REGIME-SWITCHING MODELS

A Guide to the Bank of Canada Gauss Procedures

by
Simon van Norden
and Robert Vigfusson

Bank of Canada   Banque du Canada

Regime-Switching Models

A Guide to the Bank of Canada
Gauss Procedures

Simon van Norden
E-mail: svannorden@bank-banque-canada.ca

Robert Vigfusson
E-mail: rvigfusson@bank-banque-canada.ca

International Department
Bank of Canada
Ottawa, Ontario
Canada K1A 0G9

Authors' Note

The Gauss procedures described in this paper are available from the Bank of Canada's Library Website (http://www.bank-banque-canada.ca/library) under working papers or by anonymous FTP to ftp.bank-banque-canada.ca in the directory pub/vigfuss.

The code in this paper is being released free of charge. In exchange, we ask that its use be acknowledged in any published work, including working papers and paid research papers.

The computer programs, documentation and all other information in this working paper are provided for your information only and for use entirely at your own risk. We do not warrant their accuracy nor their fitness for any purpose whatsoever.

Abstract

This paper is a user's guide to a set of Gauss procedures developed at the Bank of Canada for estimating regime-switching models. The procedures can estimate relatively quickly a wide variety of switching models and so should prove useful to the applied researcher. Sample program listings are included.

Résumé

La présente étude constitue un guide d'utilisation d'un ensemble de procédures de Gauss mises au point à la Banque du Canada en vue de l'estimation des modèles à changement de régime. Ces procédures permettent d'estimer de façon assez rapide une vaste gamme de modèles à changement de régime et devraient s'avérer utiles pour la recherche appliquée. Des échantillons de programmes sont inclus dans l'étude.

Contents

1.0 Introduction

This paper is a manual for using a set of Gauss procedures developed at the Bank of Canada to estimate regime-switching models. In these models, the observable dependent variable's behavior is state-dependent. The state (which itself is unobservable) determines the process that generates the dependent variable.

The relationship between this period's state and the previous periods' states determines what set of procedures one uses. In a first-order Markov model, hereafter referred to as the Markov model, the current state is dependent on only the last period's state. In a simple-switching model, the current state is independent of the previous periods' states.

The Bank of Canada procedures are for two state models with a single dependent variable. The model can be either a Markov model or a simple-switching model. The procedures allow switching in the parameters for any number of explanatory variables, including non-constant transition probabilities.

Several Bank of Canada working papers have used earlier versions of this code. Van Norden (1996) and van Norden and Schaller (1993) both use the simple-switching code, while Vigfusson (1996) uses the Markov switching code. A Bank of Canada working paper by Gable, van Norden, and Vigfusson (1995) provides more detail on the analytic gradients for the Markov switching model.

The Gauss procedures described in this paper are available from the Bank of Canada's Library Website (http://www.bank-banque-canada.ca/library) under working papers or by anonymous FTP to ftp.bank-banque-canada.ca in the directory pub/vigfuss. Use of this code requires Gauss version 3.1. The maximum-likelihood estimation also requires Maxlik version 3. While Gauss itself is available on many platforms, these procedures have been developed and used on Unix workstations. Some conversion may be necessary to use them on other operating systems. For example, several of the file names are longer than the DOS name length limit.

The rest of the paper is organized along the following lines. Section 2 describes the procedures used for estimating simple-switching models. Section 3 describes the procedures used for estimating Markov switching models. Section 4 describes specification testing to be used after estimation of a switching model. Section 5 contains a bibliography.

1.1 What Is New in This Version

There are some large differences between this and earlier releases of the switching procedures. A library file has been written making the procedures easier to use. The Markov switching code has seen dramatic changes. Most are related to the development of a fast algorithm for calculating full sample (or smoothed) probabilities (Kim 1994). These probabilities are necessary to make practical use of the analytic gradients that are discussed in Section 3.

2.0 Simple Switching

2.1 Maximum-Likelihood Methods

This section documents the procedures used for maximum-likelihood estimation of simple-switching and normal-mixture models. These procedures calculate the likelihood function and its as well as starting parameter estimates.

There are four files containing procedures:

SWLI_NEW.G contains likelihood function and gradients for the general switching model.

SWINIT2.G produces starting parameter values for the above.

SWCONTAM.G contains likelihood function and gradients for restricted model with identical means in both regimes.

SWINTCON.G produces starting parameter values for the above.

SWLI_NEW.G and SWINIT2.G are for the most general model, which has the form

$$Y_{1t} = X_{1t} \cdot \beta_1 + \varepsilon_{1t}$$
$$Y_{2t} = X_{2t} \cdot \beta_2 + \varepsilon_{2t}$$
$$Y_{3t} = X_{3t} \cdot \beta_3 + \varepsilon_{3t}$$
$$(Y_t = Y_{1t}) \Leftrightarrow Y_{3t} < 0 \tag{EQ 1}$$

$$\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t} \sim N\left(0, \begin{bmatrix} \sigma_1 & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right)$$

Y_t is our observable dependent variable, which is generated by a mixture of different regimes, captured by the unobservables (Y_{1t}, Y_{2t}). Y_{3t} is a latent variable that perfectly classifies Y_t into the two regimes, and X_{3t} represents whatever extra information we have to make this classification. Conditioning only on X_{3t}, the probability of being in regime 1 at t is $\Phi(-X_{3t} \cdot \beta_3)$,[1] whereas if we also condition on Y_t it becomes

---

1. Alternatively, one could use logit equations to determine the probability of being in the different regimes. See Section 2.1.1 for more information.

$$\frac{\Phi(-X_{3t} \cdot \beta_3) \cdot \phi\left(\frac{Y_t - (X_{1t} \cdot \beta_1)}{\sigma_1}\right)\frac{1}{\sigma_1}}{\Phi(-X_{3t} \cdot \beta_3) \cdot \phi\left(\frac{Y_t - (X_{1t} \cdot \beta_1)}{\sigma_1}\right)\frac{1}{\sigma_1} + \Phi(X_{3t} \cdot \beta_3) \cdot \phi\left(\frac{Y_t - (X_{2t} \cdot \beta_2)}{\sigma_2}\right)\frac{1}{\sigma_2}} \qquad \text{(EQ 2)}$$

where ϕ is the standard normal p.d.f. and Φ is the corresponding c.d.f. In the simplest case, $X_{1t} = X_{2t} = X_{3t} = 1$ and we just have a mixture of two normal distributions $N(\beta_1, \sigma_1)$ and $N(\beta_2, \sigma_2)$.

SWLI_NEW.G contains the likelihood and gradient procedures swli and swgrad. It follows the standard syntax required by maxlik and assumes that the data in the data matrix are of the form $\begin{bmatrix} Y & X_1 & X_2 & X_3 \end{bmatrix}$ and that the parameter vector Θ is of the form

$$\Theta' = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \sigma_1 & \sigma_2 \end{bmatrix} \qquad \text{(EQ 3)}$$

The user must set the global variables _n1, _n2 and _n3 to tell the procedure the dimensions of the elements in the data set and Θ.

SWINIT2.G has the procedure binitial for generating initial values for the parameters and requires the same three global variables to be set before it is called. It divides the sample into positive and negative values of Y_t, then regresses the positive on X_1 and the negative on X_2 to generate β_1, β_2.[2] β_3 comes from the regression of a scaled version of Y on X_3. This seems to work well enough if Y has a mean very close to zero. Note that it omits rows in the data set containing missing values. The procedure is called using the syntax
theta = binitial(*Xdata*)
where *Xdata* has the same form as the data matrix for swli and swgrad.

SWCONTAM.G and SWINTCON.G estimate a special case of the general model described above: where $X_1 = X_2$ *and* we wish to restrict $\beta_1 = \beta_2$. This means the model reduces to

_____

2. The ambitious user would want to program Goldfeld and Quandt (1973)'s method of moments estimator for this kind of thing, but that is definitely non-trivial. However, this procedure seems to work well for data with a mean close to zero.

$$Y_t = X_t \cdot \beta + \varepsilon_t$$

$$Y_{3t} = X_{3t} \cdot \beta_3 + \varepsilon_{3t}$$

$$\varepsilon_t \sim N(0, \sigma_1) \Leftrightarrow Y_{3t} < 0 \qquad \text{(EQ 4)}$$

$$\varepsilon_t \sim N(0, \sigma_2) \Leftrightarrow Y_{3t} \geq 0$$

$$E(\varepsilon_t, \varepsilon_{3t}) = 0$$

This says that Y and X now have the usual linear relationship, aside from a particular kind of heteroscedasticity; some errors are generated by a high-variance regime and some from a low.[3] The main use of this model is in tests of non-switching models against switching alternatives. For example, suppose we wanted to test the hypothesis that

$$Y_t = X_t \cdot \beta + \varepsilon_t$$

$$\varepsilon_t \sim N(0, \sigma) \qquad \text{(EQ 5)}$$

against the alternative of (EQ 1). Econometrically, this is difficult, since some of the parameters of (EQ 1) are unidentified under the null.[4] Typically, however, what we care about is whether there is switching in means (i.e., whether $\beta_1 = \beta_2$), since evidence switching due solely to some form of heteroscedasticity may be cured by some other model of heteroscedasticity. By estimating (EQ 4), we have a model where we allow for heteroscedasticity and where all parameters are identified under the null. Therefore, we can use standard likelihood-ratio tests of (EQ 1) versus (EQ 4).

SWCONTAM.G contains the procedures swecli and swecgrad, which are analogues of swli and swgrad in SWLI_NEW.G. The only functional differences are that they assume the data matrix is of the form $\begin{bmatrix} Y & X & X_3 \end{bmatrix}$ and the parameter vector Θ is of the form

$$\Theta' = \begin{bmatrix} \beta & \beta_3 & \sigma_1 & \sigma_2 \end{bmatrix} \qquad \text{(EQ 6)}$$

Accordingly, only _n1 and _n3 need be set before the procedure is called, and _n2 is ignored. The same applies to the procedure becinit in SWINTCON.G, which is analogous to binitial in SWINIT2.G. However, becinit uses an iterative weighted least squares (WLS) method of approximating the robust estimator of β and then regresses the final weights on X_3 to estimate β_3.

---

3. In the heteroscedasticity literature, this kind of model is formally known as the error contamination model – hence the names of the files and procedures.

4. Recent work has been done on this issue for the Markov regime-switching case (Hansen 1992).

2.1.1 Logit Probabilities

There is an alternative version of the above programs, where the above probit weights are replaced with logit weights. For the switching regression model of (EQ 1), this means

$$Pr(Y_t = Y_{2t}) = \Phi(X_{3t} \cdot \beta_3) \qquad \text{(EQ 7)}$$

is replaced with

$$Pr(Y_t = Y_{2t}) = \frac{1}{1 - e^{-X\beta_3}} \qquad \text{(EQ 8)}$$

and similarly, for the error contamination model of (EQ 4),

$$Pr(\varepsilon_t \sim N(0, \sigma_2)) = \Phi(X_{3t} \cdot \beta_3) \qquad \text{(EQ 9)}$$

is replaced with

$$Pr(\varepsilon_t \sim N(0, \sigma_2)) = \frac{1}{1 - e^{-X\beta_3}} \qquad \text{(EQ 10)}$$

Note that in both cases, the probability is still bounded asymptotically between 0 and 1, and that the probability of observing regime *2* increases with $X\beta_3$.

This slight change allows one to examine the sensitivity of estimation results to minor changes in the weights accorded to each regime, as the logit weights converge to their limits more slowly than do the probit, so they discount unlikely regimes less heavily. In addition, the exponential expressions given in (EQ 8) and (EQ 10) can be calculated more quickly and accurately than the cumulative normal distribution function, which does not allow a closed-form solution and usually requires a series of iterative approximations. Finally, the two weighting systems are similar enough that the existing procedures for generating initial parameter estimates for $(\beta_1, \beta_2, \beta_3, \sigma_1, \sigma_2)$ may be used without modification for either probit or logit models.

Procedures for use with these new models may be found in the files SWLILOGT.G and SWCONLGT.G. Each file contains procedures returning the log-likelihood function and its gradient. The names and formats of the procedures are identical to those used in the original probit procedures (swli and swgrad for SWLILOGT, swecli and swecgrad for SWCONLGT). Only the library file SWITCH.LCG needs to be modified for a user to switch from using the probit model to the logit model. One comments out the library references to the original probit procedures and removes the comments from the logit procedures.

The only lines that differ from the probit version are those calculating the weights given to each regime (phi3), and those calculating the derivative with respect to β_3, which is given by the formula

$$\frac{d}{d\beta_3}ln(L(\beta_1, \beta_2, \beta_3, \sigma_1, \sigma_2)) = \frac{(\phi(\beta_2) - \phi(\beta_1))X_3}{\{\phi(\beta_2) + \phi(\beta_1)e^{-X_3\beta_3}\} \cdot (1 + e^{X\ \beta_3})} \qquad \text{(EQ 11)}$$

where $\phi(\beta_1)$, $\phi(\beta_2)$ are the probability density functions of the residuals from regime 1 and 2 respectively. Note that the accuracy of the gradient procedures has been checked against the numerical derivatives of the likelihood function using the Gauss procedure maxgrd.

2.2 EM algorithm

Hartley (1978) described an EM algorithm to estimate the parameters of a switching regression. The algorithm is an alternative to maximum-likelihood estimation (MLE), and simply uses a combination of ordinary least squares (OLS) and WLS. The EM algorithm converges more slowly than MLE near the final estimates, but is thought to be more robust to poor starting values. This should make it a natural complement to existing MLE procedures. This algorithm is now available as a set of Gauss procedures.

We will first describe the application of the EM algorithm to the estimation of switching regression systems. The remainder of the section lists three Gauss procedures that implement this estimation technique. The first procedure, SWEM.G, performs one iteration of the EM algorithm. The second procedure, SWECMEM.G, is the analogue of SWEM.G for the case where we restrict the coefficients in the two regimes to being identical. The final procedure, SWEMMLE.G, both controls the EM estimation and integrates it with the existing MLE procedures.

2.2.1 The EM Algorithm

Suppose an observable variable (y_t) is generated by two different latent unobservable variables (y_{t1}, y_{t2}). y_{t3} is also a latent variable and classifies y_t in either regime.

$$y_{t1} = x'_{t1} \cdot \beta_1 + \varepsilon_{t1}, \ \varepsilon_{t1} \sim N(0, \sigma^2) \qquad (1)$$

$$y_{t2} = x'_{t2} \cdot \beta_2 + \varepsilon_{t2}, \ \varepsilon_{t2} \sim N(0, \sigma^2) \qquad (2)$$

$$y_{t3} = x'_{t3} \cdot \beta_3 + \varepsilon_{t3}, \ \varepsilon_{t3} \sim N(0, 1) \qquad (3)$$

$$y_t = y_{t1} \quad \text{if} \quad y_{t3} \leq 0$$

$$y_t = y_{t2} \quad \text{if} \quad y_{t3} > 0$$

The EM algorithm is an iterative method for estimating such systems that relies on simple OLS and WLS regressions. Given starting values for $(\beta_1, \beta_2, \beta_3, \sigma_1, \sigma_2)$, we begin by calculating weights for a WLS regression.[5] These weights are computed as:

$$w_1(y_t) = \lambda_t \cdot [(f_1(y_t))/(h(y_t))] \tag{4}$$

$$w_2(y_t) = (1 - \lambda_t) \cdot [(f_2(y_t))/(h(y_t))] \tag{5}$$

where $f_i(y_t)$ is the normal probability density function for each regime:

$$f_i(y_t) = \frac{1}{(2\pi)^{1/2}\sigma_i} exp\left[\frac{-1}{2\sigma_i^2}(y - (x' \cdot \beta_i))^2\right] \tag{6}$$

λ_t is the cumulative density function of ε_{3t}, which represents the probability that $y_t = y_{t1}$. It is calculated as:

$$\lambda_t = \Phi(-x_{t3} \cdot \beta_3) \tag{7}$$

where Φ is the standard normal cumulative density function. Finally $h(y_t)$ is the probability density function of the observed variable (y_t) and is calculated as:

$$h(y_t) = \lambda_t \cdot f_1(y_t) + (1 - \lambda_t) \cdot f_2(y_t) \tag{8}$$

Hartley(1979) suggests the following WLS regression to obtain estimates of β_i (i=1,2)

$$\beta_i = [X'_i \cdot W_i \cdot X_i]^{-1}[X_i' \cdot W_i \cdot y] \tag{9}$$

$$W_i = diag\{w_{1i}, ..., w_{Ti}\}, \quad i = 1, 2 \tag{10}$$

β_3 is computed using OLS:

$$\beta_3 = [X_3'X_3]^{-1}X_3'\xi_3 \tag{11}$$

where ξ_{t3} is the conditional expectation of y_{t3} given y_t.

$$\xi_{t3} = E[y_{t3}/y_t] = x_{t3}'\beta_3 - \left(w_1(y_t) \cdot \frac{f_3(0)}{\lambda_t}\right) + \left(w_2(y_t) \cdot \frac{f_3(0)}{1 - \lambda_t}\right) \tag{12}$$

Finally, to calculate the variance estimates $(\sigma_i^2; i = 1, 2)$ the following WLS regression is employed:

$$\sigma_i^2 = \left[\frac{1}{\Sigma w_i(y_t)}\right] \cdot (y - X_i\beta_i)' \cdot W_i \cdot (y - X_i\beta_i) \tag{13}$$

---

5. The aforementioned procedure, SWINIT2.G, could be used to find starting values.

These new estimates $(\beta_1, \beta_2, \beta_3, \sigma_1, \sigma_2)$ can be used as new starting values, and the calculations of equations (1)-(13) repeated. It can be shown that this iterative estimation process converges to the maximum-likelihood estimates of $(\beta_1, \beta_2, \beta_3, \sigma_1, \sigma_2)$.

2.2.2 Gauss Programs

A single iteration of the EM algorithm can be found in SWEM.G. It takes a data matrix and a set of parameter estimates, calculates equations (1)-(13), and returns the new parameter estimates. This may be used in a simple loop for model estimation.

The MLE procedures for estimating switching regressions also consider the special case where we impose the restriction $\beta_1 = \beta_2 = \beta$ in estimation. Imposing this restriction in the EM algorithm requires only slight modifications to the algorithm and procedure described above. In particular, we would now replace the 2 WLS regressions with a single WLS regression using weights $W_1/\sigma_1^2 + W_2/\sigma_2^2$ to estimate β . The unweighted residuals from this regression are then weighted by w_1 and w_2 to generate estimates of σ_1 and σ_2 , respectively. These modifications are embodied in the Gauss procedure SWEC-MEM.G, which is otherwise analogous to SWEM.G.

In order to use the above procedures, all one needs is a simple loop that updates the parameter estimates at each pass and tests them for convergence. To make the most of the relative strengths of the EM and MLE procedures, however, one might wish to begin estimation with the EM procedures, to take advantage of their greater robustness. As these estimates begin to converge, one might then want to change to MLE because of its faster convergence.[6] The procedure listed below, SWEMMLE.G, incorporates both of these features. Furthermore, it is structured to return the same information as the Gauss maxlik procedure, which allows existing MLE programs to be easily adapted to joint EM-ML estimation.

These procedures are part of the library file SWITCH.LCG. The output produced is controlled by the same global variable "__output" used to control maxlik(). There are four other global variables that control other features of the procedure. One (_emprint) controls whether the final MLE results are printed. Another (_emecm) controls whether the unrestricted or restricted ($\beta_1 = \beta_2$) versions of the model are estimated. The other two control when estimation is passed from the EM to the MLE procedures. This occurs as soon as either a maximum number (_emmaxit) of EM iterations have been completed, or when the relative change in each component of the parameter vector is less than some fraction (_emconvg).

---

6. Even more generally, if the MLE "gets stuck," one might want to change to the EM procedure for several iterations before returning to the MLE. Unfortunately, we could find no way of aborting the Gauss maxlik procedure to do this in a flexible way. However, we are talking to Aptech Systems about ways to accomplish this, and may incorporate such features in a future version of SWEMMLE.G.

2.3 A Sample Program For Simple-Switching Models

To give the user an understanding of how these procedures can be used, let's look at sample program swbasic.prg.

The first thing that is required is to have a library statement that references the maxlik and switch.lcg libraries. After this library statement, the program can then call all the procedures mentioned above.

```
library maxlik,switch.lcg
_swclass = "PROBIT";
```

The next step is to load in the data. The exact method of loading in the data is dependent on whether the data are in ASCII format or in the form of a Gauss data set. After the data have been read in, they must be in the form of: dependent variable, explanatory variables regime 1, explanatory variables regime 2, and then the probability of being in regime 1's variables.

```
@ 3)Load the data set. @
/* The dataset SWREVBUB.DAT has the format
VWRETD EXCESS BUBBLE_A BUBBLE3A BUBBLE_B BUBBLE_C BUBBLE3C */
open fdata = ^dataset;      nobs = rowsf(fdata);    nvars = colsf(fdata);
"Loading ";; nobs;; " observations and ";; nvars;;
" variables from ";; dataset; ?; ?;
crsp = readr(fdata,nobs);
?;"Using EXCESS and BUBBLE_A.";;
swd = crsp[.,2 3];nobs = rows(swd);
"This leaves ";;  nobs;; " observations.";
@============================================================@
swd = swd[.,1]~ones(nobs,1)~swd[.,2]~ones(nobs,1)~swd[.,2]~ones(nobs,1)
~((swd[.,2].^2)-meanc(swd[.,2].^2));
```

The global variables now need to be set. In particular, the globals _n1, _n2, and _n3 must be set. There are others that are optional, like the maxlik variables. Consult the user's manual for information on them.

```
@ 4)Choose Settings for the SWEM procs if desired @
 _emmaxit=100;_emconvg=0.002;

@ 5)Choose settings for the MAXLIK proc;  @
__row=0;  _mlalgr=5;  _mlstep=2;   _mlcovp=2;  _mlndmth="NEWTON";  __output=0;
_mlgtol=1E-5; _mlmiter=500;
_n1=2;  _n2=2;  _n3=2;_emprint=1;_emecm=0;
__title="A Switching Regression Model";
_mlparnm = "B_S0 " | "B_Sb " | "B_C0 " | "B_Cb " | "B_q  " | "B_qb2" | "S_S  "
| "S_C  ";
```

The next step is to provide some initial starting values.

```
let beta0 = 0.0074, -0.0146, -0.0298, -0.0533, -1.3723, 2.4154, 0.0405, 0.1334;
```

Finally, there is a call to the swemmle procedure to estimate the model.

```
{xswp,llfswp,g,h,retcode}=swemmle(beta0,swd);
```

3.0 Markov Regime Switching

The section on Markov regime-switching models consists of two parts. First, the procedures used for maximum likelihood are explained. Second, an example program is given.

The Markov switching code has had four substantial improvements. First, Hamilton's original code was generalized and made faster. Second, analytic gradients were developed that made estimation even faster. Third, a C procedure was written to further speed up estimation. Finally, Hamilton's EM algorithm (1994) has been programmed in Gauss. This algorithm can reduce greatly the time consumed by bad starting values in comparison with the other methods.

Hamilton's textbook *Time Series Analysis* (1994) devotes a chapter to Markov regime-switching models. Anyone interested in working with Markov switching models should consult it or another textbook for more details on the Markov switching models.

3.1 Maximum-Likelihood Estimation

The general regime-switching model that we consider describes the generation of a single variable y_t by *2* distinct states. In any given state *i*, y_t is generated by a linear regression model

$$y_t = x_t \cdot \beta_i + \varepsilon_{it}, \varepsilon_{it} \sim \text{i.i.d. } N(0, \sigma_i) \qquad \text{(EQ 12)}$$

where x_t is a vector that is exogenous with respect to ε_{it}. The states are assumed to follow a first-order Markov process, with transition probabilities that may vary over time according to the formula[7]

$$P(s_t = i \,|\, s_{t-1} = j) = \Phi(z_t \cdot \gamma_{ij}) \qquad \text{(EQ 13)}$$

where Φ is the standard normal cumulative distribution. The probability of being in the given state is dependent on last period's state. This is what distinguishes this from "simple" mixtures of distributions. In simple mixture models the probability of being in a given state is independent of last period's state. In the two-state model this "simple"- switching model is nested into the Markov switching model when

$$Pr(S_t = 0 | (S_{t-1} = 0)) = 1 - Pr(S_t = 1 | (S_{t-1} = 1)) \qquad \text{(EQ 14)}$$

---

7. We assume that z_t is also exogenous with respect to ε_{it}.

3.1.1 The Likelihood Function

The dependence of this period's state on the previous period's state has a major effect on the difficulty of estimating the parameters of the mixture model. To understand why, note that the p.d.f. of Y_t, given $(\mu_0, \mu_1, \sigma_0, \sigma_1, p)$ in the simple mixture model, can be written as

$$Pr(S_t = 0) \cdot pdf(Y_t|(S_t = 0)) + Pr(S_t = 1) \cdot pdf(Y_t|(S_t = 1)) \qquad \text{(EQ 15)}$$

$$p \cdot pdf(Y_t|(S_t = 0)) + (1 - p) \cdot pdf(Y_t|(S_t = 1)) \qquad \text{(EQ 16)}$$

$$p \cdot \frac{\phi\left(\dfrac{Y_t - \mu_0}{\sigma_0}\right)}{\sigma_0} + (1 - p) \cdot \frac{\phi\left(\dfrac{Y_t - \mu_1}{\sigma_1}\right)}{\sigma_1} \qquad \text{(EQ 17)}$$

where ϕ is the standard normal p.d.f. The key is that the probability of observing Y_t depends only on the parameters $(\mu_0, \mu_1, \sigma_0, \sigma_1, p)$ and not on other observations of Y_t. In the Markov switching case, however, we are unable to duplicate the step from (EQ 15) to (EQ 16). This is because the probability of being in a given state now depends on the state the period before, which is not observed. Since the p.d.f. of Y_t is necessary to calculate the model's likelihood function, this greatly complicates maximum-likelihood estimation.

Hamilton gets around this problem by replacing $Pr(S_t = 0)$ in (EQ 15) with[8]

$$Pr(S_{t-1} = 0|\{Y_{t-1}, ..., Y_1\}) \cdot p + Pr(S_{t-1} = 1|\{Y_{t-1}, ..., Y_1\}) \cdot (1 - q) \qquad \text{(EQ 18)}$$

and $Pr(S_t = 1)$ with

$$Pr(S_{t-1} = 0|\{Y_{t-1}, ..., Y_1\}) \cdot (1 - p) + Pr(S_{t-1} = 1|\{Y_{t-1}, ..., Y_1\}) \cdot q \qquad \text{(EQ 19)}$$

He then shows that $Pr(S_{t-1} = 0|\{Y_{t-1}, ..., Y_1\})$ can be calculated from $Pr(S_{t-2} = 0|\{Y_{t-2}, ..., Y_1\})$ using a simple updating formula that depends only on Y_{t-1} and $(\mu_0, \mu_1, \sigma_0, \sigma_1, p, q)$. In effect, this means we can recursively calculate the likelihood function of Y_t given that of Y_{t-1}. However, to start the process off, we must

---

8. The observant reader will notice that this means we are now dealing with the *conditional* likelihood function instead of the usual *unconditional* function. As Hamilton (1992) notes, the difference is that the former treats the initial probability $Pr(S_0 = 0)$ as given, a constant independent of the other parameters of the model. If we endogenize it by imposing (EQ 20), however, then we have the usual *unconditional* likelihood function.

supply another parameter $\rho \equiv Pr(S_0 = 0)$. In the absence of strong priors about which state we start in, this can be set equal to the unconditional probability of being in state 0, which is

$$\frac{1-q}{2-p-q} = \rho \qquad \text{(EQ 20)}$$

Alternatively, we could set ρ arbitrarily to zero or one, implying that we know with certainty the regime at time zero, or we could have ρ be a parameter of the model to be estimated.

3.1.2 The Smoothed Probabilities

Hamilton (1989) also mentions a related but more complex function he calls a "full-sample smoother." This gives the probability that $S_t = 0$ conditioning on all the available data and not just that available up to time t. In other words, the smoother gives

$$Pr(S_t = 0 | \{Y_T, ..., Y_1\}) \qquad \text{(EQ 21)}$$

instead of

$$Pr(S_t = 0 | \{Y_t, ..., Y_1\}) \qquad \text{(EQ 22)}$$

Previously, calculating the smoothed probability at any time t requires that we work out the conditional p.d.f. ($f(Y_\tau | \{Y_{\tau-1}, ..., Y_1\})$) for each observation τ from t to T and then go back and re-evaluate it for each τ under the assumption that $S_t = 0$. This means that if the number of calculations required to evaluate the likelihood function is a linear function of the number of observations, say $b \cdot T$, then the number of calculations required to calculate the full-sample smoother from 1 to T will be roughly $b \cdot \left(\frac{T^2}{2} + 1\right)$.

Kim (1994) provides a simpler method. Based upon the fact that for the last observation the smoothed and ex post probabilities are equal, he develops a recursive formula for the smoothed probabilities.

$$Pr(S_t = 0 | Y_T) = Pr(S_t = 0 | Y_t)\left(p_{00}\frac{Pr(S_{t+1} = 0 | Y_T)}{Pr(S_{t+1} = 0 | Y_{t+1})} + p_{10}\frac{Pr(S_{t+1} = 1 | Y_T)}{Pr(S_{t+1} = 1 | Y_{t+1})}\right)$$

This reduces the number of calculation from almost T^2 to about $2T$.

3.1.3 Differences Between Hamilton's and Our Code

There are some features in Hamilton's procedures that are not present in our code. As previously noted, Hamilton's allow for autocorrelation within states while the new procedures do not. His EM procedures also handle vector Markov switching systems; that is, where Y_t is a vector of variables governed by a single state variable S_t. Hamilton's programs also contain code for imposing a Bayesian prior on the likelihood function to help ensure convergence, as described in Engel and Hamilton (1990).

The new likelihood routines are much more efficiently coded than those available from Hamilton. Much of this stems from dropping the serial correlation option. However, the biggest difference comes from minimizing the amount of calculation "inside the loop." Since the likelihood function is evaluated with a non-linear recursive formula, any procedure returning the likelihood function must have a loop that calculates the likelihood function of observation t given that for t-1. However, like all interpreted computer languages, Gauss is extremely inefficient at processing instructions inside a loop. Therefore, moving as many instructions as possible outside the loop greatly speeds execution.

Table 1: **Speed Comparisons of Likelihood Functions and Maximum-Likelihood Estimation**

	Hamilton's	New
Evaluation of Likelihood Function:		
Elapsed Time in Seconds	10.38	1.15
% Time Savings	0.00	88.92
Speed-up Factor	1.00	9.03
Maximum-Likelihood Estimation:		
Elapsed Time in Seconds	1246.3	297.09
% Time Savings	0.0	76.16
Speed-up Factor	1.0	4.19

The above table shows the results of an experiment to compare the speeds of the new likelihood function procedure with that of Hamilton's with Gauss version 2.0.[9] It uses the GNP data from Hamilton (1989) and estimates a Markov mixture model with unequal variances in the two states and (EQ 20) imposed. The evaluation of the likelihood function with the new procedure is over 9 times faster than with Hamilton's procedures, and the time required to estimate this simple model is cut from over 20 to less than 5 minutes.

---

9. These results have not been updated to the current version. However, there is no reason that the newer version should produce very different results.

3.1.4 Analytic Gradients

The availability of the smoothed probabilities makes it feasible to calculate the likelihood function's analytic gradients. These gradients further increase the speed of estimation of the models.

To estimate by maximum-likelihood methods requires calculating the derivatives of the likelihood function with respect to the parameter vector.[10] (This vector is also referred to as the *score* or the *gradient* vector.) This is usually done using numerical techniques that approximate the derivative by the change in the likelihood function for small changes in the parameter vector. This is not especially efficient, however, as such techniques typically require *N+1* evaluations of the likelihood function to calculate the *N* elements of the score, and $N^2 + 1$ to calculate the Hessian (the matrix of second derivatives). The use of analytical gradients can greatly reduce the number of calculations required to evaluate either of these objects. This in turn considerably speeds up maximum-likelihood estimation of such models with no loss in accuracy.

These gradients are further detailed in Gable, van Norden and Vigfusson 1995. The gradients are derived and comparisons are made between the analytical and numerical methods.

In general, the analytic derivatives can improve calculation efficiency by roughly a factor of four.[11] The relative speed of the analytic derivatives increases with sample size, but the improvement seems to slow as samples become large.

An analysis of the code illustrates why the analytic derivatives are typically faster than the numerically calculated derivatives. The numerically calculated derivatives make P (the number of parameters) calls to the likelihood function. Most of the time the likelihood function is in turn spent in a loop over the sample size (N). One therefore could approximate the amount of calculations in the numeric derivatives as P*N.

The analytic derivatives make one call to the smoother, which in turn calculates the likelihood function and then performs an additional loop over the sample size. Hence, the smoother requires approximately 2*N calculations. Calculating the gradient then requires a loop over the number of states (S) for a set number of calculations (L). In total, therefore, this requires approximately 2*N+S*L calculations.

---

10. Alternatives to gradient-based maximum-likelihood estimation include the EM algorithm and simulated annealing. However, simulated annealing is also inefficient, and the EM algorithm has a lower rate of convergence near the optimum than do most popular gradient-based maximization methods (such as the Newton algorithm.) The score is also useful for the calculation of standard errors, and for diagnostic tests (see White 1994).

11. Our likelihood function was used for both methods of calculating the gradients. So using our likelihood function with analytic gradients should result in its being about 15-20 times faster than Hamilton's original code.

With the above analysis, we can predict that the analytic gradient procedure will be faster as either the sample size or the number of parameters increases. In the unlikely case that the number of states increases with the number of parameters held constant, there may be a relative improvement of the numeric technique. More likely, an increase in the number of states will cause a substantial increase in the number of parameters.

3.2 A Sample Program for Markov Switching Models

There is now a library file for the procedures, switch.lcg. Hence it is no longer necessary to have a number of #include statements. Rather one merely needs to have a line such as

library maxlik, switch.lcg;

The next required step is to clear the globals that will be used in the estimation. These include *nstat*, *vei*, *ix*, *iy, nrows, sameness,* and *terms.* The variable *nstat* is the number of states. At present *nstat* must be equal to two. The variable *vei* is an $nstat^2$ by 1 vector. The first *nstat* entries specify the number of variables in the corresponding level equation. The next *nstat (nstat-1)* specify the number of variables in the corresponding transition equation. The variable *ix* controls whether variances differ across states or are equal. The variable *iy* sets the starting condition for the probability of being in the different states at time 0. TABLE 2. gives the possible settings for *ix, sameness* and *iy.*

TABLE 2. Settings for *ix* and *iy*

Value	*ix*	*sameness*	*iy*
0	Variances Equal Across States	Different explanatory variables in all equations.	rho=$\frac{1-q}{2-p-q}$
1	Variances Differ Across States	Same explanatory variables in all level equations. Different in transition equations.	rho is a parameter
2		Different explanatory variables in all level equations. Same explanatory variables in transition equations.	rho =0
3		Same variables for level equations and same variables for transition equations.	rho=1
4		Same variables for level equations and transition equations.	

Each row of *terms,* an $nstat^2$ by two matrix, specifies what data are used by a particular equation. The first *nstat* rows are for the level equations. The rest are for the transition equations. The first column of *terms* specifies the first column of data used in an equation. The second specifies the last. The dependent variable is assumed to occupy the first column of the data matrix. The elements in *terms* may be specified such that the same variables are used in more than one equation.

To use analytic derivatives in the maximum-likelihood estimation, one must add the line

```
_mlgdprc= &ANGRAD;
```

Next make a call to the procedure mkvstart, which returns the value of *terms*, the vector that specifies which columns of the data matrix to use for each equation and *nrows*, the number of observations. It is assumed that the data are formatted such that the dependent variable is in the first column, then the explanatory variables for the level equations, and then the explanatory variables for the transition equations.

```
{terms}=mkvstart(DATA,vei,nstat);
```

After loading in the data and setting all the necessary globals for maximum-likelihood estimation, including starting values, make a call to the maximum-likelihood procedure. See Aptech's Maximum Likelihood manual for more information.

```
{a,b,c,d,e}=maxlik(DATA,0,&swmkv,StartValues);
```

3.3 Speeding up Estimation with C procedures

Besides the procedures already mentioned, code exists to make estimating Markov switching models faster. We speed up the calculations of the likelihood function and the smoothed probabilities by replacing the main loop with a C language procedure.

The C code actually only replaces a small portion of the code. Because C handles do-loops much more efficiently than Gauss, the C-augmented estimation is several times faster than the same procedure that uses the Gauss do-loops.

To use the C procedures, ones needs a C compiler and a version of Gauss no older than version 3.1.14. Before the C procedures can be used, their source code has to be compiled.[12] Using these procedures on a computer running Unix[13] requires the Gauss dynamic libraries feature that was introduced in version 3.1.14. The dynamic libraries commands allow Gauss to use functions written in other computer languages.

After the C procedures are compiled, using them is fairly simple. At the start of the program include a line with the Dlibrary[14] function referencing the compiled library. For example, with our compiled library libmkz.so, the line would be:

```
DLIBRARY /home/int/vigf/regswt/ccode/libmkz.so;
```

---

12. The compiler to be used will depend on the computer where the programs will be run. We used the GNU C compiler GCC.

13. The dynamic library interface is currently (as of 5/25/95) supported on the SunOS 4.1.3, Solaris 2.x, AIX, HPUX, IRIX, and OSF1 platforms (Aptech README file). In DOS, instead of the dynamic libraries, one uses the foreign language interface. See the Gauss user manual for details.

14. The Gauss on-line help system gives further details on using the DLIBRARY command.

To access the C procedures, set the global variable *Newalg* in your main program equal to one. No other changes are necessary.

3.4 EM algorithm

We have written the Gauss code for the EM algorithm for the case of constant transition probabilities (Hamilton 1994). The algorithm is useful for getting close to a maximum in spite of potentially bad starting values. The algorithm does have difficulties in final convergence to the maximum. Therefore, we recommend using the EM algorithm first to come close to the maximum and then passing this coefficient vector to the standard maxlik approach for final convergence. If *Newalg* equals one, then the EM algorithm will use a C procedure augmented version of Kim's smoother.

Something, however, is not right with our programmed version of the EM algorithm. A slight discrepancy exists between the EM algorithm's results and the maxlik package's results. In these cases, if the maxlik estimate is used as a starting value for the EM algorithm, then the EM algorithm will move away from this estimate to its own, even if the maxlik estimate has a higher likelihood. The difference between these two estimates is slight, so the EM algorithm continues to be useful for coming close to a maximum. We are looking for a reason for the difference and would appreciate feedback on possible solutions.

3.5 Future Developments

The analytic gradients and likelihood function currently can be used only for a two-state system. The code has been written such that the change to an n-state system, while non-trivial, is certainly feasible. The code for the derivatives of (EQ 12) is already capable of handling n-states. The global variables *vei, terms*, and *nstat* should facilitate the transition to the n-state system.

The EM algorithm currently estimates only models with constant transition probabilities. Diebold, Lee and Weinbach (1994), however, have an EM algorithm for the non-constant transition probabilities model that we hope to include in a latter release of this software.

4.0 Specification Tests for Regime-Switching and Other Models

White (1987, 1994) and Hamilton (1990, 1996) suggested a set of specification tests based on the serial correlation properties of the gradient vectors at a given set of parameter estimates. These tests can be adapted to the case of simple-switching regressions. They allow one to construct simple tests for serial correlation, ARCH and Markov switching.

Section 4.1 reviews the econometric basis of these tests. Section 4.1.1 lists a Gauss procedure that can be used to calculate such tests for virtually any model whose gradient can be calculated. Section 4.2 discusses the application and interpretation of the misspecification

tests in the context of switching regressions, while Section 4.2.1 lists a Gauss procedure that applies a series of such tests to a standard switching regression.

4.1 The Theory of Score-Based Specification Tests

Let $L(y_t|x_t, \theta)$ denote the likelihood function of a given observation y_t conditional on some information set x_t and a choice of parameters θ. The score $h_t(\theta)$ is simply the gradient of $L(y_t|x_t, \theta)$ with respect to θ. Subject to a few regularity conditions (such as no corner solutions and ergodicity), it can be shown that at the true parameter estimates θ_0, $h_t(\theta_0)$ should be unforecastable on the basis of any information available at $t-1$, which includes $h_{t-1}(\theta_0)$. Since $h_t(\theta)$ is a vector with the same dimensions as θ (say, mx1), this implies that the absence of any first-order serial correlation in $h_t(\theta_0)$ alone gives us mxm testable restrictions. While in theory, we could test all of these conditions, or test even higher-order restrictions on its serial correlation, in practice, most of these restrictions are very difficult to interpret. In what follows, attention will limited to those restrictions that are most easily understood.

White (1987) constructs the general test by listing those l elements of the mxm matrix $h_t(\theta) \cdot h_t(\theta)'$ that we wish to test in the lx1 vector $c_t(\theta)$. He then lets $\hat{\theta}$ denote our maximum-likelihood estimate of θ, and lets $\hat{A}$ be the (2,2) subblock of the *inverse* of the partitioned matrix

$$T^{-1} \cdot \begin{bmatrix} \displaystyle\sum_{t=1}^{T} h_t(\hat{\theta}) \cdot h_t(\hat{\theta})' & \displaystyle\sum_{t=1}^{T} h_t(\hat{\theta}) \cdot c_t(\hat{\theta})' \\ \displaystyle\sum_{t=1}^{T} c_t(\hat{\theta}) \cdot h_t(\hat{\theta})' & \displaystyle\sum_{t=1}^{T} c_t(\hat{\theta}) \cdot c_t(\hat{\theta})' \end{bmatrix} \tag{14}$$

where T is the sample size. In this case, he shows that if the model is correctly specified, the matrix product

$$T^{-1} \cdot \left[\sum_{t=1}^{T} c_t(\hat{\theta}) \right]' \cdot \hat{A} \cdot \left[\sum_{t=1}^{T} c_t(\hat{\theta}) \right] \tag{15}$$

will have a $\chi^2(l)$ asymptotic distribution.

4.1.1 General Procedure for Score-Based Tests: wmisspec()

The Gauss procedure, wmisspec(), may be used to calculate the test statistics defined by (15). Since the theory is completely general, this can be applied to almost any econometric model for which we can calculate the gradient vector.[15] For any of these models, we can select any parameter or combination of parameters to use in our test for misspecification.

Accordingly, this procedure may be useful in a wide variety of situations. The procedure does not require that the model be estimated by maximum likelihood. All that is required is the actual gradient matrix. For some models, such as the switching-regression case considered below, this can simply be done using a procedure that calculates gradients analytically. Otherwise, it will be necessary to supply a procedure that calculates the likelihood function, and use the gradp() function in Gauss to provide numerical gradient estimates.

As written, the procedure requires as arguments only the *T*x*m* matrix of gradients evaluated at the parameter estimates, and a column vector that lists the numbers of the columns in the gradient vector that we wish to include in the test.[16] It then returns the test statistic and its χ^2 marginal significance level.

4.2 Misspecification Tests in Switching Regressions

The specification tests are similar whether one is testing simple switching or Markov switching. If we assume that the level equations each contain a constant, then testing whether each of these gradients are autocorrelated amounts to testing whether there is omitted serial correlation in ε_{t1} or ε_{t2}. The intuition here is that serial correlation in these gradients implies that we tend to find "runs" where the constant should be higher or lower, which in this context implies persistence in the residuals, or serial correlation. For the simple-switching case, testing the gradient of the probability of being in a regime gives a test for neglected Markov-like switching effects, since serial correlation here implies that regimes seem to be more persistent than indicated. For the Markov switching case, testing the gradient of the constant term in the transition equation gives a test for neglected higher-order Markov switching effects. Finally, tests of the gradients of σ_S and σ_C amount to tests for first-order regime-specific ARCH effects, since persistence here implies that the volatility in each regime seems to vary over time in a way captured by a first-order autoregression.

4.2.1 Testing Simple-Switching Models: swmisspc()

The Gauss program swmisspc() calculates and prints the test statistics discussed in the previous section for the simple-switching model. It is designed to be used immediately after the estimation of a switching regression by maxlik(). It accepts as arguments the final parameter estimates and the data set used by swli() and swgrad(), and prints the test statistics and their marginal significance levels.

The procedure assumes that swgrad() has been defined, as have the global variables *_n1, _n2,* and *_n3*. Finally, the procedure also assumes that the first regressor in each equation is a constant.

---

15. More precisely, the model must satisfy certain regularity conditions, such as ergodicity. See White (1987).

16. In the notation of the previous section, this means that $c_t(\theta)$ can include only diagonal elements of $h_t(\theta) \cdot h_t(\theta)'$. However, these are the ones that are most likely to be of interest.

4.2.2 Testing Markov Switching Models: mkvspec()

The Gauss program mkvspec() calculates and prints the test statistics or the Markov switching model. It is designed to be used immediately after estimation using the same data set as and final estimates from maxlik. The procedure assumes that the global variables discussed in subsection 3.2 have been defined and that the first regressor in each equations is a constant.

5.0 Bibliography

Diebold, Francis X. Joon Haeng Lee and Gretchen C. Weinbach. 1994. "Regime Switching with Time Varying Transition Probabilities." in C. Hargreaves *Nonstationary Time Series Analysis and Cointegration* Oxford, Oxford University Press.

Durland, J. Michael and Thomas H. McCurdy. 1994. "Duration Dependent Transition in a Markov Model of U.S. GNP Growth." *Journal of Business & Economic Statistics* July 12:3:279-288.

Elliot, Robert. J. Lakhdar Aggoun. and John B. Moore. 1995. *Hidden Markov Models Estimation and Control.* Springer Verlag. New York.

Engel, Charles 1994. "Can the Markov Switching Model forecast exchange rates?" *Journal of International Economics* 36 pp 151-165.

Engel, Charles and James D. Hamilton 1990. "Long Swings in the Dollar: Are They in the Data and Do Markets Know It?" *American Economic Review* Sept. 80:4:689-713.

Gable, Jeff. Simon van Norden and Robert Vigfusson 1995 "Analytical Derivatives for Markov-Switching Models." Bank of Canada Working Paper 95-7.

Goldfeld, S.M. and R.E. Quandt. 1973. "A Markov model for switching regressions." *Journal of Econometrics.* 1:3-16

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57:357-84

—————. 1992. "Estimation, Inference, and Forecasting of Time Series Subject to Changes in Regime." in C.R. Rao and G.S. Maddala *Handbook of Statistics.* Volume 10

—————. 1994. *Time Series Analysis*. Princeton: Princeton University Press.

—————. 1996. "Specification testing in Markov-switching time-series models." *Journal of Econometrics* January. 70:1:127-157.

Hansen, B. E. 1992. "The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model Of GNP." *Journal of Applied Econometrics.* Vol. 7 Supplement.

Hartley, Michael J. 1978. "Comment on Quandt and Ramsey", *Journal of the American Statistical Association*, December

Kim, Chaing-Jin. 1994. "Dynamic Linear Models with Markov-Switching." *Journal of Econometric.* 60:1-22

Lam, Pok-sang. 1990. "The Hamilton Model with a General Autoregressive Component Estimation and Comparison with Other Models of Economic Time Series." *Journal of Monetary Economics.* 26:409-32

van Norden, Simon. 1996. "Regime Switching as a Test for Exchange Rate Bubbles" *Journal of Applied Econometrics* forthcoming

van Norden, Simon and H. Schaller. 1993. "Speculative Behaviour, Regime Switching and Stock Market Fundamentals" Bank of Canada Working Paper 93-2

Vigfusson, Robert. 1996. "Switching Between Chartists and Fundamentalists." Bank of Canada Working Paper 96-1

White, Halbert. 1987. "Specification Testing in Dynamic Models." in *Advances in Econometrics, Fifth World Congress, Volume II*, Truman F. Bewley ed., Cambridge, Cambridge Press.

—————. 1994. *Estimation, Inference, and Specification Analysis* Econometric Society Monographs. Cambridge University Press.

Bank of Canada Working Papers

1996

96-1	Switching Between Chartists and Fundamentalists: A Markov Regime-Switching Approach	R. Vigfusson
96-2	Decomposing U.S. Nominal Interest Rates into Expected Inflation and Ex-Ante Real Interest Rates Using Structural VAR Methodology	P. St-Amant
96-3	Regime-Switching Models: A Guide to the Bank of Canada Gauss Procedures	S. van Norden and R. Vigfusson

1995

95-1	Deriving Agents' Inflation Forecasts from the Term Structure of Interest Rates	C. Ragan
95-2	Estimating and Projecting Potential Output Using Structural VAR Methodology: The Case of the Mexican Economy	A. DeSerres, A. Guay and P. St-Amant
95-3	Empirical Evidence on the Cost of Adjustment and Dynamic Labour Demand	R. A. Amano
95-4	Government Debt and Deficits in Canada: A Macro Simulation Analysis	T. Macklem, D. Rose and R. Tetlow
95-5	Changes in the Inflation Process in Canada: Evidence and Implications	D. Hostland
95-6	Inflation, Learning and Monetary Policy Regimes in the G-7 Economies	N. Ricketts and D. Rose
95-7	Analytical Derivatives for Markov Switching Models	J. Gable, S. van Norden and R. Vigfusson
95-8	Exchange Rates and Oil Prices	R. A. Amano and S. van Norden
95-9	Selection of the Truncation Lag in Structural VARs (or VECMS) with Long-Run Restrictions	A. DeSerres and A. Guay
95-10	The Canadian Experience with Weighted Monetary Aggregates	D. Longworth and J. Atta-Mensah
95-11	Long-Run Demand for M1	S. Hendry
95-12	The Empirical Performance of Alternative Monetary and Liquidity Aggregates	J. Atta-Mensah