

Common Large Innovations Across Nonlinear Time Series

Richard Paap*
Econometric Institute
Erasmus University Rotterdam

and

Philip Hans Franses
Econometric Institute
Erasmus University Rotterdam

Econometric Institute Report 2002-09

Abstract

We propose a multivariate nonlinear econometric time series model, which can be used to examine if there is common nonlinearity across economic variables. The model is a multivariate censored latent effects autoregression. The key feature of this model is that nonlinearity appears as separate innovation-like variables. Common nonlinearity can then be easily defined as the presence of common innovations. We discuss representation, inference, estimation and diagnostics. We illustrate the model for US and Canadian unemployment and find that US innovation variables have an effect on Canadian unemployment, and not the other way around, and also that there is no common nonlinearity across the unemployment variables.

Key words: Nonlinearity, common features, censored latent effects autoregression

*Corresponding author: Erasmus University Rotterdam, Econometric Institute (H11-2), P.O. Box 1738, NL-3000 DR Rotterdam, the Netherlands, e-mail: paap@few.eur.nl. We thank Dennis Fok for helpful discussions. The first author thanks the Netherlands Organization for Scientific Research for its financial support.

1 Introduction

It is now widely accepted that nonlinear time series models can capture cyclical fluctuations in economic time series variables. There are by now many such models, see for example the surveys in Granger and Teräsvirta (1993) and Franses and van Dijk (2001), among others. Examples of popular models are the (smooth) threshold autoregressions, Markov switching models, and artificial neural networks.

In many instances the nonlinear models are considered for univariate time series. This may be due to the potentially large number of parameters involved in straightforward multivariate versions of the above models. It may therefore be relevant to examine if two or more time series have common nonlinear features. An additional and economic motivation for this examination is that this could give insights as to whether two or more time series display common cyclical fluctuations.

Testing for common nonlinearity is often not an easy exercise. The main reason for this is that most nonlinear econometric time series models incorporate the nonlinearity in the parameters. For example, the smooth transition autoregression basically assumes that the autoregressive parameters are varying over time, see Granger and Teräsvirta (1993), among others. Hence, when one aims to examine common nonlinearity, in a sense one has to look for common varying parameters. Anderson and Vahid (1998) propose a method to do so, but the resulting method is based on linear regression models, which only to some extent approximate the original nonlinear model.

In this paper we propose to circumvent the above noted problems by considering a new nonlinear time series model, which basically assumes that nonlinearity can be summarized by an additional innovation-type variable. The univariate version of this, so-called, censored latent effects autoregression [CLEAR], is put forward in Franses and Paap (2002). In words, this model consists of an autoregression with two error terms. The first error term is a standard white noise process, while the second error term only once in a while gets non-zero values, depending on the outcome of a censored regression model. In Franses and Paap (2002) the CLEAR model is applied to US unemployment, where

unexpected large positive increases are explained by a censored regression with a single leading indicator variable. Once the right-hand side of this censored regression exceeds a stochastic threshold level, a positive innovation is added to the autoregressive model. Together, the CLEAR model can capture the nonlinear features in US unemployment very well, while the number of parameters is rather small. As the nonlinear feature enters the model through a variable, the CLEAR model seems well suited for investigating common linearity. Indeed, when a linear combination of two or more variables cancels these additional variables, one has encountered common nonlinearity.

The outline of our paper is as follows. In Section 2 we outline the basics of the univariate CLEAR model and a suitable representation of its multivariate version. In Section 3, we briefly discuss unconditional inference. In Section 4, we treat parameter estimation, which can be done by Maximum Likelihood, and we deal with conditional inference. In Section 5, we illustrate our model for US and Canadian unemployment. We find that US innovation variables have an effect on Canadian unemployment, and not the other way around, and also that there is no common nonlinearity across the unemployment variables. In Section 6, we conclude with some remarks.

2 CLEAR models

In this section we discuss various features of the univariate CLEAR model, and a suitable representation of its multivariate extension.

2.1 A Univariate CLEAR model

Franses and Paap (2002) propose a censored latent effects autoregressive [CLEAR] model to describe the salient features of US unemployment. A univariate CLEAR model of order p for a univariate time series $\{y_t\}_{t=1}^T$ is given by

$$y_t = \sum_{i=1}^p \alpha_i y_{t-i} + x_t' \gamma + v_t + \varepsilon_t, \quad (1)$$

where $\varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$, y_{p-1}, \dots, y_1 are fixed, and where x_t is a $(k \times 1)$ vector of exogenous variables including an intercept. The variable v_t is added as an extra innovation

which helps to describe the sharp rise in unemployment, which is usually associated with recessions. Therefore, v_t is imposed to be a censored variable modelled by

$$v_t = \begin{cases} x'_t\beta + u_t & \text{if } x'_t\beta + u_t \geq 0 \\ 0 & \text{if } x'_t\beta + u_t < 0 \end{cases} \quad (2)$$

with $u_t \sim \text{NID}(0, \sigma_u^2)$ and where β and γ are $(k \times 1)$ parameter vectors. The disturbances ε_t and u_t are assumed to be uncorrelated, although various extensions are possible. Also applications other than unemployment may require other versions of (2).

The effects of the exogenous variables on y_t are given by γ when $v_t = 0$, while $\gamma + \beta$ represent the effects of the exogenous variables when $v_t \geq 0$. Hence, the CLEAR model is a nonlinear model as it allows for a time-varying effect of the explanatory variables. The nonlinear part of the model concerns the variable v_t . As (2) implies that only positive values of v_t are added to the error term, the above CLEAR(p) model contains an explicit description of what might be viewed as an exogenous positive innovation outlier generating mechanism. The variable v_t is zero unless $x'_t\beta$ exceeds a stochastic threshold level $-u_t$, where u_t is a normal random variable. When the threshold is exceeded, v_t takes a positive value.

Franses and Paap (2002) successfully describe the US unemployment rate using for x_t a lagged value of a single leading indicator variable. In their application it turns out that the γ may be restricted to be zero, but that the censored variable v_t cannot be removed from the model. Graphs of the estimates of v_t show that the innovation outlier generating variable closely follows the cyclical pattern in US unemployment.

2.2 A Multivariate CLEAR model

The univariate CLEAR model (1) and (2) can be generalized to a multivariate CLEAR. For example, a bivariate CLEAR(p) for the vector time series $(y_{1t}, y_{2t})'$, $t = 1, \dots, T$ may be represented by

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \sum_{i=1}^p A_i \begin{pmatrix} y_{1t-i} \\ y_{2t-i} \end{pmatrix} + \begin{pmatrix} x'_{1t}\gamma_{11} & x'_{2t}\gamma_{12} \\ x'_{1t}\gamma_{21} & x'_{2t}\gamma_{22} \end{pmatrix} + \begin{pmatrix} 1 & \delta_{12} \\ \delta_{21} & 1 \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \quad (3)$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})' \sim \text{NID}(\mathbf{0}, \Sigma_\varepsilon)$ with Σ_ε a (2×2) covariance matrix, and where A_i , $i = 1, \dots, p$, are (2×2) parameter matrices containing the autoregressive parameters. The

x_{it} variables are $(k_i \times 1)$ vectors of exogenous stationary variables including an intercept.

The two innovation outlier mechanisms are modelled by

$$v_{it} = \begin{cases} x'_{it}\beta_i + u_{it} & \text{if } x'_{i,t}\beta_i + u_{i,t} \geq 0 \\ 0 & \text{if } x'_{i,t}\beta_i + u_{i,t} < 0, \end{cases} \quad (4)$$

where $u_{it} \sim \text{NID}(\mathbf{0}, \sigma_{u_i}^2)$ for $i = 1, 2$, and where $\beta_i, \gamma_{ii}, \gamma_{i1}, \gamma_{2i}$, are $(k_i \times 1)$ parameter vectors. The δ_{12} and δ_{21} parameters model the cross-effect of the v_{it} variables on the series y_t , and therefore we can impose that $E[u_{1t}u_{2t}] = 0$. Note that this bivariate model would be suitable for modeling unemployment series again. Other applications of the model may require other variations of (4).

Common Large Innovations

The key feature of our bivariate CLEAR model, which makes it distinct from alternative multivariate nonlinear models, advocated in for example Philips (1991), Krolzig (1997), Diebold and Rudebusch (1996) and Kim and Nelson (1998), is that we introduce nonlinearity in y_{1t} and y_{2t} through an innovation-type variable. When this variable is absent, the model is linear. This feature makes it well-suited for studying common nonlinearity.

There are several interesting restricted versions of (3) and (4), which somehow concern common properties across the two time series y_{1t} and y_{2t} . An extreme case is that y_{1t} and y_{2t} have absolutely no common nonlinear features at time t . This implies that the δ_{12} and δ_{21} parameters are zero.

It may also be the case that only one of the δ parameters equals zero. Then, one of the series is not affected by the innovation-type variable of the other series. It may even be the case that the series have a common outlier generating mechanism. This occurs if a linear combination of y_{1t} and y_{2t} is a linear process, while the individual series are nonlinear. From (3) it is easy to see that the process $y_{2t} - \delta_{21}y_{1t}$ with $\delta_{21} \neq 0$ is a linear process if $v_{2t} = \delta_{21}v_{1t}$. The model (3) then simplifies to

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \sum_{i=1}^p A_i \begin{pmatrix} y_{1t-i} \\ y_{2t-i} \end{pmatrix} + \begin{pmatrix} x'_{1t}\gamma_{11} & x'_{2t}\gamma_{12} \\ x'_{1t}\gamma_{21} & x'_{2t}\gamma_{22} \end{pmatrix} + \begin{pmatrix} 1 \\ \delta_{21} \end{pmatrix} v_{1t} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \quad (5)$$

where v_{1t} is defined as in (4). Note that also the δ_{12} disappears from the model. This model assumes a common factor, and it can be viewed to correspond with the common

trend analysis in Stock and Watson (1988). A shock v_{1t} affects both series at the same time but the size of the shock may be different due to the δ_{21} parameter. A special case amounts to $\delta_{21} = 1$, where the difference between y_{1t} and y_{2t} is a linear process. In this case the size of the shock is also equal. Note that model (5) with (4) is nested in the general model (3) with (4).

Other interesting cases concern the occurrence of the shocks. An interesting bivariate CLEAR model, which is slightly different from (3) to (4) imposes that shocks v_{1t} and v_{2t} enter the vector autoregression simultaneously. This innovation outlier mechanism is given by

$$\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} = \begin{cases} \begin{pmatrix} x'_{1t}\beta_1 \\ x'_{2t}\beta_2 \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} & \text{if } x'_{1t}\beta_1 + u_{1t} \geq 0 \wedge x'_{2t}\beta_2 + u_{2t} \geq 0 \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \text{elsewhere.} \end{cases} \quad (6)$$

In this specification it is not possible that v_{1t} enters (3) while v_{2t} does not and *vice versa*. The size of the shocks may differ, of course, and therefore we do not label this model as one with common nonlinearity but one might call it synchronous nonlinearity. Note that this specification is not nested in (4), and hence it is not possible to use a standard likelihood ratio test to compare the models. Although the synchronous specification (6) seems to be a restricted version of (4), the maximum likelihood value of the model with (6) does not have to be smaller than the specification with (4). To compare the two different specifications, one can use tests for non-nested hypotheses, see for example Santos Silva (2001) and Gouriéroux and Monfort (1994) for a discussion.

It is of course possible to have synchronous innovation-type variables but no common variables. In this case the model equals (3) with $\delta_{12} = \delta_{21} = 0$ and with (6). The positive innovations occur at the same time but there is no relation between them. Although the synchronous model (6) is not nested in the general model (3) and (4), the common innovation outlier mechanism model (5) is nested within (3) and (6).

In the remainder of this paper we focus on estimation and inference for model specification (3) with (4) as the analysis of specification (3) with (6) can be done in a similar way. In our empirical section below we will of course compare both models.

3 Unconditional Inference

The v_{it} variables in our bivariate CLEAR model are unobserved and hence we can only make probability statements about their realizations. The probability that $v_{it} = 0$ for $i = 1, 2$ equals the probability that $x'_{it}\beta_i + u_{it} < 0$, that is,

$$\Pr[v_{it} = 0|x_{it}; \theta] = \int_{-\infty}^{-x'_{it}\beta_i} \frac{1}{\sigma_{u_i}} \phi(u_{it}/\sigma_{u_i}) du_{it} = \Phi\left(\frac{-x'_{it}\beta_i}{\sigma_{u_i}}\right) = \Phi_{it}, \quad (7)$$

where θ summarizes the model parameters, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the density function and the cumulative density function of the standard normal distribution, respectively. The probability that $v_{it} > 0$ is of course given by $(1 - \Phi_{it})$. These probabilities can be used to predict the presence of one or two innovation-type variables at time t .

To predict the size of these variables, one can use the expected value of v_{it} . As the v_{it} variables are censored, their expectations are not simply given by $x'_{it}\beta_i$. The expected value of the innovation-type variable v_{it} follows from

$$E[v_{it}|x_{it}; \theta] = x'_{it}\beta_i(1 - \Phi_{it}) + \sigma_{u_i}\phi_{it}, \quad \text{for } i = 1, 2, \quad (8)$$

with $\phi_{it} = \phi(-x'_{it}\beta_i/\sigma_{u_i})$, where we have used the results on the moments of truncated normal random variables, see Johnson and Kotz (1970, p. 81–83) and Gouriéroux and Monfort (1995, p. 483). The variance of the variable v_{it} is equal to

$$V[v_{it}|x_{it}; \theta] = \sigma_{u_i}^2(1 - \Phi_{it}) + x'_{it}\beta_i E[v_{it}|x_{it}; \theta] - E[v_{it}|x_{it}; \theta]^2 \quad (9)$$

for $i = 1, 2$, see Franses and Paap (2002) for details. Hence, the censored regressions introduce heteroskedasticity in the model.

4 Inference

In this section we discuss maximum likelihood estimation of the model parameters and conditional inference on the unobserved outlier mechanism.

4.1 Maximum Likelihood Estimation

The model parameters of the bivariate CLEAR model (3) to (4) are summarized by $\theta = (A_i, \dots, A_p, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \delta_{12}, \delta_{21}, \Sigma_\varepsilon, \beta_1, \beta_2, \sigma_{u_1}^2, \sigma_{u_2}^2)$. To estimate these parameters, we use a maximum likelihood estimator. For notational convenience we write (3) in matrix notation

$$Y_t = \sum_{i=1}^p A_i Y_{t-i} + \Gamma X_t + DV_t + \varepsilon_t, \quad (10)$$

where $Y_t = (y_{1t}, y_{2t})'$, $X_t = (x'_{1t}, x'_{2t})'$, and $V_t = (v_{1t}, v_{2t})'$. The matrices D and Γ contain the δ and γ parameters.

To derive the likelihood function, we first consider the conditional pdf of Y_t given its past $Y^{t-1} = \{Y_{t-1}, \dots, Y_1\}$, X_t , and given V_t . This function is given by

$$f(Y_t|Y^{t-1}, X_t, V_t; \theta) = \frac{1}{2\pi} |\Sigma_\varepsilon|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \left(Y_t - \sum_{i=1}^p A_i Y_{t-i} - X_t \Gamma - DV_t\right)' \Sigma_\varepsilon^{-1} \left(Y_t - \sum_{i=1}^p A_i Y_{t-i} - X_t \Gamma - DV_t\right)\right). \quad (11)$$

To obtain the pdf of Y_t conditional on Y^{t-1} but unconditional on V_t , we have to integrate over the unknown error processes u_{it} in the censored regressions. The pdf of Y_t can be decomposed in four parts corresponding to whether the v_{it} terms are zero or not

$$f(Y_t|Y^{t-1}, X_t; \theta) = g_{00} + g_{01} + g_{10} + g_{11}, \quad (12)$$

where

$$\begin{aligned} g_{00} &= \Pr[v_{1t} = 0|x_{1t}; \theta] \Pr[v_{2t} = 0|x_{2t}; \theta] f(Y_t|Y^{t-1}, X_t, V_t; \theta)|_{v_{1t}=0, v_{2t}=0} \\ g_{01} &= \Pr[v_{1t} = 0|x_{1t}; \theta] \int_{-x'_{2t}\beta_2}^{\infty} \frac{\phi(u_{2t}/\sigma_{u_2})}{\sigma_{u_2}} f(Y_t|Y^{t-1}, X_t, V_t; \theta)|_{v_{1t}=0, v_{2t}=x'_{2t}\beta_2+u_{2t}} du_{2t} \\ g_{10} &= \Pr[v_{2t} = 0|x_{2t}; \theta] \int_{-x'_{1t}\beta_1}^{\infty} \frac{\phi(u_{1t}/\sigma_{u_1})}{\sigma_{u_1}} f(Y_t|Y^{t-1}, X_t, V_t; \theta)|_{v_{1t}=x'_{1t}\beta_1+u_{1t}, v_{2t}=0} du_{1t} \\ g_{11} &= \int_{-x'_{1t}\beta_1}^{\infty} \int_{-x'_{2t}\beta_2}^{\infty} \frac{\phi(u_{1t}/\sigma_{u_1})}{\sigma_{u_1}} \frac{\phi(u_{2t}/\sigma_{u_2})}{\sigma_{u_2}} f(Y_t|Y^{t-1}, X_t, V_t; \theta)|_{v_{1t}=x'_{1t}\beta_1+u_{1t}, v_{2t}=x'_{2t}\beta_2+u_{2t}} du_{1t} du_{2t}. \end{aligned} \quad (13)$$

The log likelihood function is now simply the sum of the log of the unconditional pdfs, that is,

$$\mathcal{L}(Y^T|X_T; \theta) = \sum_{t=1}^T \ln f(Y_t|Y^{t-1}, X_t; \theta). \quad (14)$$

This log likelihood can be maximized using standard numerical optimization algorithms like Newton-Raphson. In this paper we opt for the BHHH algorithm of Berndt *et al.* (1974). To ensure that the covariance matrix Σ_ε is positive definite we write Σ_ε as $Q'Q$ where Q is a matrix with one of the off-diagonals equal to zero.

4.2 Conditional Inference

In Section 3 we already discussed inference on the unobserved variables in the model. This inference is unconditional on the value of Y_t and can therefore be used in forecast exercises. In this section, we consider probabilistic inference conditional on the observed values of the time series Y_t . First, we consider the conditional probabilities that the elements of the unobserved vector V_t equal zero. For instance, the conditional probability that $v_{1t} = v_{2t} = 0$ given Y_t and x_t equals

$$\Pr[v_{1t} = 0 \wedge v_{2t} = 0 | Y_t, X_t; \theta] = \frac{\Pr[v_{1t} = 0 | x_{1t}; \theta] \Pr[v_{2t} = 0 | x_{2t}; \theta] f(Y_t | Y^{t-1}, X_t, V_t; \theta)|_{v_{1t}=0, v_{2t}=0}}{f(Y_t | Y^{t-1}, X_t; \theta)}. \quad (15)$$

The other conditional probabilities are simply

$$\begin{aligned} \Pr[v_{1t} = 0 \wedge v_{2t} > 0 | Y_t, X_t; \theta] &= \frac{g_{01}}{f(Y_t | Y^{t-1}, X_t; \theta)} \\ \Pr[v_{1t} > 0 \wedge v_{2t} = 0 | Y_t, X_t; \theta] &= \frac{g_{10}}{f(Y_t | Y^{t-1}, X_t; \theta)} \\ \Pr[v_{1t} > 0 \wedge v_{2t} > 0 | Y_t, X_t; \theta] &= \frac{g_{11}}{f(Y_t | Y^{t-1}, X_t; \theta)}. \end{aligned} \quad (16)$$

Hence, the marginal probabilities that $v_{1t} = 0$ and $v_{2t} = 0$ given Y_t and X_t equal

$$\begin{aligned} \Pr[v_{1t} = 0 | Y_t, X_t; \theta] &= \Pr[v_{1t} = 0 \wedge v_{2t} = 0 | Y_t, X_t; \theta] + \Pr[v_{1t} = 0 \wedge v_{2t} > 0 | Y_t, X_t; \theta] \\ \Pr[v_{2t} = 0 | Y_t, X_t; \theta] &= \Pr[v_{1t} = 0 \wedge v_{2t} = 0 | Y_t, X_t; \theta] + \Pr[v_{1t} > 0 \wedge v_{2t} = 0 | Y_t, X_t; \theta]. \end{aligned} \quad (17)$$

These conditional probabilities indicate whether it is likely that a positive shock affects (one of) the series at time t . They can be used to give a business cycle chronology, see Franses and Paap (2002) for an example.

An estimate of the magnitude of the shock at time t follows from the conditional expectation of v_t given Y_t and x_t , that is,

$$\begin{aligned}
\mathbb{E}[v_{1t}|Y^t, X_t; \theta] &= \mathbb{E}[v_{1t}|v_{1t} = 0, Y^t, X_t; \theta] \Pr[v_{1t} = 0|Y^t, X_t; \theta] \\
&\quad + \mathbb{E}[v_{1t}|v_{1t} > 0, Y^t, X_t; \theta] \Pr[v_{1t} > 0|Y^t, X_t; \theta] \\
&= \mathbb{E}[v_{1t}|v_{1t} > 0 \wedge v_{2t} = 0, Y^t, X_t; \theta] \Pr[v_{1t} > 0 \wedge v_{2t} = 0|Y^t, X_t; \theta] \\
&\quad + \mathbb{E}[v_{1t}|v_{1t} > 0 \wedge v_{2t} > 0, Y^t, X_t; \theta] \Pr[v_{1t} > 0 \wedge v_{2t} > 0|Y^t, X_t; \theta],
\end{aligned} \tag{18}$$

and

$$\begin{aligned}
\mathbb{E}[v_{2t}|Y^t, X_t; \theta] &= \mathbb{E}[v_{2t}|v_{1t} = 0 \wedge v_{2t} > 0, Y^t, X_t] \Pr[v_{1t} = 0 \wedge v_{2t} > 0|Y^t, X_t; \theta] \\
&\quad + \mathbb{E}[v_{2t}|v_{1t} > 0 \wedge v_{2t} > 0, Y^t, X_t; \theta] \Pr[v_{1t} > 0 \wedge v_{2t} > 0|Y^t, X_t; \theta].
\end{aligned} \tag{19}$$

A typical element in the expression for these conditional expectations is for instance

$$\begin{aligned}
&\mathbb{E}[v_{1t}|v_{1t} > 0 \wedge v_{2t} > 0, Y^t, X_t; \theta] \Pr[v_{1t} > 0 \wedge v_{2t} > 0|Y^t, X_t; \theta] \\
&= [f(Y_t|Y^{t-1}, X_t; \theta)]^{-1} \int_{-x'_{1t}\beta_1}^{\infty} \int_{-x'_{2t}\beta_2}^{\infty} (x'_{1t}\beta_1 + u_{1t}) \frac{\phi(u_{1t}/\sigma_{u_1})}{\sigma_{u_1}} \frac{\phi(u_{2t}/\sigma_{u_2})}{\sigma_{u_2}} \\
&\quad f(Y_t|Y^{t-1}, X_t, V_t; \theta)|_{v_{1t}=x'_{1t}\beta_1+u_{1t}, v_{2t}=x'_{2t}\beta_2+u_{2t}} du_{1t} du_{2t}
\end{aligned} \tag{20}$$

and can be evaluated using the simplification in (28) in the Appendix below and the results in Rosenbaum (1961) or Maddala (1983, p. 368) concerning the expectation of a truncated bivariate normal random variable. The same holds for the other expectations in (18) and (19).

Residuals

Finally, to analyze the fit of the model, one may consider residuals. Residuals are defined as the difference between Y_t and the expectation of Y_t given Y^{t-1} and X_t evaluated in the parameter estimates. This expectation equals

$$\mathbb{E}[Y_t|X_t; \hat{\theta}] = \sum_{i=1}^p \hat{A}_i Y_{t-i} + X_t \hat{\Gamma} + \hat{D} \mathbb{E}[V_t|X_t; \hat{\theta}], \tag{21}$$

where the unconditional expectation of V_t is given in (8) for $i = 1, 2$. Note that the result in (9) implies that the residuals are heteroskedastic.

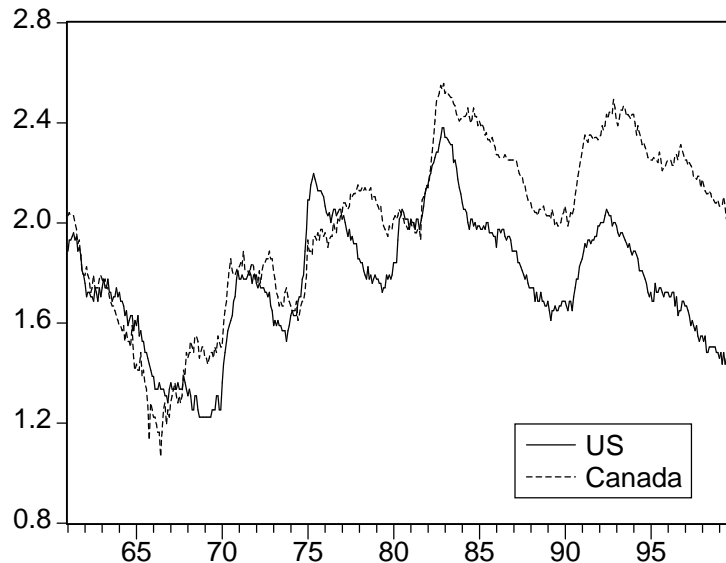


Figure 1: The logarithm of US and Canadian civilian unemployment rate, 1961.01–1999.12

5 Unemployment in the US and Canada

To illustrate our bivariate CLEAR model, we consider an application to seasonally adjusted monthly observed civilian unemployment rates of the United States [US] and Canada, 1961.01–1999.12. The results in Franses and Paap (2002) suggest that a univariate CLEAR is suitable for describing the US series, and hence our choice for a multivariate illustration. Figure 1 shows a graph of the natural log of both unemployment rate series. For both series we notice short periods characterized by large increases in unemployment, which are usually associated with recessions, and longer periods with a slow decline in the unemployment rate, which might match with expansions. The recessionary periods in both series seem to occur roughly in the same periods, except for the period between 1975 and 1980.

To describe both series, we consider the bivariate CLEAR model (3) and (4) for $i = 1, 2$. We examine whether the large increases in unemployment can be captured by two censored latent variables v_{1t} and v_{2t} . Following Franses and Paap (2002), we use, as explanatory variables for the censored latent effects, the lagged values of leading indicator

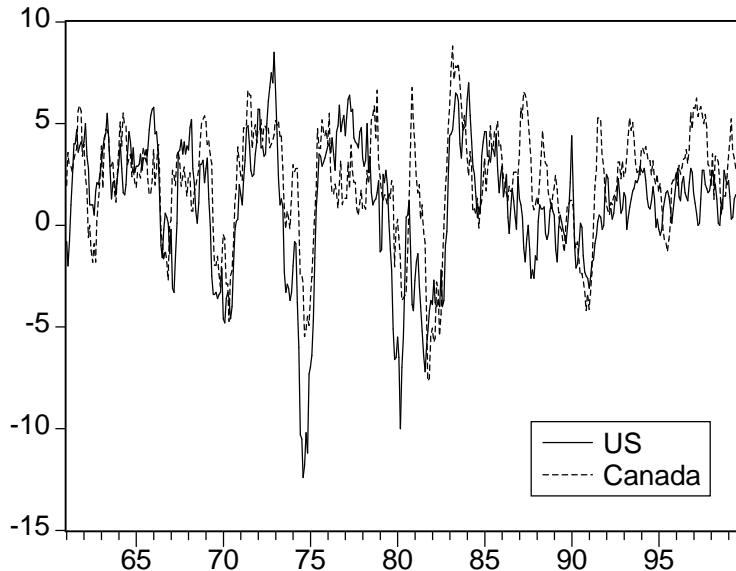


Figure 2: The US leading indicator US (XLI-2) and the bi-annual growth percentages of the Canadian composite leading indicator, 1961.01–1999.12

variables. Their results suggest that for the US the experimental leading indicator $XLI-2_t$ of Stock and Watson (1993) is suitable. For the Canadian series, we use the composite leading indicator series of Canada denoted by CLI_t , which is obtained from Datastream. This CLI_t series display exponential growth and has to be transformed to become approximately stationary. Based on our experience in Franses and Paap (2002), we opt for the bi-annual growth rate of CLI_t denoted as $\Delta_6 \log CLI_t$. Figure 2 displays the two leading indicator series, which clearly bear resemblance.

To estimate our bivariate CLEAR model, we need the order p and appropriate lags of the leading indicator variables. To determine this lag structure, we consider a $VARX(p)$ model which is equal to (3) with $v_{1t} = v_{2t} = 0$. We allow for lag orders up to 12 and we consider for each lag order the appropriate lag structure of the indicator variables. It turns out that for all values of p , $XLI-2_{t-1}$ and $\Delta_6 \log CLI_{t-2}$ produce the smallest determinant of the residual covariance matrix. Given this lag structure for the exogenous variables, the subsequent BIC value indicates that an AR order of 2 is best.

We estimate a bivariate CLEAR model of order 2, where Y_t consist of log unemployment rate of the US and Canada, respectively. The ML parameter estimates, with White's (1980) heteroskedastic-consistent standard errors in parentheses, are found to be

$$Y_t = \begin{pmatrix} 0.028 \\ (0.001) \\ -0.014 \\ (0.011) \end{pmatrix} + \begin{pmatrix} 0.759 & -0.064 \\ (0.046) & (0.050) \\ 0.056 & 0.844 \\ (0.057) & (0.063) \end{pmatrix} Y_{t-1} + \begin{pmatrix} 0.230 & 0.067 \\ (0.045) & (0.039) \\ -0.046 & 0.153 \\ (0.057) & (0.062) \end{pmatrix} Y_{t-2} \\ + \begin{pmatrix} -0.329 & -0.250 \\ (0.072) & (0.077) \\ -0.021 & -0.142 \\ (0.098) & (0.086) \end{pmatrix} \begin{pmatrix} 100 \times \text{XLI-2}_{t-1} \\ \Delta_6 \log \text{CLI}_{t-2} \end{pmatrix} + \begin{pmatrix} 1 & 0.364 \\ & (0.225) \\ 0.767 & 1 \\ (0.163) & \end{pmatrix} V_t + \varepsilon_t, \quad (22)$$

with

$$\Sigma_\varepsilon = \frac{1}{1000} \begin{pmatrix} 0.588 & 0.115 \\ 0.115 & 0.491 \end{pmatrix}, \quad (23)$$

and

$$v_{1t} = -0.073 - 2.355 \text{XLI-2}_{t-1} \times 100 + u_{1t}, \quad u_{1t} \sim \text{NID}(0, 0.050^2) \\ (0.014) \quad (0.443) \\ v_{2t} = -0.005 - 0.706 \Delta_6 \log \text{CLI}_{t-2} + u_{2t}, \quad u_{2t} \sim \text{NID}(0, 0.028^2). \quad (24) \\ (0.012) \quad (0.201)$$

The maximum log likelihood value of this model equals 2048.66.

To check whether the lag order selected by BIC is sufficient, we also estimate a model of order 3. The maximum likelihood value of this model equals 2051.55 and the corresponding LR statistic equals 5.78. This statistic is not significant when compared with the 95 percentile of a $\chi^2(4)$ distribution and hence the lag order of 2 does not seem underspecified. The LR statistic for the significance of the second order autoregressive parameters equals 62.89 and hence the model cannot be simplified to a first order model.

As the order of the estimated bivariate CLEAR model seems satisfactory, we try to analyze whether the estimated model can be simplified into a model with common aspects. First, we test whether we may delete the leading indicator variables from the VAR part (22), that is, we test whether the four γ parameters are zero. The LR statistic equals $-2 \times (2029.79 - 2048.66) = 37.74$, which is significant compared to the 95% percentile of a $\chi^2(4)$ distribution. Hence we cannot delete the two indicator variables from (22).

To analyze the role of the innovation generating censored regression, we test whether the Canadian shocks (v_{2t}) have no effect on US unemployment. The LR test for $\delta_{12} = 0$ equals $-2 \times (2047.06 - 2048.66) = 3.20$ and hence this restriction seems valid. The LR test for $\delta_{21} = 0$ equals $-2 \times (2039.99 - 2048.66) = 17.34$ and hence the innovation outlier mechanism of the US seems to affect Canadian unemployment. The joint restriction $\delta_{12} = \delta_{21} = 0$ is also rejected. In words, the cyclical pattern in Canadian unemployment bears similarities with that in the US, but often not the other way around.

Fixing $\delta_{12} = 0$, we may test for the presence of a common innovation outlier mechanism which corresponds to $v_{2t} = 0$. The LR statistic for $\beta_2 = 0$ and $\sigma_{u_2}^2 = 0$ equals $-2 \times (2024.47 - 2047.06) = 45.18$. As the total restriction contains a single one-sided alternative ($\sigma_{u_2}^2 = 0$ versus $\sigma_{u_2}^2 > 0$) and two two-sided alternatives ($\beta = 0$ versus $\beta \neq 0$), this test statistic is asymptotically $\frac{1}{2}\chi^2(2) + \frac{1}{2}\chi^2(3)$ distributed, see Wolak (1989, p. 19–20). The 95% percentile of this mixture distribution is 7.80 and hence we reject the presence of a common innovation outlier mechanism.

Finally, we consider a CLEAR model with a nonlinear synchronous cycle, that is, (3) with (6). Remember that this model is not nested in the general bivariate CLEAR specification estimated in (22) and (24) and hence a standard LR test does not apply. The value of the maximum log likelihood value equals 2036.86 and it is smaller than for the model (3) with (24). The nonlinear synchronous cycle representation is not preferred.

In sum, the test results suggest that there is no common outlier mechanism, but that the outlier mechanism of Canada does not affect the US series. The parameter estimates of the final model supported by our test results are

$$\begin{aligned}
Y_t = & \begin{pmatrix} 0.024 \\ (0.001) \\ -0.014 \\ (0.014) \end{pmatrix} + \begin{pmatrix} 0.769 & -0.052 \\ (0.046) & (0.040) \\ 0.058 & 0.857 \\ (0.014) & (0.069) \end{pmatrix} Y_{t-1} + \begin{pmatrix} 0.220 & 0.056 \\ (0.046) & (0.039) \\ -0.048 & 0.141 \\ (0.012) & (0.069) \end{pmatrix} Y_{t-2} \\
& + \begin{pmatrix} -0.327 & -0.306 \\ (0.071) & (0.054) \\ -0.028 & -0.178 \\ (0.124) & (0.133) \end{pmatrix} \begin{pmatrix} 100 \times \text{XLI-2}_{t-1} \\ \Delta_6 \log \text{CLI}_{t-2} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0.815 & 1 \\ (0.239) \end{pmatrix} V_t + \varepsilon_t, \quad (25)
\end{aligned}$$

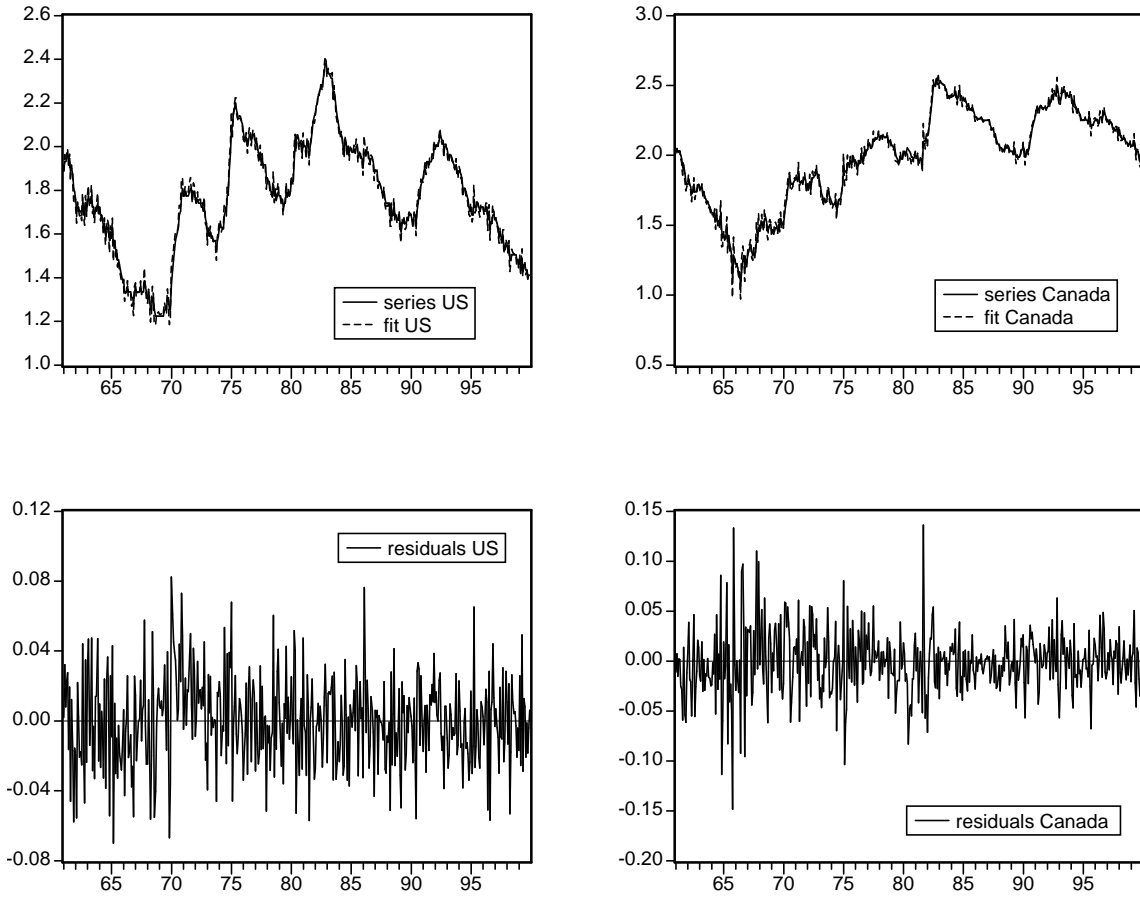


Figure 3: Fit and residuals based on (21)

with

$$\Sigma_\varepsilon = \frac{1}{1000} \begin{pmatrix} 0.600 & 0.114 \\ 0.115 & 0.504 \end{pmatrix}, \quad (26)$$

and

$$\begin{aligned} v_{1t} &= -0.064 - 2.278 \text{ XLI-2}_{t-1} \times 100 + u_{1t}, & u_{1t} &\sim \text{NID}(0, 0.047^2) \\ & (0.022) & (0.527) & \\ v_{2t} &= -0.010 - 0.672 \Delta_6 \log \text{CLI}_{t-2} + u_{2t}, & u_{2t} &\sim \text{NID}(0, 0.032^2) \end{aligned}, \quad (27)$$

$$\begin{aligned} & (0.020) & (0.221) & \end{aligned}$$

where again heteroskedastic-consistent standard errors appear in parentheses.

In Figure 3 we display the fit and the estimated residuals of the final model. Evidently, and even with not too many parameters, the multivariate CLEAR model provides

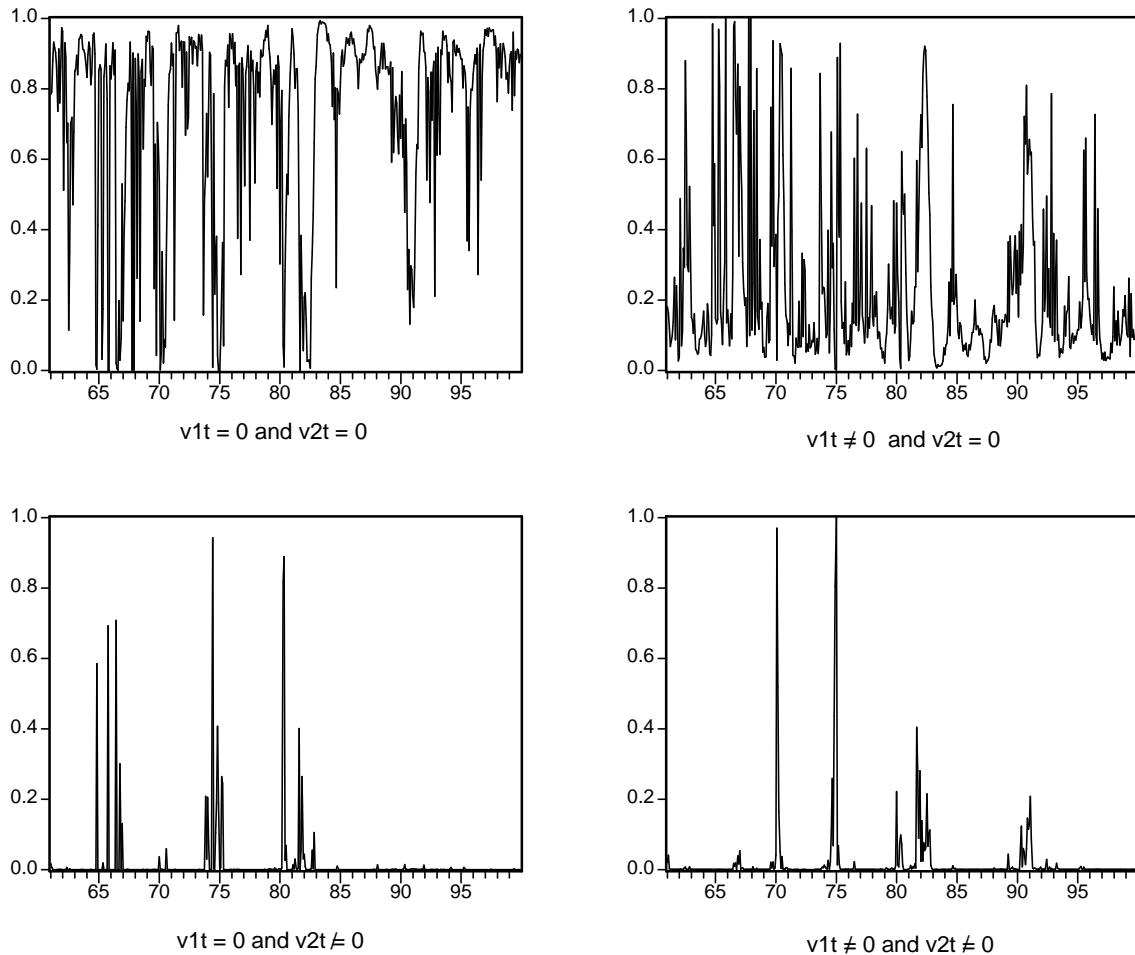


Figure 4: Estimated conditional probabilities $\Pr[v_{1t} \wedge v_{2t} | Y^t, X_t, \hat{\theta}]$

an excellent fit. As there are no obvious large residuals around the sharp increases in unemployment, the bivariate CLEAR model with the leading indicator variables seems to predict the start of recessionary periods rather well.

In Figure 4, we depict the estimated conditional probabilities. The left upper graph gives the probability that there are no positive innovations to both series, while the right bottom graph concern the probability that both variables both have large and positive innovations. These graphs show the familiar cyclical patterns around the oil crises and the recession in the 90s. The left bottom graph suggests that additional to the large shocks, the US only suffered from a few shocks which did not hit Canada, while the top

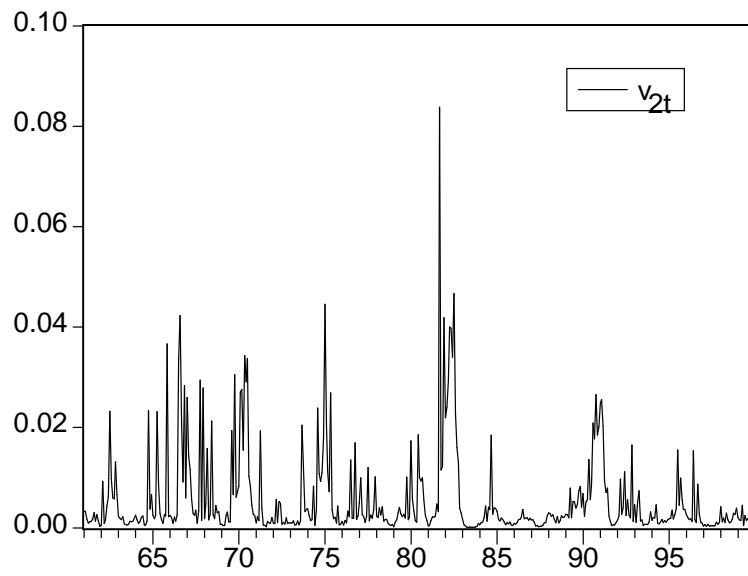
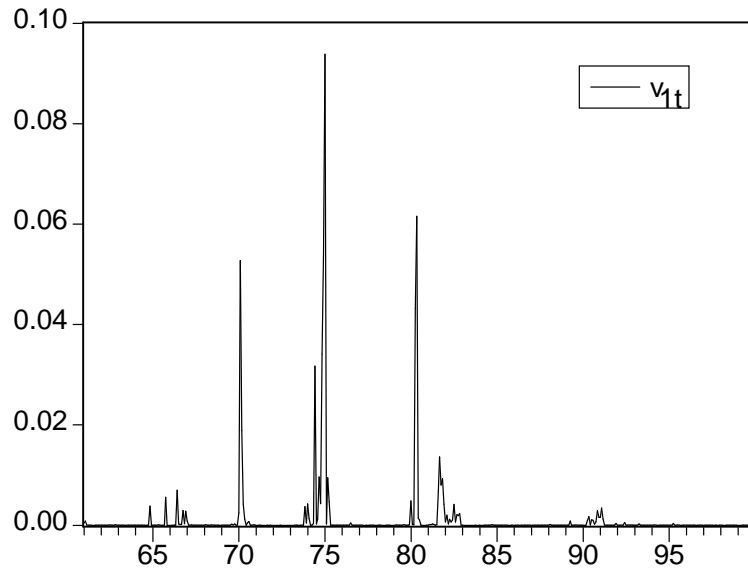


Figure 5: $E[v_{1t}|Y^t, X_t, \hat{\theta}]$ and $E[v_{2t}|Y^t, X_t, \hat{\theta}]$

right graph shows that Canada display much more idiosyncratic variation.

Finally, Figure 5 gives the estimated conditional expectations of the innovations. The top panel suggests that the increases in US unemployment were largest around the first oil crisis and relatively minor during the Iraq invasion in Kuwait. For Canada we observe rather different results, see the bottom panel of Figure 5, as here the second oil crisis clearly generated the largest positive shocks to unemployment.

6 Concluding Remarks

In this paper we put forward a new multivariate nonlinear time series model, which has the promising feature that it facilitates the analysis of common nonlinearity. This convenience is due to the fact that the model includes nonlinearity as additional variables. When linear combinations of these variables do not have such an innovation variable, these series share their nonlinearity. The model actually described nonlinear data very well, as could be seen from the illustration to US and Canadian unemployment. The illustration further suggested that nonlinear features of the US economy get reflected through Canadian unemployment, but not the other way around.

Even though the present paper only considered a bivariate illustration, it could already be seen that the multivariate censored latent autoregression is rather parsimonious. This seems a relevant feature when examining more than two time series. Furthermore, a smaller amount of parameters may also increase forecasting accuracy. We leave both issues for further research.

A Appendix

To simplify the computation of the three integrals in the log likelihood we use the following general result

$$\begin{aligned}
& u' \Sigma_u^{-1} u + (Su - a)' \Sigma_\varepsilon^{-1} (Su - a) \\
&= u' \Sigma_u^{-1} u + u' S' \Sigma_\varepsilon^{-1} Su - a' \Sigma_\varepsilon^{-1} Su - u' S' \Sigma_\varepsilon^{-1} a + a' \Sigma_\varepsilon^{-1} a \\
&= u' (\Sigma_u^{-1} + S' \Sigma_\varepsilon^{-1} S) u + a' \Sigma_\varepsilon^{-1} Su - u' S' \Sigma_\varepsilon^{-1} a + a' \Sigma_\varepsilon^{-1} a \\
&= (u - b)' (\Sigma_u^{-1} + S' \Sigma_\varepsilon^{-1} S) (u - b) - b' (\Sigma_u^{-1} + S' \Sigma_\varepsilon^{-1} S) b + a' \Sigma_\varepsilon^{-1} a
\end{aligned} \tag{28}$$

with $b = (\Sigma_u^{-1} + S' \Sigma_\varepsilon^{-1} S)^{-1} S' \Sigma_\varepsilon^{-1} a$. This result allows us to write the product of the three normal pdfs in the three integrals which both depend on u_t as the product of a normal pdf which depends on u_t and a remainder that does not depend on u_t . Hence, each integral can be expressed in terms of a CDF of a bivariate normal distribution.

References

- Anderson, H. and F. Vahid (1998), Testing multiple equation systems for common non-linear components, *Journal of Econometrics*, **84**, 1–36.
- Berndt, E., B. Hall, E. Hall, and J. Hausman (1974), Estimation and Inference in Non-linear Structural Models, *Annals of Economic and Social Measurement*, **3**, 653–665.
- Diebold, F. and G. Rudebusch (1996), Measuring Business Cycles: A Modern Perspective, *Review of Economics and Statistics*, **78**, 67–77.
- Franses, P. and R. Paap (2002), Censored Latent Effects Autoregression, with an Application to US Unemployment, *Journal of Applied Econometrics*, forthcoming.
- Franses, P. and D. van Dijk (2001), *Nonlinear Time Series Models in Empirical Finance*, Cambridge University Press, Cambridge.
- Gourieroux, C. and A. Monfort (1994), Testing Non-Nested Hypotheses, in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, vol. IV, chap. 44, North-Holland, Amsterdam, pp. 2583–2637.
- Gourieroux, C. and A. Monfort (1995), *Statistics and Econometric Models*, vol. 2, Cambridge University Press, Cambridge.
- Granger, C. and T. Teräsvirta (1993), *Modelling Nonlinear Economic Relations*, Oxford University Press, Oxford.
- Johnson, N. and S. Kotz (1970), *Distributions in Statistics: Continuous Univariate Distributions*, Houghton Mifflin, Boston.
- Kim, C.-J. and C. Nelson (1998), Business Cycle Turning Points, a New Coincident Index, and Tests of Duration Dependence on a Dynamic Factor Model with Regime Switching, *Review of Economics and Statistics*, **80**, 188–201.
- Krolzig, H.-M. (1997), *Markov Switching Vector Autoregressions: Modelling, Statistical Inference, and an Application to Business Cycle Inference*, Springer, Berlin.

- Maddala, G. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, vol. 3 of *Econometric Society Monographs*, Cambridge University Press, Cambridge.
- Philips, K. (1991), A Two-Country Model for Stochastic Output with Changes in Regime, *Journal of International Economics*, **31**, 121–142.
- Rosenbaum, S. (1961), Moments of a Truncated Bivariate Normal Distribution, *Journal of the Royal Statistical Society B*, **23**, 405–408.
- Santos Silva, J. (2001), A Score Test for Non-Nested Hypotheses with Applications to Discrete Data Models, *Journal of Applied Econometrics*, **16**, 577–597.
- Stock, J. and M. Watson (1988), Testing for Common Trends, *Journal of the American Statistical Association*, **83**, 1097–1107.
- Stock, J. and M. Watson (1993), A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience, in J. Stock and M. Watson (eds.), *Business Cycles, Indicators, and Forecasting*, chap. 2, The University of Chicago Press, Chicago, pp. 95–147.
- White, W. (1980), A Heteroscedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity, *Econometrica*, **48**, 817–838.
- Wolak, F. (1989), Local and Global Testing of Linear and Nonlinear Inequality Constraints in Nonlinear Econometric Models, *Econometric Theory*, **5**, 1–35.