Carlo study. Numbers in the table are based on the 0.05 significance level. We note that the low frequency cyclical series with the camel's back pattern are all classified as stationary using unit root tests with four or more lags. The particular construction of the camel's back logically means that all series built on that frame are nonstationary, but there exist stationary higher-order autoregressive processes that could approach the same pattern to any degree of accuracy. Hence, one should not find fault with the unit root test for rejecting the null hypothesis of nonstationarity for the Y and X series that are based on the camel's back pattern. The Y series in the second set of experiments do contain a purely nonstationary component, and again the series are not recognized as nonstationary by the unit root tests.[18]

Many wrong answers, by contrast, are produced by the unit root tests in the cases of the random walks with stochastic trend. Note that many realizations of the Y series are classified incorrectly as stationary if four or more lags are used.

Table 17: Unit root tests

	X				Y			
lags	0	1	4	8	0	1	4	8
random walks with trend (see table 1)	30	30	5	22	2	2	18	27
random walks with trend (see table 2)					4	5	14	26
camel's back series (see table 9)	4	100	100	100	6	0	100	100
camel's back series (see table 10)					46	33	100	100

Note: the figures in the table indicate the percentage of cases in which the null hypothesis of a unit root was rejected using the Augmented Dickey-Fuller test.

The results in table 17 confirm much earlier evidence that unit root tests for single series have low power. This is also true for the results of the Augmented Dickey-Fuller tests that were already incorporated in tables 1 through 4. These tests suffer from an additional disadvantage because their

[18] The standard deviation of the permanent shift in the random walk that is added to X equals 0.05; the permanent changes in the growth rate have a standard deviation of 0.00707. Over 100 periods this produces on average a deviation of 3.839 between the levels of X and Y, a distance that should be compared to the range of the basic camel's back which goes from 2.544 to 18.449.

null hypothesis is that of nonstationarity. For nonstationarity occurs in infinitely many forms, implying the risk that in the case at hand the model of nonstationarity differs from those considered by Dickey and Fuller. In fact, the table in their (1979) paper considers first order autoregressive processes with values around 1 for the autoregressive coefficient, even though one could claim that the time series models considered in this paper which are of the ARIMA(0,2,2) or ARIMA(2,2,2) variety are closer to macroeconomic time series. In my view, the state-space methodology has an important advantage because it does not require any decision about the number of unit roots, but instead transforms the problem into that of properly estimating one or more hyperparameters that indicate the importance of temporary versus permanent shocks to the system.

The cointegration technique can be used to test whether there exists a unitary elasticity between X and Y in these Monte Carlo experiments, in the sense that conditional forecasts of Y should be changed proportionally to changes in one's forecast of X. That happens to be correct for all the Monte Carlo experiments. Results for the cointegration tests are in table 18. The tests are based on a bivariate regression in which first differences of X and Y are simultaneously regressed on a series of lagged first differences of X and Y as well as on the lagged levels of the two variables (see Johansen, 1991). Formally:

$$(11) \qquad \Delta Z_t = \sum_{i=1}^{k-1} \Gamma_i \Delta Z_{t-i} + \Pi Z_{t-k} + \mu + \varepsilon_t,$$

where Z is a vector containing Y and X and ε_t $(t = 1,\ldots,T)$ are independent 2-dimensional Gaussian variables with mean zero and variance matrix Λ. In our case, $k = 8$. The parameters $\Gamma_1,\ldots,\Gamma_{k-1}$, μ and Λ are assumed to vary without restrictions, and the hypothesis of a single r cointegration vector is formulated as the following restriction on Π:

$$(12) \qquad \Pi = \alpha\beta',$$

where β, the cointegration vector, and α, the adjustment coefficients, are 2×1 matrices. The test is then performed on the elements of β. The hypothesis of a unitary elasticity requires $\beta_1 = -\beta_2$.

Table 18: Cointegration

	reject hypothesis of no cointegration	accept unit elasticity
cointegrated random walks with trends	82%	86.6%
non-cointegrated random walks with trends	27%	40.7%
cointegrated cyclical series	98%	62.2%
non-cointegrated cyclical series	92%	7.6%

Note: the figures in the first column indicate the percentages of cases in which the null hypothesis of no cointegration could be rejected at the 0.05 level, those in the last column refer to the percentage of cases in which, given the rejection of the null, a unit elasticity between X and Y could not be rejected.

The results in table 18 show that in the first set of 100 experiments the cointegration test correctly classifies 82 out of the hundred cases as being cointegrated and subsequently does not reject the hypothesis of a unitary elasticity at the 0.05 level in 7 out of 8 cases. However, once the random walks with stochastic trends are no longer cointegrated, the cointegration technique becomes incapable of producing useful results for the test of the unitary elasticity. The percentages indicated in the table are almost identical to the split over rejection versus non-rejection of the null hypothesis of a unitary elasticity in case the cointegration hypothesis is indeed rejected. In the third and fourth set of experiments involving cyclical variables the cointegration technique classifies virtually all experiments as exhibiting cointegration. It is interesting to note how in the 100 cases in which the cyclical variables are indeed cointegrated the test quite frequently does not reject the null hypothesis of a unitary elasticity, but how again in the case of non-cointegrated series the test is not at all useful in illuminating the question whether the elasticity equals 1.0.

It seems to be the case that the test is not well designed to investigate the long term elasticity between two variables X and Y in case X and Y are not cointegrated.[19] Nevertheless, this is a context that very frequently

[19] McCallum (1984) has correctly emphasized how ambiguous definitions of elasticities can become in the case of stationary variables. It is ironic that ols works best for stationary variables when elasticities are often ill-defined or unobservable, whilst nonstationary series offer a better context to precisely define elasticities, but are hard to tackle using traditional ols methods.

occurs in macroeconomics, for instance when one speaks about exact price compensation in wages. Series for inflation and growth in nominal wages need not be cointegrated; nevertheless it makes sense to talk about 1-1 compensation in wages for inflation. The cointegration technique does not seem capable of dealing with such instances.

8. Conclusions

In a recent article Lawrence Summers (1991) complains about the scant progress made during the past two decades in applied macroeconomics. Without necessarily agreeing with Summers's analysis, let alone his recommendations, many economists will share his sense of disappointment. The present paper reconsiders four prominent econometric techniques that have been used in many of the papers in applied macroeconomics during this period. I have studies these four techniques in a number of Monte Carlo experiments using nonstationary time series that were quite closely related over the longer term, but subject to different short-term components. Nonstationarity creates well known problems for ordinary least squares and its many applications. At the same time, the "errors-in-variables" context makes it impossible to interpret the regression findings: does the fact that Y causes X mean that Y changes before X, or does it signify that observations on X are more seriously contaminated by high-frequency noise than observations on Y?

The paper combines experiments using causality tests, vector autoregressions, unit root tests and cointegration with implementation of a multivariate Kalman filter. This allows for a comparison between the autoregressive-moving average interpretation of time series and the state-space formulation. Even though the two modelling techniques are formally equivalent, the state-space representation emphasizes the often useful distinction between the observed series and its "underlying" level, and allows for an interpretation in which a time series does not change over time because of a single type of innovation with a complicated dynamic impulse reaction function, but because different types of disturbances, temporary or permanent, occur simultaneously.

Unsurprising is the general finding that observations on two time series without additional structure are insufficient to draw interesting conclusions about cause and effect, but can only be used to compare different predictive formulas. A more interesting finding is that none of the four econometric fashions is capable of delivering a useful estimate of the unitary elasticity between the constructed time series, unless a simple moving average is applied to one of the series. It seems fair to say that moving averages and Almon lags went out of fashion with the introduction of multicoefficient autoregressions and vector autoregressions, but in the Monte Carlo simulations in this paper, the old-fashioned Almon lag technique is capable of producing useful results, whereas unrestricted estimates of many coefficients are of no help.

The multivariate Kalman filter also requires some smoothing of the explanatory variable in order to produce useful forecasts and estimates of the unitary elasticity between X and Y. The principal advantage of the Kalman filter in this context is that it circumvents the difficult choice between working with levels, first differences or second differences. In the implementation chosen for this paper, the Kalman filter is capable of dealing with any series that can usefully be described with an ARIMA(0,2,2) model, and leaves it to the data to determine the relative importance of the three types of basic innovations (temporary shocks to the level, permanent shocks to the level, permanent shocks to the rate of growth) that occur in such a model.

The contrast between regression techniques and Kalman filter estimation also helps in the discussion of simulations based on a vector autoregression. The Kalman filter (or state-space formulation) indicates first why it is incorrect to talk about simulation experiments in terms of a single shock that takes after a tranquil period, and second, shows that the estimated response function must logically be dependent upon the size of the shock. Both these insights are impossible to get out of the autoregressive-moving average representation of the same process. Experiments in the paper show that at least sometimes uncertainty about the classification of the innovation as temporary or permanent is much more important for the margins of uncertainty surrounding the innovation experiment than uncertainty about the coefficients in the corresponding vector autoregression. Differences between the computed uncertainty deriving from the true character of the innovation on the one hand, and model parameter uncertainty on the other hand are so large, that this finding should generalize to other contexts.

References

BOX, GEORGE E.P. AND GWILYM M. JENKINS, 1970, "Time series analysis: forecasting and control", Holden-Day, San Francisco.

CHRISTIANO, LAWRENCE J. AND MARTIN EICHENBAUM, 1990, "Unit roots in real gnp: do we know, and do we care?", Carnegie-Rochester Conference Series on Public Policy, Vol. 32, pp. 7-82.

DEMPSTER, ARTHUR P., N.M. LAIRD, AND D.B. RUBIN, 1977, "Maximum likelihood from incomplete data via the EM algorithm", Journal of the Royal Statistical Society, Series B, Vol. 39, pp. 1-22.

DICKEY, DAVID A. AND WAYNE A. FULLER, 1979, "Distribution of the estimators for autoregressive time series with a unit root", Journal of the American Statistical Association, Vol. 74, No. 366, pp. 427-31.

GEWEKE, JOHN, RICHARD MEESE AND WARREN DENT, 1982, "Comparing alternative tests of causality in temporal systems", Journal of Econometrics, Vol. 21, pp. 161-94.

GRANGER, CLIVE W.J., 1969, "Investigating causal relations by econometric models and cross-spectral models", Econometrica, Vol. 37, pp. 424-38.

JOHANSEN, SØREN, 1991, "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models", Econometrica, Vol. 59, pp. 1551-80.

KLOEK, TEUN AND HERMAN K. VAN DIJK, 1978, "Bayesian estimates of equation system parameters: an application of integration by Monte Carlo", Econometrica, Vol. 46, pp. 1-20.

LJUNG, LENNART AND TORSTEN SÖDERSTRÖM, 1983, "Theory and practice of recursive identification", The MIT Press.

MAYBECK, PETER S., 1979, "Stochastic models, estimation, and control volume 1", Mathematics in Science and Engineering, Vol. 141-1, Academic Press.

MAYBECK, PETER S., 1982, "Stochastic models, estimation, and control volume 2", Mathematics in Science and Engineering, Vol. 141-2, Academic Press.

MCCALLUM, BENNETT T., 1984, "On low-frequency estimates of long-run relationships in macroeconomics", Journal of Monetary Economics, Vol. 14, pp. 3-14.

MENDEL, JERRY M., 1987, "Lessons in digital estimation theory", Prentice Hall, Englewood Cliffs, New Jersey.

NELSON, CHARLES R., 1988, "Spurious trend and cycle in the state-space decomposition of a time series with a unit root", Journal of Economic Dynamics and Control, Vol. 12, pp. 475-88.

NELSON, CHARLES R. AND CHANG-JIN KIM, 1988, "The time-varying-parameter model as an alternative to ARCH for modeling changing conditional variance: the case of Lucas hypothesis", N.B.E.R. Technical Working Paper Series, No. 70, September.

NELSON, CHARLES R. AND CHARLES I. PLOSSER, 1982, "Trends and random walks in macroeconomic time series: some evidence and implications", Journal of Monetary Economics, Vol. 10, No. 2, pp. 139-62.

PAGAN, ADRIAN, 1980, "Some identification and estimation results for regression models with stochastically varying coefficients", Journal of Econometrics, Vol. 13, pp. 341-63.

SHUMWAY, ROBERT H. AND D.S. STOFFER, 1982, "An approach to time series smoothing and forecasting using the EM algorithm", Journal of Time Series Analysis, Vol. 3, No. 4, pp. 253-64.

STENGEL, ROBERT F., 1986, "Stochastic optimal control", John Wiley & Sons, New York.

STEYN, IVO, 1987, "Recursive estimation of parameters in the Kalman filter", Unpublished Paper, University of Amsterdam.

SUMMERS, LAWRENCE H., 1991, "The scientific illusion in empirical macroeconomics", Scandinavian Journal of Economics, Vol. 93, pp. 129-48.

WATSON, MARK W. AND ROBERT F. ENGLE, 1983, "Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models", Journal of Econometrics, Vol. 23, pp. 385-400