Three Essays on Econometrics

by

Jiahui Wang

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

University of Washington

1997

Approved by__________________________________

(Chairperson of Supervisory Committee)

____

____

Program Authorized
to Offer Degree Economics

Date 6-12-97

UMI Number: 9736399

University of Washington

Abstract

Three Essays on Econometrics

by Jiahui Wang

Chairperson of Supervisory Committee: *Assistant Professor Eric Zivot*
Department of Economics

This dissertation is composed of three chapters on modern econometric topics. Chapter 1 studies the problem of making inference on structural parameters in instrumental variables regression with weak instruments. It has become well known in the literature that the instrumental variables estimates are strongly biased and the Wald type test statistics are nonpivotal and thus could have zero coverage probability when the instruments selected are very weak. Using a local-to-zero assumption for the coefficients of instruments, Chapter 1 shows that likelihood type statistics for testing hypothesis on the structural parameters are asymptotically (boundedly) pivotal, thus valid inference can be made using the asymptotic (bounding) distribution of the likelihood type statistics. Characteristics of the confidence sets for the structural parameters are also analyzed.

Chapter 2 develops a Bayesian time series model of multiple structural changes, which allows changes in both the time trend and the variance term. In spite of the complex structure the model allows, it is shown in Chapter 2 that the Bayesian Gibbs sampling technique can be employed to estimate the model in a straightforward way. The model selection problem regarding the number of structural changes is also considered using the posterior odds ratio.

Chapter 3 presents a high order Markov switching model, where the underlying high order Markov chain is modeled after Raftery's parsimonious formulation. The model is used to investigate the statistical distributions of daily foreign exchange rate data. For most exchange rates high order Markov switching appears to exist in the data. However, the forecasts from the high order Markov switching models do not have obvious advantage over alternative models such as the first order Markov switching model and GARCH model with conditionally t-distributed errors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Chapter 1 of this dissertation was written when the author was a research assistant for Professors Charles R. Nelson, Richard Startz and Eric Zivot, and evolved from the work done by Charles R. Nelson, Richard Startz and Eric Zivot. Without Eric Zivot's excellent guidance and cooperation, this chapter would not be possible. The author wants to thank Change-Jin Kim of Korea University for inspiring conversations which led to Chapter 2. Charles Engel's suggestion has helped improve the presentation of Chapter 3. Finally, the author wants to express his gratitude to his parents, Chuanfu Wang and Xiuzhen Gu, who went through all the pain and hardship bringing him up.

Chapter 1

INFERENCE ON STRUCTURAL PARAMETERS IN INSTRUMENTAL VARIABLES REGRESSION WITH WEAK INSTRUMENTS

1.1 Introduction

Instrumental variables (IV) estimation with weak instruments has recently captured the attention of applied and theoretical econometricians. Research has uncovered two problems associated with inference on structural parameters when IV methods are used and there are weak instruments. First, Nelson and Startz[39, 40], Maddala and Jeong[36], Bound, Jaeger and Baker[11] have shown that in the presence of weak instruments IV estimates can be strongly biased in the same direction as OLS estimates and IV-based inference can be highly misleading. Second, Dufour[17] showed that in a limited information simultaneous equations model (LISEM) with very weak instruments Wald statistics for testing hypothesis about structural parameters are not asymptotically pivotal and thus the standard asymptotically justified "estimate $\pm 2\cdot$ asymptotic standard error" type confidence interval for a structural parameter can have zero coverage probability.

To understand these problems, Staiger and Stock[56], hereinafter SS, derived alternative asymptotic distributions for TSLS (two-stage-least-squares) and LIML (limited-information-maximum-likelihood) estimators of structural parameters in a LISEM based on a local-to-zero assumption for the coefficients on the instruments in the reduced form equation. Under fairly weak conditions, SS showed that the TSLS and LIML estimators of the structural parameters and Wald statistics do not have standard asymptotic distributions but rather have distributions that depend on nuisance parameters in a complicated way. Inference on structural parameters based on the SS results is complicated by the fact that some of the nuisance parameters appearing in the asymptotic distributions are not consistently estimable. One solution to this problem is to follow Dufour[17] and base inference of structural parameters using test statistics that have asymptotically pivotal or boundedly pivotal distributions.

Since SS and Dufour[17]'s results show that Wald statistics based on TSLS and LIML are not asymptotically pivotal or boundedly pivotal, in this chapter we will consider the construction of asymptotically valid confidence sets for structural parameters based on inverting Lagrange multiplier (LM) and likelihood ratio (LR) type statistics. The approach is motivated by Dufour[17]'s observation that some LR statistics for hypotheses on nearly nonidentified parameters are boundedly pivotal under certain conditions. We can show, using SS's local-to-zero asymptotics, that the asymptotic distributions of the TSLS LM statistic and the LIML LR statistic are asymptotically boundedly pivotal. Hence, the analysis in this chapter shows that certain LM statistics as well as LR statistics provide asymptotically valid inference on structural parameters in a LISEM framework.

Based on the asymptotic (bounding) distributions of the likelihood type test statistics, we follow Dufour[17] and consider the construction of valid $(1-\alpha)\cdot 100\%$ confidence sets for structural parameters. Dufour[17] proved that under general regularity conditions, any valid confidence set with coverage probability $(1-\alpha)$ for a locally almost unidentified (LAU) parameter should be unbounded with probability close to $(1-\alpha)$. For the single right hand side (RHS) endogenous variable case, this result is verified using the goodness-of-fit statistics from the first-stage regression. Generally we note that valid confidence sets for the structural parameters can be obtained by inverting test statistics based on the asymptotic (bounding) distribution which does not depend on nuisance parameters.

The rest of this chapter is organized as follows. Section 1.2 lays down the Limited Information Simultaneous Equations Model, and introduces the local-to-zero assumptions made by SS. In Section 1.3 we derive the asymptotic distributions of LM and LR type statistics for testing hypothesis on structural parameters. In Section 1.4 we first consider the construction of valid confidence intervals for the structural parameter in the single RHS endogenous variable case and verify Dufour[17]'s result that any valid $(1-\alpha)\cdot 100\%$ confidence set for a LAU structural parameter is unbounded with probability $(1-\alpha)$. Then a characterization of valid joint confidence sets is given in the multiple RHS endogenous variables case. In Section 1.5 the finite sample performance of the various test statistics is investigated by Monte Carlo experiments. Section 1.6 concludes the chapter. The proof of the results in this chapter is relegated to the Appendix.

1.2 *The Local-to-Zero LISEM Framework*

The LISEM model we consider is:

$$\mathbf{y} = \mathbf{Y}\boldsymbol{\beta} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{u} \tag{1.1}$$

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{X}\boldsymbol{\Phi} + \mathbf{V} \tag{1.2}$$

where $\mathbf{y}$ and $\mathbf{Y}$ are a $T\times 1$ and a $T\times n$ matrix of endogenous variables, respectively. $\mathbf{X}$ is a $T\times k_1$ matrix of included exogenous variables, $\mathbf{Z}$ is a $T\times k_2$ matrix of excluded exogenous variables (or instruments), $\mathbf{u}$ is a $T\times 1$ vector of structural errors and $\mathbf{V}$ is a $T\times n$ matrix of reduced form errors. Let $k = k_1 + k_2$ and let $\bar{\mathbf{Z}} = [\mathbf{X}\ \mathbf{Z}]$, which is assumed to be of full column rank and uncorrelated with $\mathbf{u}$ and $\mathbf{V}$ with

$$\mathrm{E}(\bar{\mathbf{Z}}_t\bar{\mathbf{Z}}_t{}') = \begin{bmatrix} \mathrm{E}(\mathbf{X}_t\mathbf{X}_t') & \mathrm{E}(\mathbf{X}_t\mathbf{Z}_t') \\ \mathrm{E}(\mathbf{Z}_t\mathbf{X}_t') & \mathrm{E}(\mathbf{Z}_t\mathbf{Z}_t') \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{XX} & \mathbf{Q}_{XZ} \\ \mathbf{Q}_{ZX} & \mathbf{Q}_{ZZ} \end{bmatrix} = \mathbf{Q}.$$

The error terms u_t and $\mathbf{V}_t$ are assumed to have zero mean, and to be serially uncorrelated and homoskedastic with

$$\text{var}\begin{bmatrix} u_t \\ \mathbf{V}_t \end{bmatrix} = \begin{bmatrix} \sigma_{uu} & \Sigma_{uV} \\ \Sigma_{Vu} & \Sigma_{VV} \end{bmatrix} = \mathbf{\Sigma}.$$

We call the model just-identified when $k_2 = n$ and the model over-identified when $k_2 > n$.

Since we are only interested in making inference on $\boldsymbol{\beta}$, it is convenient to transform (1.1)-(1.2) using the Frisch-Waugh-Lovell theorem[1] as follows:

$$\mathbf{y}^{\perp} = \mathbf{Y}^{\perp}\boldsymbol{\beta} + \mathbf{u}^{\perp}, \tag{1.3}$$

$$\mathbf{Y}^{\perp} = \mathbf{Z}^{\perp}\mathbf{\Pi} + \mathbf{V}^{\perp}, \tag{1.4}$$

where $\mathbf{A}^{\perp} = \mathbf{M}_X\mathbf{A} = (\mathbf{I}_T - \mathbf{P}_X)\mathbf{A}$, for any conformable matrix $\mathbf{A}$, and $\mathbf{P}_B = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$ denotes the matrix that projects onto the space spanned by any full rank matrix $\mathbf{B}$.

Let "$\xrightarrow{p}$" denote convergence in probability and "$\xrightarrow{d}$" denote convergence in distribution. As in SS, we use the following two assumptions:

Assumption 1 $\mathbf{\Pi} = \mathbf{\Pi}_T = \mathbf{C}/\sqrt{T}$, *where* $\mathbf{C}$ *is a fixed* $k_2 \times n$ *matrix.*

Assumption 2 *The following limits hold jointly:*

(1) $(\mathbf{u}'\mathbf{u}/T, \mathbf{V}'\mathbf{u}/T, \mathbf{V}'\mathbf{V}/T) \xrightarrow{p} (\sigma_{uu}, \Sigma_{Vu}, \Sigma_{VV})$*;*

(2) $\bar{\mathbf{Z}}'\bar{\mathbf{Z}}/T \xrightarrow{p} \mathbf{Q}$*;*

(3) $(\mathbf{X}'\mathbf{u}/\sqrt{T}, \mathbf{Z}'\mathbf{u}/\sqrt{T}, \mathbf{X}'\mathbf{V}/\sqrt{T}, \mathbf{Z}'\mathbf{V}/\sqrt{T}) \xrightarrow{d} (\mathbf{\Psi}_{Xu}, \mathbf{\Psi}_{Zu}, \mathbf{\Psi}_{XV}, \mathbf{\Psi}_{ZV})$, *where* $\mathbf{\Psi} \equiv (\mathbf{\Psi}'_{Xu}, \mathbf{\Psi}'_{Zu}, \text{vec}(\mathbf{\Psi}_{XV})', \text{vec}(\mathbf{\Psi}_{ZV})')'$ *is distributed as* $N(0, \mathbf{\Sigma} \otimes \mathbf{Q})$.

Assumption 1 is SS's local-to-zero assumption for the reduced form coefficients $\mathbf{\Pi}$. It is a device that allows the instruments $\mathbf{Z}$ to remain weakly correlated with the

[1] For example, see Davidson and MacKinnon[16].

included endogenous variables $\mathbf{Y}$ as the sample size grows large. As a result. the Wald statistic for testing $\mathbf{\Pi} = 0$ in (2) is $O_p(1)$, whereas in the standard fixed-$\mathbf{\Pi}$ asymptotics the Wald statistic tends to infinity. The local-to-zero assumption for $\mathbf{\Pi}$ also implies that the structural parameters $\boldsymbol{\beta}$ are asymptotically nearly nonidentified. Phillips[50] showed that there is a discontinuity in the asymptotic distribution of the IV estimate of $\boldsymbol{\beta}$ when $\boldsymbol{\beta}$ is nonidentified ($\mathbf{\Pi} = 0$). The local-to-zero approach of SS allows the study of the asymptotic distribution in the neighborhood of the parameter configurations where $\boldsymbol{\beta}$ ceases to be identified. The approach is useful in that it illustrates the dependence of the distribution on nuisance parameters in the neighborhood of points where the identification conditions fail. Indeed. SS were able to characterize situations in which standard asymptotic theory is applicable and conditions in which it is not. Assumption 2 imposes weak moment conditions on the exogenous variables and error terms.

Under Assumptions 1 and 2, SS showed that the convergence results of sample moments based on (1.3)-(1.4) are different from the fixed-$\mathbf{\Pi}$ case. The following lemma gives the main results we will need for the analysis in this chapter.

Lemma 1 *Suppose that Assumptions 1 and 2 hold for the model (1.1)-(1.2). Then the following convergence results hold jointly as* $T \to \infty$:[2]

(a) $\mathbf{u}^{\perp\prime}\mathbf{u}^{\perp}/T \xrightarrow{p} \sigma_{uu}$,

(b) $\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{u}^{\perp} \xrightarrow{d} \Sigma_{VV}^{\frac{1}{2}\prime}(\boldsymbol{\lambda} + \mathbf{Z}_V)'\mathbf{Z}_u\sigma_{uu}^{\frac{1}{2}} = \Sigma_{VV}^{\frac{1}{2}\prime}\nu_2\sigma_{uu}^{\frac{1}{2}}$,

(c) $\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{Y}^{\perp} \xrightarrow{d} \Sigma_{VV}^{\frac{1}{2}\prime}(\boldsymbol{\lambda} + \mathbf{Z}_V)'(\boldsymbol{\lambda} + \mathbf{Z}_V)\Sigma_{VV}^{\frac{1}{2}} = \Sigma_{VV}^{\frac{1}{2}\prime}\nu_1\Sigma_{VV}^{\frac{1}{2}}$.

(d) $T\mathbf{u}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{u}^{\perp}/\mathbf{u}^{\perp\prime}\mathbf{u}^{\perp} \xrightarrow{d} \mathbf{Z}_u'\mathbf{Z}_u$,

where $\boldsymbol{\lambda} = \mathbf{\Omega}^{\frac{1}{2}}\mathbf{C}\Sigma_{VV}^{-\frac{1}{2}}$, $\mathbf{Z}_u = \mathbf{\Omega}^{-\frac{1}{2}\prime}(\mathbf{\Psi}_{Zu} - \mathbf{Q}_{ZX}\mathbf{Q}_{XX}^{-1}\mathbf{\Psi}_{Xu})\sigma_{uu}^{-\frac{1}{2}}$. $\mathbf{Z}_V = \mathbf{\Omega}^{-\frac{1}{2}\prime}(\mathbf{\Psi}_{ZV} - \mathbf{Q}_{ZX}\mathbf{Q}_{XX}^{-1}\mathbf{\Psi}_{XV})\Sigma_{VV}^{-\frac{1}{2}}$, $\mathbf{\Omega} = \mathbf{Q}_{ZZ} - \mathbf{Q}_{ZX}\mathbf{Q}_{XX}^{-1}\mathbf{Q}_{XZ}$, $\nu_1 = (\boldsymbol{\lambda} + \mathbf{Z}_V)'(\boldsymbol{\lambda} + \mathbf{Z}_V)$ *and* $\nu_2 = (\boldsymbol{\lambda} + \mathbf{Z}_V)'\mathbf{Z}_u$. *The random vector* $(\mathbf{Z}_u, \mathrm{vec}(\mathbf{Z}_V)')' \sim N(0, \bar{\mathbf{\Sigma}} \otimes I_{k_2})$ *and* $\bar{\mathbf{\Sigma}}$ *is an* $(n+1) \times (n+1)$ *matrix with* $\bar{\mathbf{\Sigma}}_{11} = 1$, $\bar{\mathbf{\Sigma}}_{22} = I_n$, *and* $\bar{\mathbf{\Sigma}}_{21} = \bar{\mathbf{\Sigma}}_{12}' = \boldsymbol{\rho} = \Sigma_{VV}^{-\frac{1}{2}\prime}\Sigma_{Vu}\sigma_{uu}^{-\frac{1}{2}}$.

---

[2] See SS for details.

where $\bar{\Sigma}$ is partitioned conformably with Σ.

1.3 Test Statistics and Limiting Distributions

In this section let us consider testing hypotheses about $\boldsymbol{\beta}$ using LM and LR type test statistics based on GMM and LIML estimation techniques. Consider first GMM estimation. Given the model (1.1)-(1.2) is linear, GMM estimation is equivalent to TSLS estimation. Let $\mathbf{u}^{\perp}(\boldsymbol{\beta}) = \mathbf{y}^{\perp} - \mathbf{Y}^{\perp}\boldsymbol{\beta}$. Then the sample moment condition for GMM estimation is given by $m_T(\boldsymbol{\beta}) = \mathbf{Z}^{\perp\prime}\mathbf{u}^{\perp}(\boldsymbol{\beta}) = 0$. Under Assumption 2 $\sqrt{T}m(\boldsymbol{\beta}) \xrightarrow{d} N(0, \mathbf{W})$, where $\mathbf{W} = \sigma_{uu}\boldsymbol{\Omega}$. Let $\mathbf{W}_T$ be a consistent estimator of $\mathbf{W}$, then the efficient GMM estimator of $\boldsymbol{\beta}$ is found by minimizing the quadratic form $m_T(\boldsymbol{\beta})'\mathbf{W}_T^{-1}m_T(\boldsymbol{\beta}) = \mathbf{u}^{\perp}(\boldsymbol{\beta})'\mathbf{Z}^{\perp}\mathbf{W}_T^{-1}\mathbf{Z}^{\perp\prime}\mathbf{u}^{\perp}(\boldsymbol{\beta})$. Using $\mathbf{W}_T = \hat{\sigma}_{uu}\mathbf{Z}^{\perp\prime}\mathbf{Z}^{\perp}/T$ as the weighting matrix, where $\hat{\sigma}_{uu}$ is an estimator of σ_{uu}, gives $\hat{\boldsymbol{\beta}}_{GMM} = \hat{\boldsymbol{\beta}}_{TSLS} = (\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^{\perp}}\mathbf{Y}^{\perp})^{-1}\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^{\perp}}\mathbf{y}^{\perp}$.

Newey and West[42] showed how to construct the trio of statistics - Wald, LM, D (or LR) - in the context of GMM for testing general nonlinear hypotheses about $\boldsymbol{\beta}$. For the simple hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ vs. $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, the three test statistics take the form:

$$\begin{aligned}
D_{GMM} &= [\mathbf{u}^{\perp}(\boldsymbol{\beta}_0)'\mathbf{P}_{Z^{\perp}}\mathbf{u}^{\perp}(\boldsymbol{\beta}_0) - \mathbf{u}^{\perp}(\hat{\boldsymbol{\beta}}_{GMM})'\mathbf{P}_{Z^{\perp}}\mathbf{u}^{\perp}(\hat{\boldsymbol{\beta}}_{GMM})]/\hat{\sigma}_{uu}, \\
LM_{GMM} &= \mathbf{u}^{\perp}(\boldsymbol{\beta}_0)'\mathbf{P}_{Z^{\perp}}\mathbf{Y}^{\perp}(\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^{\perp}}\mathbf{Y}^{\perp})^{-1}\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^{\perp}}\mathbf{u}^{\perp}(\boldsymbol{\beta}_0)/\hat{\sigma}_{uu}, \\
Wald_{GMM} &= (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{GMM})'\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^{\perp}}\mathbf{Y}^{\perp}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{GMM})/\hat{\sigma}_{uu}.
\end{aligned}$$

Newey and West[42] showed that these three statistics are numerically identical as long as the same estimator of σ_{uu} is used. Hence we use $\mathcal{GMM}$ to denote the common value of the three statistics for testing the hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

In the literature σ_{uu} is typicaly estimated using the TSLS estimator given by $\hat{\sigma}_{uu} = \mathbf{u}^{\perp}(\hat{\boldsymbol{\beta}}_{GMM})'\mathbf{u}^{\perp}(\hat{\boldsymbol{\beta}}_{GMM})/T$. In this case, $\mathcal{GMM}$ is the TSLS Wald statistic considered by SS. SS showed that under Assumptions 1 and 2 the TSLS estimators of

β and σ_{uu} are not consistent. In fact, they converge to random variables due to the asymptotic near nonidentification of β. Moreover, the $\mathcal{GMM}$ ($Wald_{TSLS}$) statistic does not have a χ^2 limiting distribution but rather has a distribution that depends on nuisance parameters such as the multiple correlation coefficient between u_t and $\mathbf{V}_t$ and the noncentrality parameter of the asymptotic distribution of the Wald statistic for testing $\mathbf{\Pi} = 0$ in equation (1.2).[3] Part of the nonstandard limiting behavior of $\mathcal{GMM}$ is attributable to the nonstandard limiting behavior of $\hat{\sigma}_{uu}$. This effect, however, can be eliminated by using $\hat{\sigma}_{uu,0} = \mathbf{u}^{\perp}(\beta_0)'\mathbf{u}^{\perp}(\beta_0)/T$, which is a consistent estimator of σ_{uu} under the null hypothesis, instead of using the TSLS estimate $\hat{\sigma}_{uu}$. Let $\mathcal{GMM}_0$ denote the $\mathcal{GMM}$ statistic computed using $\hat{\sigma}_{uu,0}$. Note that $\mathcal{GMM}_0$ has the natural interpretation of an LM statistic since all of the components of the statistic are based on the value of β under the null hypothesis. In addition, $\mathcal{GMM}_0$ is equal to T times the uncentered R^2 from the regression of $\mathbf{u}^{\perp}(\beta_0)$ on $\mathbf{P}_{Z^{\perp}}\mathbf{Y}^{\perp}$. The $\mathcal{GMM}_0$ statistic was not considered by SS or Dufour[17] and is new.

Next consider LIML estimation of β. SS focused on the k-class formulation of the LIML estimator and considered Wald statistics based on this formulation. As with TSLS estimation, SS showed that under Assumptions 1 and 2 the LIML estimator is inconsistent, has a nonstandard asymptotic distribution and the asymptotic distribution of LIML-based Wald statistics are nonpivotal.

Dufour[17] showed that the LR statistic based on the likelihood function of the unrestricted reduced form of (1.1)-(1.2) is boundedly pivotal in finite samples assuming normal errors. The bounding distribution, however, is based on the distribution of the Wilks statistic and is not particularly tight in models with many instruments. Instead of constructing a bound for the distribution of the LR statistic from the unrestricted reduced form let us consider the possibility of constructing a bounding distribution for the LR statistic computed directly from the concentrated likelihood function for

---

[3] See SS for details.

$\boldsymbol{\beta}$.

The log likelihood function of (1.1)-(1.2) concentrated with respect to the parameters $\boldsymbol{\Sigma}$, $\boldsymbol{\gamma}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Pi}$[4] is given by

$$\mathcal{L}^c(\boldsymbol{\beta}) = -\frac{nT}{2}\ln(2\pi) - \frac{T}{2}\ln k(\boldsymbol{\beta}) - \frac{T}{2}\ln|\bar{\mathbf{Y}}^{\perp\prime}\mathbf{M}_{\bar{Z}^{\perp}}\bar{\mathbf{Y}}^{\perp}|, \tag{1.5}$$

where $\bar{\mathbf{Y}} = [\mathbf{y}\ \mathbf{Y}]$ and

$$k(\boldsymbol{\beta}) = \frac{(\mathbf{y}^{\perp} - \mathbf{Y}^{\perp}\boldsymbol{\beta})'(\mathbf{y}^{\perp} - \mathbf{Y}^{\perp}\boldsymbol{\beta})}{(\mathbf{y}^{\perp} - \mathbf{Y}^{\perp}\boldsymbol{\beta})'\mathbf{M}_{Z^{\perp}}(\mathbf{y}^{\perp} - \mathbf{Y}^{\perp}\boldsymbol{\beta})}. \tag{1.6}$$

The LIML estimator of $\boldsymbol{\beta}$ is found by maximizing (1.5) which is equivalent to minimizing $k(\boldsymbol{\beta})$. It can be shown that the minimized value, $k(\hat{\boldsymbol{\beta}}_{LIML}) \equiv \hat{k}_{LIML}$, is the smallest root of the determinantal equation $\left|\bar{\mathbf{Y}}'\mathbf{M}_X\bar{\mathbf{Y}} - k\bar{\mathbf{Y}}\mathbf{M}_{\bar{\mathbf{Z}}}\bar{\mathbf{Y}}\right| = 0$. For testing the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ *vs.* $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, the LR statistic based on (1.5) is

$$LR_{LIML} = T\ln k(\boldsymbol{\beta}_0) - T\ln \hat{k}_{LIML}. \tag{1.7}$$

The following theorem gives the local-to-zero asymptotic distributions of $\mathcal{GMM}_0$ and LR_{LIML} and shows that these distributions are boundedly pivotal.

Theorem 1 *Suppose that Assumptions 1 and 2 hold for the model (1.1)-(1.2). Under the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, as $T \to \infty$.*

(a) $\mathcal{GMM}_0 \xrightarrow{d} \nu_2'\nu_1^{-1}\nu_2$.

(b) $LR_{LIML} \xrightarrow{d} \mathbf{Z}_u'\mathbf{Z}_u - \xi^*$, *where ξ^* is the smallest root of the determinantal equation* $\left|\Xi_0^* - k\cdot\bar{\boldsymbol{\Sigma}}\right| = 0$ *and* $\Xi_0^* = [\mathbf{Z}_u\ (\boldsymbol{\lambda} + \mathbf{Z}_V)]'[\mathbf{Z}_u\ (\boldsymbol{\lambda} + \mathbf{Z}_V)]$.

(c) When $k_2 = n$, the limiting distributions in parts (a) and (b) reduce to a $\chi^2(k_2)$ distribution; when $k_2 > n$, the asymptotic distributions are bounded from above by a $\chi^2(k_2)$ distribution.

Theorem 1 shows that the asymptotic distributions of $\mathcal{GMM}_0$ and LR_{LIML} are bounded by a $\chi^2(k_2)$ distribution. Some intuition for this result can be given by

---

[4] See Davidson and MacKinnon[16] page 647.

linking these statistics to the Anderson-Rubin (AR) statistic, originally proposed by Anderson and Rubin[2], for testing the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \; vs. \; \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$[5]. The AR statistic is given by

$$AR(\boldsymbol{\beta}_0) = \frac{\mathbf{u}^{\perp}(\boldsymbol{\beta}_0)'\mathbf{P}_{Z^{\perp}}\mathbf{u}^{\perp}(\boldsymbol{\beta}_0)/k_2}{\mathbf{u}^{\perp}(\boldsymbol{\beta}_0)'\mathbf{M}_{Z^{\perp}}\mathbf{u}^{\perp}(\boldsymbol{\beta}_0)/(T-k)},$$

which has an exact F distribution in finite samples with normal errors and, under Assumptions 1 and 2, SS showed that $k_2 AR(\boldsymbol{\beta}_0) \xrightarrow{d} \chi^2(k_2)$. Let $c_T = Tk_2/(T-k)$ so that $k(\boldsymbol{\beta}) = 1 + T^{-1}c_T AR(\boldsymbol{\beta})$. After some algebra it can be shown that $\mathcal{GMM}_0 = c_T k(\boldsymbol{\beta}_0)^{-1} AR(\boldsymbol{\beta}_0) - \mathbf{u}^{\perp}(\hat{\boldsymbol{\beta}}_{GMM})'\mathbf{P}_{Z^{\perp}}\mathbf{u}^{\perp}(\hat{\boldsymbol{\beta}}_{GMM})/\hat{\sigma}_{uu.0}$ and $LR_{LIML} = c_T AR(\boldsymbol{\beta}_0) - c_T AR(\hat{\boldsymbol{\beta}}_{LIML}) + o_p(1)$. Now, under Assumptions 1 and 2 $c_T \to k_2$, $k(\boldsymbol{\beta}_0) \xrightarrow{p} 1$, $c_T AR(\boldsymbol{\beta}_0) \xrightarrow{d} \chi^2(k_2)$ and the second terms in the expressions for $\mathcal{GMM}_0$ and LR_{LIML} are both positive with probability one in the overidentified case and zero in the exactly identified case. Hence $\mathcal{GMM}_0$, LR_{LIML} and $AR(\boldsymbol{\beta}_0)$ are asymptotically equivalent in the exactly identified case and in the overidentified case $\mathcal{GMM}_0$ and LR_{LIML} are less than $AR(\boldsymbol{\beta}_0)$ asymptotically and are therefore bounded by a $\chi^2(k_2)$ distribution.

1.4 Construction of Valid Confidence Intervals

We have shown that $\mathcal{GMM}_0$ and LR_{LIML} for testing the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ are asymptotically bounded by a $\chi^2(k_2)$ distribution. Thus, valid inference can be made based on this asymptotic bounding distribution. In principle we can also invert these test statistics to obtain valid joint confidence sets for $\boldsymbol{\beta}$.

Simple manipulations of the test statistics yield the $(1-\alpha)\cdot 100\%$ joint confidence set for $\boldsymbol{\beta}$ as follows:

$$C(\boldsymbol{\beta}; 1-\alpha) = \{\boldsymbol{\beta} : f(\boldsymbol{\beta}) \leq 0\},$$

where

$$f(\boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{Y}^{\perp\prime}\mathbf{H}\mathbf{Y}^{\perp}\boldsymbol{\beta} - 2\mathbf{y}^{\perp\prime}\mathbf{H}\mathbf{Y}^{\perp}\boldsymbol{\beta} + \mathbf{y}^{\perp\prime}\mathbf{H}\mathbf{y}^{\perp}. \tag{1.8}$$

[5] A referee of *Econometrica pointed out this result.*

The matrix $\mathbf{H}$ depends on the particular test statistic used. For LR_{LIML}, $\mathbf{H} = I_T - \hat{k}e^{\chi^2_{1-\alpha}(k_2)/T}\mathbf{M}_{Z^\perp}$; for $\mathcal{GMM}_0$, $\mathbf{H} = \mathbf{P}_{Z^\perp}\mathbf{Y}^\perp(\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{Y}^\perp)^{-1}\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp} - I_T\chi^2_{1-\alpha}(k_2)/T$, where $\chi^2_{1-\alpha}(k_2)$ denotes the $(1-\alpha)\cdot 100\%$ quantile of a $\chi^2(k_2)$ distribution.

1.4.1 The Single RHS Endogenous Variable Case

When there is only one RHS endogenous variable, i.e., $n = 1$, Dufour[17] showed that valid $(1-\alpha)\cdot 100\%$ confidence sets for a LAU parameter must be unbounded with probability $(1-\alpha)$. We now demonstrate this result for confidence sets formed by inverting the AR, LR_{LIML} and $\mathcal{GMM}_0$ statistics for testing $H_0 : \beta = \beta_0$. As shown in Nelson, Startz and Zivot[41], the inversion of the AR, LR_{LIML}, and $\mathcal{GMM}_0$ statistics requires solving a quadratic inequality of the form

$$a\beta^2 + b\beta + c \leq 0,$$

from which it follows that the confidence sets will be unbounded when the coefficient a is negative, cover the entire real line when $b^2 - 4ac < 0$ as well, and they can even be empty when $a > 0$ and $b^2 - 4ac < 0$. Following this line of argument, we now give the conditions under which the confidence intervals obtained by inverting from the AR, LR_{LIML} and $\mathcal{GMM}_0$ statistics will be unbounded:

Proposition 1 *In the LISEM model given by (1.1)-(1.2),*[6]

(a) The $(1-\alpha)\cdot 100\%$ confidence interval $C_{AR}(\beta; 1-\alpha)$ obtained by inverting the AR statistic will be unbounded when the first stage F statistic for testing $H_0 : \mathbf{\Pi} = 0$ is insignificant at level α, i.e., $F_{\Pi=0} \leq F_{1-\alpha}(k_2, T-k)$.

(b) When the model is nearly non-identified, the $(1-\alpha)\cdot 100\%$ confidence interval $C_{AR}(\beta; 1-\alpha)$ will be unbounded with probability $(1-\alpha)$.

---

[6] When inverting the test statistics, the $\chi^2(k_2)$ bounding distribution is used when necessary, and the corresponding confidence intervals will be conservative as a result.

(c) The $(1-\alpha)\cdot 100\%$ *(conservative) confidence interval* $C_{LR}(\beta;1-\alpha)$ *obtained by inverting the* LR_{LIML} *statistic will be unbounded when the first stage* F *statistic for testing* $H_0:\mathbf{\Pi}=0$ *is insignificant at level* α*, i.e.,* $F_{\mathbf{\Pi}=0}\leq F_{1-\alpha}(k_2,T-k)$.

(d) The $(1-\alpha)\cdot 100\%$ *(conservative) confidence interval* $C_{LM}(\beta;1-\alpha)$ *obtained by inverting the* $\mathcal{GMM}_0$ *statistics will be unbounded when the first stage* LM *statistic for testing* $H_0:\mathbf{\Pi}=0$ *is insignificant at level* α*, i.e.,* $LM_{\mathbf{\Pi}=0}\leq\chi^2_{1-\alpha}(k_2)$.

(e) When the model is nearly non-identified, the $(1-\alpha)\cdot 100\%$ *(conservative) confidence intervals* $C_{LR}(\beta;1-\alpha)$ *and* $C_{LM}(\beta;1-\alpha)$ *will asymptotically be unbounded with probability close to* $(1-\alpha)$.

Result (a) is consistent with the common wisdom that the first stage F statistic provides a pretest for the relevance of the instruments. Results (b) and (e) confirm Dufour[17]'s statement that valid confidence sets for β must be unbounded with probability close to $(1-\alpha)$ in nearly non-identified models. Results (c) and (d) show that either the first stage F statistic or the first stage LM statistic (the TR^2 statistic) can be used as a pretest when LR or LM type test statistics are used to test the hypothesis $\mathbf{\Pi}=0$. If these first stage test statistics are insignificant, which implies that the instruments are "weak", the confidence intervals for β will typically be unbounded. Results from Monte Carlo experiments verifying these statements can be found in Nelson, Startz and Zivot[41].

1.4.2 An Alternative Inversion

As pointed out by Dufour[17], in the presence of normality, the likelihood ratio criterion[7] for testing $\beta=\beta_0$ is bounded by the Wilks statistic, a statistic commonly used

[7] The likelihood ratio statistics used by Dufour[17] are defined as the ratio of two maximized likelihoods, but the likelihood ratio statistics we have described so far are defined as -2 times the logarithm of the ratio. To avoid confusion, we will refer to the likelihood ratio statistics used by Dufour[17] as the likelihood ratio criterion (LRC) hereinafter.

in the analysis of the multivariate linear model. Hence, valid confidence intervals can also be obtained by inverting the likelihood ratio criterion for the LIML model. To see the point, consider the unrestricted reduced-form of the limited information model:

$$\bar{\mathbf{Y}} = \bar{\mathbf{Z}}\mathbf{B} + \epsilon,$$

where $\mathbf{B}$ is the matrix of the reduced-form coefficients, and ϵ is the vector of reduced-form error terms. The likelihood ratio criterion for testing the null hypothesis $\mathbf{B} = \bar{\mathbf{B}}$ is given by:

$$\overline{LRC} = \frac{L(\bar{\mathbf{B}})}{L(\mathbf{B})},$$

where $L(\bar{\mathbf{B}})$ is the maximum of the likelihood function of the multivariate model under the null hypothesis, and $L(\mathbf{B})$ is the maximum of the likelihood function under the alternative hypothesis. Similarly, the likelihood ratio criterion for testing the null hypothesis $\beta = \beta_0$ in the LIML framework is given by:

$$LRC = \frac{L(\beta_0)}{L(\beta)},$$

where $L(\beta_0)$ is the maximum of the likelihood function of the LIML model under the null hypothesis, and $L(\beta)$ is the maximum of the likelihood function under the alternative hypothesis. Since

$$L(\bar{\mathbf{B}}) \leq L(\beta_0) \leq L(\beta) \leq L(\mathbf{B}),$$

it follows that:

$$LRC = \frac{L(\beta_0)}{L(\beta)} \geq \frac{L(\bar{\mathbf{B}})}{L(\mathbf{B})} = \overline{LRC}.$$

$\overline{LRC}^{\frac{2}{T}}$ is the Wilks statistic, whose distribution is the distribution of the product $\prod_{i=1}^{p} b_i$, where p is the number of endogenous variables in the multivariate model, and b_i's are independent Beta random variables (see Rao[53], or Chapter 8 in Anderson[1]):

$$b_i \sim Beta\left(\frac{T-k-p+i}{2}, \frac{k}{2}\right), \; i = 1, \ldots, p.$$

thus the distribution of the Wilks statistic can be easily determined by simulation. Since $LRC = \hat{k}/k(\beta_0)$, where k is defined in equation (1.6), a conservative confidence interval for β can be obtained by solving the following inequality:

$$\Pr\{\frac{\hat{k}}{k(\beta)} \geq W_\alpha(T-k-2,k)\} \geq \Pr\{\overline{LRC} \geq W_\alpha(T-k-2,k)\} = 1-\alpha.$$

where $W_\alpha(T-k-2,k)$ is the α quantile of the distribution of the Wilks statistic. For example, a conservative 95% confidence interval for β is given by all values of β that satisfy

$$k(\beta) \leq \frac{\hat{k}}{W_{0.05}(T-k-2,k)} = C_W.$$

In comparison, conservative 95% LR_{LIML} confidence sets for β based on the $\chi^2(k_2)$ bounding distribution consist of all values of β satisfying[8]

$$k(\beta) \leq \hat{k}\exp\{\frac{\chi^2_{95\%}(k_2)}{T}\} = C_{LR}$$

By simulating $W_{0.05}(T-k-2,k)$ for various values of T and k, we find that $1/W_{0.05}(T-k-2,k)$ is always larger than $\exp\{\chi^2_{95\%}(k_2)/T\}$ and so confidence sets formed by inverting LR_{LIML} are always smaller than confidence sets formed by inverting LRC.

More generally, a transformation of the Wilks statistic follows a $\chi^2(2k)$ distribution asymptotically, i.e., [9]

$$-2\ln\overline{LRC} \Rightarrow \chi^2(2k).$$

This suggests that asymptotically valid confidence sets can be constructed based on LR_{LIML}, using $\chi^2(2k)$ as a bounding distribution. However, this obviously yields "wider" confidence sets than those based on the $\chi^2(k_2)$ bounding distribution.

---

[8] See Appendix B.

[9] See Section 8.5 in Anderson[1].

1.4.3 The Multiple RHS Endogenous Variables Case

When there are multiple RHS endogenous variables in the structural equation, i.e., $n > 1$, the function $f(\boldsymbol{\beta})$ in (1.8) defines either an ellipsoid or a hyperboloid in an n-dimensional space, as long as $\mathbf{Y}^{\perp\prime}\mathbf{H}\mathbf{Y}^{\perp}$ is nonsingular. Thus the joint confidence set $C(\boldsymbol{\beta}; 1-\alpha)$ will be either the "inside" or the "outside" of the ellipsoid (hyperboloid). Using a geometric argument,[10] it can be easily shown that the shape of the confidence set is determined by the eigenvalues of $\mathbf{Y}^{\perp\prime}\mathbf{H}\mathbf{Y}^{\perp}$, and the sign of the constant $\mathbf{y}^{\perp\prime}\mathbf{H}\mathbf{y}^{\perp} - \mathbf{y}^{\perp\prime}\mathbf{H}\mathbf{Y}^{\perp}(\mathbf{Y}^{\perp\prime}\mathbf{H}\mathbf{Y}^{\perp})^{-1}\mathbf{Y}^{\perp\prime}\mathbf{H}\mathbf{y}^{\perp}$ determines whether the confidence set lies inside or outside of the shape.

To illustrate the idea, suppose we have two right hand side endogenous variables ($n = 2$). The contour plots of $f(\boldsymbol{\beta})$ are shown in Figure 1.1. The joint confidence set for $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ is given by the set where $f(\boldsymbol{\beta}) \leq 0$. In Case I, both eigenvalues of $\mathbf{Y}^{\perp\prime}\mathbf{H}\mathbf{Y}^{\perp}$ are positive, and the confidence set is given by the inside of the ellipse; in Case II, both eigenvalues are negative, and the confidence set is given by the outside of the ellipse; in Case III and Case IV, one eigenvalue is positive and the other negative, thus the confidence set is given by either the "outside" or the "inside" of the hyperbola, depending upon whether the saddle path is above or below zero.

The above analysis is applied to the whole vector of $\boldsymbol{\beta}$. If we are interested in some elements of $\boldsymbol{\beta}$, or some function of $\boldsymbol{\beta}$, a projection method can be employed to make inference on the parameters of interest as suggested in Dufour[17] and Dufour and Kiviet[18]. For example, consider again the two RHS endogenous variables case illustrated in Figure 1.1 and suppose we are interested in testing the null hypothesis $H_0 : \beta_2 = \beta_2^0$. Generally, a valid test will involve the following: if the infimum of $f(\beta_1, \beta_2^0)$ is greater than zero, we reject the hypothesis; if the supremum of $f(\beta_1, \beta_2^0)$ is less than zero, we accept the hypothesis (or more precisely, cannot reject the hypothesis); otherwise, the evidence is inconclusive. Thus, in Case I, the projection

---

[10] For example, see Olmsted [45].

method involves the minimization of $f(\beta_1, \beta_2^0)$ over β_1, and if the minimum value is greater than zero, we reject the hypothesis. In Case II, the projection method involves the maximization of $f(\beta_1, \beta_2^0)$ over β_1, and if the maximum value is less than zero, we accept the hypothesis.

1.5 Monte Carlo Results

Monte Carlo experiments are conducted to evaluate the finite sample approximations to the asymptotic distributions of $\mathcal{GMM}_0$ and LR_{LIML} and to compare the coverage probabilities of 95% confidence sets constructed by inverting the AR, $\mathcal{GMM}_0$, LR_{LIML} and $Wald_{TSLS}$ test statistics when $n = 1$. The experiments are set up according to SS's design I. That is, data are generated from (1.1)-(1.2) with $\beta = 0, \gamma = 0$, $\mathbf{X}$ a vector of ones, $\mathbf{Z} \sim$ iid $N(0, I_{k_2})$ and $(u_i, V_i)' \sim$ iid $N(0, \bar{\Sigma})$. Results are reported for $k_2 = 1, 4$; $\rho = 0.5$, 0.99; $\lambda'\lambda/k_2 = 0$, 0.25, 1, 10 and are summarized in TABLE 1.1 and Figure 1.2.

The finite sample and asymptotic probability density functions of $\mathcal{GMM}_0$ (LM) and LR_{LIML} are plotted in Figure 1.2 for the case where $\rho = 0.99$, $\lambda'\lambda/k_2 = 0$. Recall that when $k_2 = 1$, the asymptotic distribution of both $\mathcal{GMM}_0$ and LR_{LIML} is a $\chi^2(1)$, and we note that when there are only five observations per instrument ($T/k_2 = 5$), the finite sample approximations are a little off, but when we have twenty observations per instrument ($T/k_2 = 5$), the finite sample approximations are very close to the asymptotic distributions. When $k_2 = 4$, the asymptotic distributions of $\mathcal{GMM}_0$ and LR_{LIML} are different, and again, when there are only fiver observations per instrument, the finite sample approximations are a little off, but the approximations become very good when we have twenty observations per instrument.

TABLE 1.1 gives the finite sample coverage rates of nominal 95% confidence sets formed by inverting the AR, $\mathcal{GMM}_0$, LR_{LIML} and $Wald_{TSLS}$ test statistics. When $k_2 = 1$, the asymptotic distributions of AR, $\mathcal{GMM}_0$, LR_{LIML} are $\chi^2(1)$ whereas the distribution of $Wald_{TSLS}$ depends on the nuisance parameters ρ, $\lambda'\lambda/k_2$ and k_2. The finite sample coverage rate of $\mathcal{GMM}_0$ is 95% for both values of ρ whereas the coverage rates of AR and LR_{LIML} are close to 85% when we have five observations per instrument and get to about 95% when we have twenty observations per instrument. The coverage rate of $Wald_{TSLS}$ is sensitive to the value of ρ and $\lambda'\lambda/k_2$ and. as noted by Hall, Rudebusch and Wilcox[26], Nelson, Startz and Zivot[41] and SS. is worst when ρ is close to 1 and $\lambda'\lambda/k_2$ is close to zero. When $k_2 = 4$, the asymptotic distributions of $\mathcal{GMM}_0$, LR_{LIML} and $Wald_{TSLS}$ are nonpivotal but the standard $\chi^2(1)$ distribution works well for $\mathcal{GMM}_0$ and LR_{LIML} if $\lambda'\lambda/k_2 = 10$. In all cases. the asymptotic $\chi^2(k_2)$ bounding distribution for $\mathcal{GMM}_0$ and LR_{LIML} holds in finite samples.

The $\mathcal{GMM}_0$ and LR_{LIML} confidence sets based on $\chi^2(k_2)$ critical values may be large if k_2 is large, so the power of the $\mathcal{GMM}_0$ and LR_{LIML} tests may be poor particularly if $\lambda'\lambda/k_2$ is large. To avoid such a possibility, Nelson. Startz and Zivot[41] suggested constructing $\mathcal{GMM}_0$ and LR_{LIML} confidence sets based on a pre-test of the significance of $\mathbf{\Pi}$ in the first stage regression (1.2). In particular, if $F_{\mathbf{\Pi}=0} < F(k_2, T - k_1 - k_2)$ (or $LM_{\mathbf{\Pi}=0} < \chi^2(k_2)$) use the $\chi^2(k_2)$ critical values to construct the confidence sets and if $F_{\pi=0} > F(k_2, T - k_1 - k_2)$ (or $LM_{\pi=0} > \chi^2(k_2)$) the $\chi^2(1)$critical values are used. The results for this pre-test based scheme is given in TABLE 1.1 under the columns labeled "switching". The coverage rates for LR_{LIML} are almost exact and the coverage rates for $\mathcal{GMM}_0$ are a little off only for the case when $\lambda'\lambda/k_2 = 1$.

1.6 Conclusion

For nearly non-identified models, Dufour[17] argued that confidence sets based on the familiar "estimate $\pm$ 2 $\cdot$ asymptotic standard error" may be highly misleading. The paper by SS demonstrated this phenomenon in instrumental variables regression with weak instruments. The asymptotic results in SS for t-statistics based on TSLS and

LIML, while highly informative, are not straightforward to apply in practice since they involve unknown values of nuisance parameters. In this chapter, we have shown how valid inference about the structural parameters can be conducted using the LR and LM type statistics which are asymptotically (boundedly) pivotal. The construction of valid confidence sets is also considered. The Monte Carlo results confirm that the method proposed in this chapter works well.

TABLE 1.1. Finte Sample Coverage Rates with 5% Nominal Size

$T / k_2 = 5$

Parameters			$\chi^2(1)$			$\chi^2(k_2)$			Switching	
ρ	$\lambda'\lambda / k_2$	k_2	Wald	GMM_0	LR	k_2 AR	GMM_0	LR	GMM_0	LR
0.50	0.00	1	0.88	0.95	0.84	0.85				
0.50	0.25	1	0.88	0.95	0.84	0.86				
0.50	1.00	1	0.87	0.95	0.84	0.85				
0.50	10.00	1	0.82	0.95	0.84	0.85				
0.50	0.00	4	0.81	0.85	0.72	0.90	1.00	0.95	0.98	0.93
0.50	0.25	4	0.82	0.87	0.75	0.90	1.00	0.96	0.98	0.93
0.50	1.00	4	0.85	0.91	0.82	0.90	1.00	0.98	0.97	0.93
0.50	10.00	4	0.91	0.94	0.92	0.90	1.00	0.99	0.94	0.92
0.99	0.00	1	0.31	0.95	0.84	0.86				
0.99	0.25	1	0.53	0.95	0.84	0.85				
0.99	1.00	1	0.67	0.95	0.84	0.85				
0.99	10.00	1	0.79	0.95	0.84	0.86				
0.99	0.00	4	0.01	0.51	0.72	0.90	0.97	0.95	0.96	0.91
0.99	0.25	4	0.13	0.62	0.92	0.90	0.98	0.99	0.93	0.96
0.99	1.00	4	0.42	0.78	0.93	0.90	0.99	1.00	0.86	0.96
0.99	10.00	4	0.85	0.93	0.93	0.90	1.00	1.00	0.93	0.94

$T / k_2 = 20$

Parameters			$\chi^2(1)$			$\chi^2(k_2)$			Switching	
ρ	$\lambda'\lambda / k_2$	k_2	Wald	GMM_0	LR	k_2 AR	GMM_0	LR	GMM_0	LR
0.50	0.00	1	0.97	0.95	0.93	0.94				
0.50	0.25	1	0.97	0.95	0.93	0.93				
0.50	1.00	1	0.96	0.95	0.93	0.93				
0.50	10.00	1	0.94	0.95	0.94	0.94				
0.50	0.00	4	0.86	0.86	0.77	0.94	0.99	0.97	0.97	0.96
0.50	0.25	4	0.86	0.88	0.80	0.94	0.99	0.98	0.97	0.95
0.50	1.00	4	0.88	0.91	0.87	0.94	1.00	0.99	0.95	0.95
0.50	10.00	4	0.93	0.95	0.94	0.94	1.00	1.00	0.94	0.94
0.99	0.00	1	0.36	0.94	0.93	0.93				
0.99	0.25	1	0.65	0.95	0.93	0.93				
0.99	1.00	1	0.79	0.94	0.93	0.93				
0.99	10.00	1	0.90	0.95	0.93	0.93				
0.99	0.00	4	0.01	0.57	0.77	0.94	0.95	0.98	0.95	0.94
0.99	0.25	4	0.14	0.66	0.94	0.94	0.96	1.00	0.90	0.97
0.99	1.00	4	0.44	0.80	0.95	0.94	0.99	1.00	0.83	0.97
0.99	10.00	4	0.87	0.93	0.95	0.94	1.00	1.00	0.93	0.95

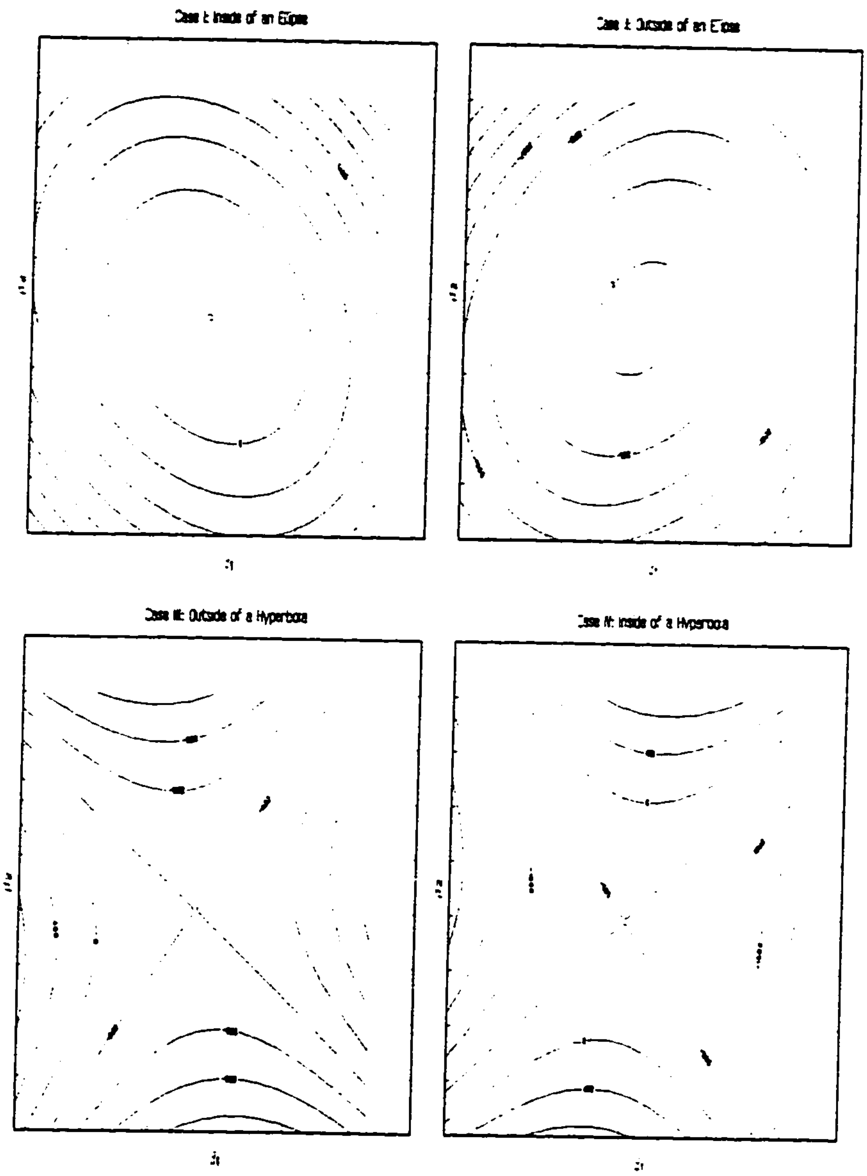

Figure 1.1: Contour Plots of Function $f(\beta)$

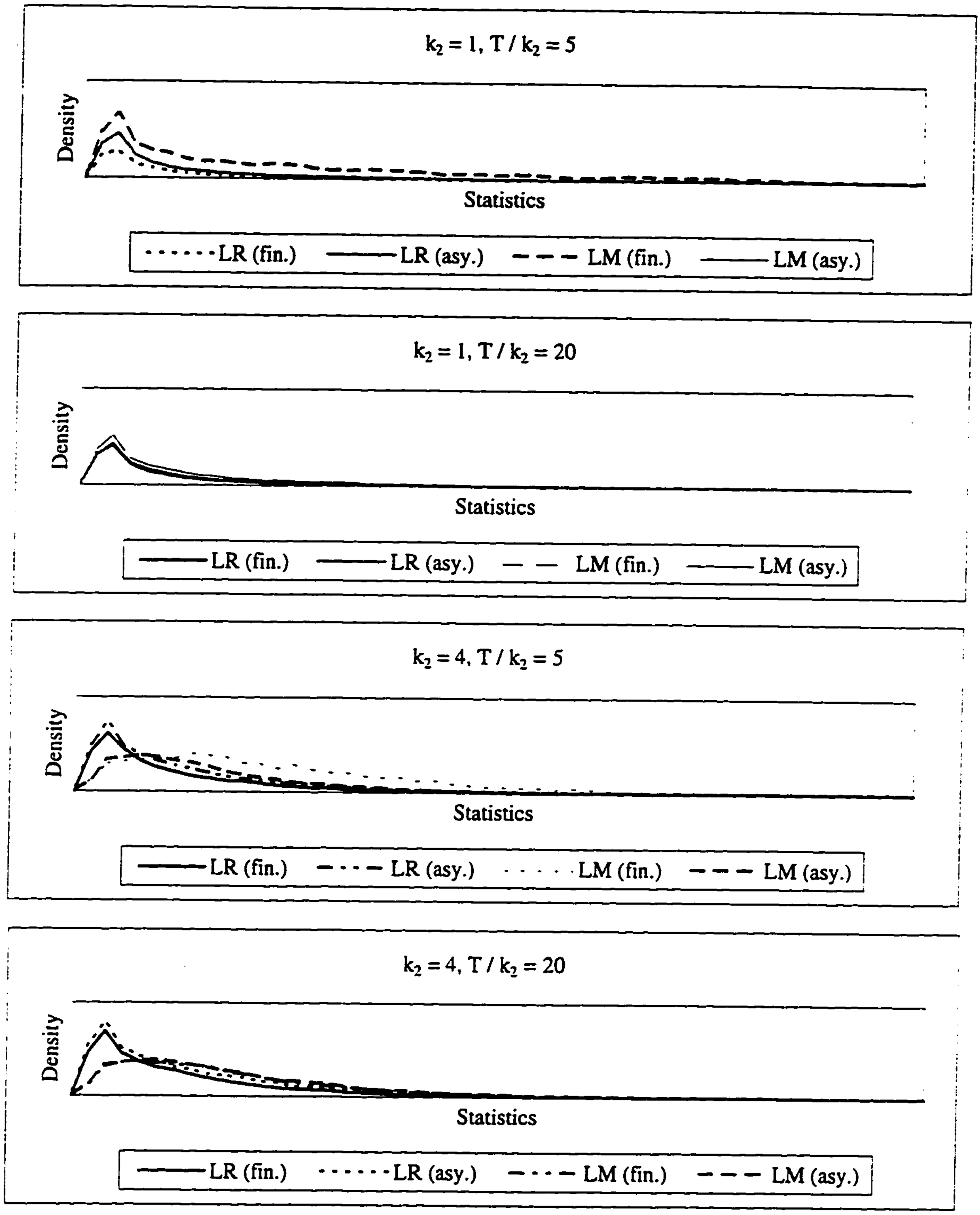

Figure 1.2 – Approximations of LR and GMM_0 Statistics

Chapter 2

A BAYESIAN TIME SERIES MODEL OF MULTIPLE STRUCTURAL CHANGES

2.1 Introduction

Since Perron[47] argued that most macroeconomic time series might be stationary around a broken trend, the detection of structural changes in time series data has attracted econometricians' attention. The follow-up literature emphasizes the testing of the unit-root hypothesis when there might be structural changes in the time series. For example, Banerjee, Lumsdaine and Stock[8], Perron and Vogelsang[48] and Zivot and Andrews[63] derived the asymptotic theory in this context. Zivot and Phillips[64] considered the same problem in a Bayesian framework. Andrews[4] analyzed the estimation and hypothesis testing of structural changes in general using generalized method of moments in linear and nonlinear models.

One drawback of these models is that only a one-time structural change is allowed in the trend of the time series. This assumption may be too restrictive for macroeconomic time series on account of two reasons: first, there are more than one historical event that might have led to a structural change: second, structural changes could occur in the variance term as well, for example, Engle[22] and Schwert[55] documented the changes of conditional volatility of some macroeconomic variables over time. However, the accommodation of multiple structural changes into a time series model is complicated. For instance, the estimators of multiple break points suggested by Bai and Perron[6] and Bai[5] have nonstandard distributions, which makes hypothesis testing difficult.

A related literature uses the intervention analysis advocated by Box and Tiao[12] to detect outliers and level shifts in time series data, for example, see Tsay[59] and the papers cited therein. In this framework, the usual procedure is "specification-estimation-detection-removal", which could handle multiple outliers and level shifts. However, the distributions of the test statistics are intractable when the true parameters are unknown.

The advantage of Bayesian approach to estimating structural changes has been acknowledged for a long time, as mentioned in Chapter 19 of Judge. Griffiths. Hill Lütkepohl and Lee[30]. For example, Raftery[52] suggested that Bayesian analysis in this context is "technically simpler", allows for "finite-sample inferences which are optimal given the framework", and allows "non-nested model comparisons". For a comprehensive review of the Bayesian econometric literature on structural changes. see Broemeling and Tsurumi[13].

In the Bayesian literature the structural change points have been modeled either explicitly or implicitly. For example, Raftery[52] suggested Markov transition distribution could be used to model the change points implicitly. Phillips and Smith[49] used a jump diffusion process to determine the number of structural changes. In contrast. Carlin, Gelfand and Smith[14] considered models with one structural change point by assuming a discrete uniform prior distribution for the change point to model a structural change explicitly.

In econometric analysis, we usually have some prior knowledge about the number of structural changes but we do not know for sure when the structural changes occurred. An econometric analysis will thus involve testing hypothesis about the number of structural changes and estimating the timing of structural changes. Therefore in this chapter we consider a time series model in which multiple structural changes are modeled explicitly with the number of structural changes assumed to be known. The structural changes could occur in both the time trend term and variance term. so the model nests many types of structural changes considered by economists. The estimation is made possible by the Gibbs sampling technique. Sample draws from Gibbs sampling also make it straightforward to estimate the numerical moments of the unknown parameters. Posterior odds ratio can be used for model selection. For example, when the number of structural changes is unknown, it can be determined by comparing posterior odds ratios of different specifications.

The rest of this chapter is organized as follows. In Section 2.2, the time series model of multiple structural changes is introduced and some notations are established. In Section 2.3 the algorithm of Gibbs sampling is developed for the purpose of estimating the model presented in Section 2.2. Section 2.4 gives a review of how to obtain numerical moments of the unknown parameters using the output from Gibbs sampling, then discusses the issue of model selection in our context. Section 2.5 presents some results from the simulation experiments to show the validity of the method developed in this chapter. In Section 2.6, this method is applied to two empirical data sets. First, we will investigate the structural changes in the regime of Chinese exchange rate. Second, we will consider the structural changes in the time series of U.S. real GDP from 1870 to 1994. Section 2.7 concludes the chapter.

2.2 A Time Series Model of Multiple Structural Changes

Consider a time series model with a linear time trend in the following form:

$$y_t = a_t + b_t t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_r y_{t-r} + w_t u_t, \tag{2.1}$$

where $u_t|\mathbf{\Omega}_{t-1}$ follows a standard normal distribution, with $\mathbf{\Omega}_t$ denoting the information set at time t, and $t = 1, 2, \cdots, T$. We assume that a_t, b_t and w_t are subject to structural changes, and structural changes could happen at m $(m < T)$ points of time: $1 < k_1 < k_2 < \cdots < k_m \leq T$, so that the observations can be divided into $m + 1$ regimes. For each regime i, where $i = 1, 2, \cdots, m + 1$, the parameters a_t, b_t and w_t are given by:

$$a_t = \alpha_i, b_t = \beta_i, w_t = \sigma_i, \text{ for } k_{i-1} \leq t < k_i,$$

where we define $k_0 = 1$ and $k_{m+1} = T + 1$, and we let $\mathbf{k} = (k_1, \cdots, k_m)'$.

The model given by (2.1) nests many types of structural changes considered by economists. For example, when there are changes only in a_t, the structural changes are usually known as "level shifts"; when there are changes only in b_t, the structural changes are usually known as "innovative outliers". In addition, we can also have structural changes in the variance term (or more precisely, the standard deviation) w_t. In spite of the complexity of the model, it can be written in a standard form by means of an indicator variable. Let I_A be an indicator variable such that I_A is one if A is true and zero otherwise, then equation (2.1) can be written as:

$$y_t = \sum_{i=1}^{m+1} \mathrm{I}_{k_{i-1} \le t < k_i}(\alpha_i + \beta_i t) + \sum_{j=1}^{r} y_{t-j}\phi_j + w_t u_t,$$

or

$$y_t = \mathbf{x}_t'\mathbf{B} + w_t u_t, \tag{2.2}$$

where $\mathbf{x}_t = (\mathrm{I}_{k_0 \le t < k_1}, \cdots, \mathrm{I}_{k_m \le t < k_{m+1}}, t\mathrm{I}_{k_0 \le t < k_1}, \cdots, t\mathrm{I}_{k_m \le t < k_{m+1}}, y_{t-1}, \cdots, y_{t-r})'$, and $\mathbf{B} = (\alpha_1, \cdots, \alpha_{m+1}, \beta_1, \cdots, \beta_{m+1}, \phi_1, \cdots, \phi_r)'$. Equation (2.2) is a standard linear regression model with heteroscedasticity, from which the likelihood function of the unknown parameters can be readily obtained as follows:

$$L(\theta|\mathbf{Y}, \mathbf{Y}_0) \propto (\prod_{t=1}^{T} w_t)^{-1} \exp\{-\frac{1}{2}\sum_{t=1}^{T} \frac{(y_t - \mathbf{x}_t'\mathbf{B})^2}{w_t^2}\} \tag{2.3}$$

$$\propto |\mathbf{W}|^{-1} \exp\{-\frac{1}{2}(\mathbf{Y} - \mathbf{XB})'\mathbf{W}^{-2}(\mathbf{Y} - \mathbf{XB})\}, \tag{2.4}$$

where $\propto$ denotes "proportional to", $\mathbf{Y}_0$ is the r initial observations of y_t, $\mathbf{Y} = (y_1, \cdots, y_T)'$, $\mathbf{W}$ is a diagonal matrix with $(w_1, w_2, \cdots, w_T)$ on the diagonal, $\mathbf{X}$ is an $T \times (2m + 2 + r)$ matrix with t-th row given by $\mathbf{x}_t'$, and $\boldsymbol{\theta}$ includes all the unknown parameters. For the remainder of this chapter, the conditioning on $\mathbf{Y}_0$ will be suppressed for notational brevity.

2.3 Bayesian Estimation with Gibbs Sampling

From a Bayesian perspective, the inference of the unknown parameters $\boldsymbol{\theta}$ can be made using the posterior distribution of $\boldsymbol{\theta}$. Once we know the posterior joint distribution of $\boldsymbol{\theta}$ and marginal distribution of elements of $\boldsymbol{\theta}$, we can obtain estimates of the unknown parameters and test hypotheses about those parameters. By Bayes theorem. the posterior density function of $\boldsymbol{\theta}$ is given by:

$$f(\boldsymbol{\theta}|\mathbf{Y}) = \frac{f(\mathbf{Y}|\boldsymbol{\theta})f_0(\boldsymbol{\theta})}{f(\mathbf{Y})} \propto L(\boldsymbol{\theta}|\mathbf{Y})f_0(\boldsymbol{\theta}),$$

where $f_0(\boldsymbol{\theta})$ denotes the prior density function of $\boldsymbol{\theta}$. Therefore, given a prior specification $f_0(\boldsymbol{\theta})$, our knowledge of the unknown parameters $\boldsymbol{\theta}$. which is embodied in the posterior density function $f(\boldsymbol{\theta}|\mathbf{Y})$, can be updated using information contained in $L(\boldsymbol{\theta}|\mathbf{Y})$, the likelihood function. Obviously, to derive the posterior distribution of $\boldsymbol{\theta}$ we need to specify a prior distribution for $\boldsymbol{\theta}$. For convenience. we will assume independence among groups of elements of $\boldsymbol{\theta}$ such that:

$$f_0(\boldsymbol{\theta}) = f_0(\mathbf{k})f_0(\mathbf{B})\prod_{i=1}^{m+1} f_0(\sigma_i^2). \tag{2.5}$$

From the likelihood function (2.3) or (2.4) it is obvious that the natural conjugate prior for $\mathbf{B}$ should be a multivariate normal distribution $N(\mathbf{B}_0, \Sigma_B)$, and the natural conjugate prior for σ_i^2 should be an inverted Gamma distribution $IG(v_0, \lambda_0)$ with the probability density function given by:

$$f_0(\sigma_i^2) \propto (\sigma_i^2)^{-(v_0-1)}\exp\{-\frac{\lambda_0}{\sigma_i^2}\}. \tag{2.6}$$

since for those observations in regime i, $w_t = \sigma_i$, for $i = 1, \cdots, m+1$. The prior distribution of $\mathbf{k}$ is assumed to be discrete uniform over all ordered subsequences of $(2, 3, \cdots, T)$ of length m.

Given those prior distributions, the posterior distribution of $\boldsymbol{\theta}$ is not of a standard form. However. sample draws from this posterior distribution can still be generated using Gibbs sampling. Gibbs sampling has been used extensively in statistics and econometrics in recent years since the seminar papers by German and German[24], Tanner and Wong[58] and Gelfand and Smith[23]. For an introduction to Gibbs sampling and related Markov Chain Monte Carlo methods. see Tanner[57] and Chib and Greenberg[15].

The idea of Gibbs sampling is: although the joint distribution of a multivariate random vector $\boldsymbol{\theta}$, $f(\boldsymbol{\theta}|\mathbf{Y})$, may be unknown, samples of $\boldsymbol{\theta}$ can still be generated if the draws from the full conditional distributions $f(\theta_i|\mathbf{Y}, [\boldsymbol{\theta}-\theta_i])$. where $[\boldsymbol{\theta}-\theta_i]$ denotes the rest of the vector $\boldsymbol{\theta}$ other than θ_i, can be easily generated. By iteratively generating samples from the complete full conditional distributions. the draws of the random variables from the full conditional distributions will finally converge to draws from the joint distribution $f(\boldsymbol{\theta}|\mathbf{Y})$, and the draws of a particular element of $\boldsymbol{\theta}$ will converge to draws from the marginal distribution of that element.[1] Before we show how to generate samples from the joint distribution, let us show that the full conditional distributions are standard ones.

First, let us consider the full conditional distribution of k_i, for $i = 1, 2, \cdots, m$:

$$f(k_i|\mathbf{Y}, [\boldsymbol{\theta} - k_i]) \propto L(\boldsymbol{\theta}|\mathbf{Y})f_0(\mathbf{k}) \propto L(k_i, [\boldsymbol{\theta} - k_i]|\mathbf{Y}), \tag{2.7}$$

because we assume the change points are discretely uniform. Thus sample draws of k_i from the conditional distribution $f(k_i|\mathbf{Y}, [\boldsymbol{\theta} - k_i])$ can be generated as a multinomial random variable with the sample space given by (k_{i-1}, k_{i+1}) and probabilities given by $L(k_i, [\boldsymbol{\theta} - k_i]|\mathbf{Y})$ up to a constant.

Second, let us consider the full conditional distribution of $\mathbf{B}$:

$$f(\mathbf{B}|\mathbf{Y}, [\boldsymbol{\theta} - \mathbf{B}]) \propto L(\boldsymbol{\theta}|\mathbf{Y})f_0(\mathbf{B}) \sim N(\boldsymbol{\Psi}_B(\boldsymbol{\Sigma}_B^{-1}\mathbf{B}_0 + \mathbf{X}'\mathbf{W}^{-2}\mathbf{Y}), \boldsymbol{\Psi}_B), \tag{2.8}$$

where $\boldsymbol{\Psi}_B = (\boldsymbol{\Sigma}_B^{-1} + \mathbf{X}'\mathbf{W}^{-2}\mathbf{X})^{-1}$. This can be easily checked using (2.4). Thus sample draws of $\mathbf{B}$ from the full conditional distribution $f(\mathbf{B}|\mathbf{Y}, [\boldsymbol{\theta} - \mathbf{B}])$ can be

[1] For proof of these results, see Liu, Wong and Kong[34, 35].

generated as a multivariate normal random vector.

Finally let us consider the full conditional distribution of σ_i^2:

$$f(\sigma_i^2|\mathbf{Y},[\boldsymbol{\theta}-\sigma_i^2]) \propto L(\boldsymbol{\theta}|\mathbf{Y})f_0(\sigma_i^2) \sim IG(v_i,\lambda_i), \tag{2.9}$$

where $v_i = v_0 + n_i/2$, $\lambda_i = \lambda_0 + (\mathbf{Y}^i - \mathbf{X}^i\mathbf{B})'(\mathbf{Y}^i - \mathbf{X}^i\mathbf{B})/2$, n_i is the number of observations in regime i, and $\mathbf{Y}^i$ is an $n_i \times 1$ vector of those y_t's in regime i. and $\mathbf{X}^i$ is an $n_i \times (2m+2+r)$ matrix with rows given by those $\mathbf{x}_t'$ in regime i. This can be easily checked using (2.3) and (2.6). Thus sample draws of σ_i^2 from the full conditional distribution $f(\sigma_i^2|\mathbf{Y},[\boldsymbol{\theta}-\sigma_i])$ can be generated as an inverted Gamma random variable, for $i = 1,2,\cdots,m+1$.

Now given these full conditional distributions, we summarize the algorithm of Gibbs sampling for generating sample draws from the posterior distribution $f(\boldsymbol{\theta}|\mathbf{Y})$:

Step 1 Specify starting values $\boldsymbol{\theta}^{(0)} = (\mathbf{k}^{(0)\prime},\mathbf{B}^{(0)\prime},\boldsymbol{\sigma}^{(0)\prime})'$, where $\boldsymbol{\sigma} = (\sigma_1,\cdots,\sigma_{m+1})'$. and set iteration number $j = 1$.

Step 2 Generate a draw of the first change point k_1 from the full conditional distribution $f(k_1^{(j)}|\mathbf{Y},k_2^{(j-1)},\mathbf{B}^{(j-1)},\boldsymbol{\sigma}^{(j-1)})$ given in (2.7) as a multinomial random variable.

Step i: $i = 3,\cdots,m+1$ Generate a draw of the $(i-1)$-th change point $k_{i-1}^{(j)}$ from the full conditional distribution $f(k_{i-1}^{(j)}|\mathbf{Y},k_{i-2}^{(j)},k_i^{(j-1)},\mathbf{B}^{(j-1)},\boldsymbol{\sigma}^{(j-1)})$ given in (2.7) as a multinomial random variable.

Step $m+2$ Generate a draw of $\mathbf{B}^{(j)}$ from the full conditional distribution $f(\mathbf{B}^{(j)}|\mathbf{Y},\mathbf{k}^{(j)},\boldsymbol{\sigma}^{(j-1)})$ given in (2.8) as a multivariate normal random vector.

Step $m+3$ Generate a draw of $(\sigma_i^2)^{(j)}$ from the full conditional distribution $f((\sigma_i^2)^{(j)}|\mathbf{Y},\mathbf{k}^{(j)},\mathbf{B}^{(j)})$ as given in (2.9), for $i = 1,2,\cdots,m+1$. as an inverted Gamma random variable.

Step $m+4$ Set the iteration number $j = j+1$ and go to Step 2.

After the algorithm converges, we can save the Gibbs draws as draws from the joint distribution $f(\boldsymbol{\theta}|\mathbf{Y})$, and sample draws of each element of $\boldsymbol{\theta}$ can be treated as draws from the marginal distribution of that element.

2.4 *Bayesian Inference via Gibbs Sampling*

2.4.1 The Estimation of Numerical Moments of Unknown Parameters

In practice there are two ways to obtain sample draws from the joint distribution $f(\boldsymbol{\theta}|\mathbf{Y})$. One way is to save a sample draw after the convergence has occurred, then restart the iteration to obtain another sample draw. This process can be repeated M times to obtain M sample draws. Sample draws obtained this way are independent. The other way is to continue the same iteration for M times after the convergence has occurred to obtain M sample draws. Sample draws obtained this way are correlated. The second method is employed in this chapter.

When the second method is employed, under the conditions of convergence, the M sample draws of any element θ_i of $\boldsymbol{\theta}$ are correlated observations from a stationary and ergodic process whose distribution is given by the posterior marginal distribution of θ_i.[2] Thus, given the sample draws $(\theta_i^{(1)}, \theta_i^{(2)}, \cdots, \theta_i^{(M)})$ of θ_i from Gibbs sampling, the numerical mean of the M sample draws:

$$\hat{\theta}_i = \frac{1}{M}\sum_{j=1}^{M}\theta_i^{(j)}$$

is a consistent estimator of the mean of θ_i. To obtain an estimate of the variance of

[2] Some authors prefer to take every q-th draw after the convergence has occurred, which is described as "thinning the chain", to obtain "independent" draws. However, the usual ways of picking q are "ad hoc", and the impacts of "thinning the chain" may be bad, as argued by Geyer[25]. Therefore, we choose to save all the draws in this chapter.

θ_i, the following estimate is used in this chapter:

$$\hat{\text{var}}(\theta_i) = \sum_{j=-(M_w-1)}^{M_w-1} (1 - \frac{|j|}{M_w})\rho_j,$$

where ρ_j is the j-th order autocovariance of θ_i. This formula calculates the spectral density function of θ_i evaluated at frequency zero, using a Bartlett lag window with bandwidth M_w. Throughout this chapter, I will choose $M_w = 4 \cdot (T/100)^{\frac{1}{4}}$. For other alternative estimates. see Andrews[3], Newey and West[43] and Geyer [25].

2.4.2 Model Selection

In the Bayesian framework, hypothesis testing or model selection can be conducted by computing posterior odds ratio. For example, given a sample observation $\mathbf{Y} = (y_1, y_2, \cdots, y_T)'$. to test the hypothesis H_i against the alternative hypothesis H_j. the posterior odds ratio is given by:

$$K_{ij} = \frac{\Pr(H_i|\mathbf{Y})}{\Pr(H_j|\mathbf{Y})} = \frac{\Pr(H_i)}{\Pr(H_j)} \frac{f(\mathbf{Y}|H_i)}{f(\mathbf{Y}|H_j)}.$$

The first fraction in the last equation is the prior odds ratio, and the second fraction is the Bayes factor. which is the ratio of two marginal likelihood values. If we specify the prior odds ratio to be one to signify our ignorance. the posterior odds ratio reduces to the familiar likelihood ratio. If we further specify a "symmetric" loss structure,[3] we would accept the hypothesis H_i if K_{ij} is greater than 1 and reject the hypothesis H_i if K_{ij} is less than 1.

When the number of structural changes is unknown, posterior odds ratio can be used to determine the number of structural changes in the data. Let H_i denote the hypothesis that there are i structural changes in the data, i.e., $m = i$. To determine how many structural changes there are, we can calculate the posterior odds ratio K_{ij} for the values of i and j we are interested in. However, to be able to calculate the

[3] See Chapter 10 in Zellner[62].

posterior odds ratio, we first need a reasonable prior odds ratio. For this purpose, let us assume that a structural change would occur with probability p at any point of time. Obviously this assumption is consistent with our discrete uniform prior for the change points **k**. Under this assumption, the prior probability $\Pr(H_i)$ is given by:

$$\begin{aligned}\Pr(H_i) &= \int_0^1 \binom{T}{i} p^i (1-p)^{T-i} dp \\ &= \binom{T}{i} B(i+1, T-i+1) \\ &= \binom{T}{i} \frac{\Gamma(i+1)\Gamma(T-i+1)}{\Gamma(T+2)},\end{aligned}$$

where $B(\cdot,\cdot)$ and $\Gamma(\cdot)$ denote the Beta function and the Gamma function respectively. Since $i+1$ and $T-i+1$ both are positive integers, we further get:

$$\Pr(H_i) = \binom{T}{i} \frac{i! \cdot (T-i)!}{(T+1)\cdot T!} = \frac{1}{T+1},$$

for $i = 0, 1, \cdots, T$. Therefore, a prior odds ratio of one is consistent with our assumption that a structural change may occur with probability p at any point of time. To calculate the Bayes factor, we need the marginal likelihood value $f(\mathbf{Y}|H_i)$ for each hypothesis H_i, which does not have a tractable expression. However, it can be easily estimated as the harmonic mean of the likelihood values from the Gibbs sampling:[4]

$$f(\mathbf{Y}|m=i) = [\frac{1}{M}\sum_{j=1}^{M} f(\mathbf{Y}|m=i, \theta^{(j)})^{-1}]^{-1}. \tag{2.10}$$

Now using a prior odds ratio of one, the natural logarithm of the posterior odds ratio can be expressed as:

$$\log_e K_{ij} = \log_e f(\mathbf{Y}|m=i) - \log_e f(\mathbf{Y}|m=j).$$

Thus the number of structural changes can be determined by comparing the logarithms of the marginal likelihood values for each hypothesis. The magnitude of the

[4] See Newton and Raftery[44] for reference. For other methods of calculating the Bayes factor. see Kass and Raftery[32].

difference reveals our confidence in favor of one hypothesis against another. Kass and Raftery[32] provide a rule of thumb for interpreting the magnitude of this value, which is reproduced in TABLE 2.1 for convenience.

2.5 *Application to Simulated Data*

To confirm the validity of the Bayesian approach presented in the last couple of sections, in this section we apply the method to a simulated time series. The time series y_t is generated according to equation (2.1) with $\mathbf{k} = (103, 143)'$, $\alpha_1 = \alpha_2 = \alpha_3 = 3$, $\beta_1 = \beta_2 = 0.9$, $\beta_3 = 0.8$, $\sigma_1 = 3$, $\sigma_2 = \sigma_3 = 1$, $\phi_1 = 0.6$, and $\phi_2 = -0.2$. The first 50 observations were discarded, and we saved the next 150 observations, so the actual change points are $\mathbf{k} = (53, 93)'$. In this simulated series, the first structural change is designed to be a decrease in the variance term, and the second structural change is designed to be an innovative outlier only.

The approach presented in Sections 2.3 and 2.4 is employed to estimate a model with two autoregressive terms and two structural changes. To signify our ignorance of the data generating process, diffuse priors are used for $f_0(\mathbf{B})$ and $f_0(\sigma_i^2)$ such that $\mathbf{B}_0 = 0$, $v_0 = 2.001$, $\lambda_0 = 0.001$, and $\mathbf{\Sigma}_B$ is set to a diagonal matrix with elements 1000 on the diagonal. The starting values of $\mathbf{k}$ are set at the two (approximately) equidistant points between 2 and 150. Then the starting values of $\mathbf{B}$ and σ_i^2 are computed as in a standard linear model with group heteroscedasticity, for example, see Chapter 11 of Judge, Griffiths, Hill, Lütkepohl and Lee[30]. After running the Gibbs sampler with those starting values for 200 times, we save the next 2000 draws for inference. The posterior probabilities of the two change points are plotted in Figure 2.1, together with the simulated time series. Numerical moments of other unknown parameters are reported in TABLE 2.2, together with the 95% empirical confidence intervals.[5]

[5] We also calculated the 95% highest posterior density (HPD) regions according to Wei and Tanner[60], which turned out to be close to the 95% empirical confidence intervals in most cases.

From Figure 2.1, it is obvious that the second change point is estimated perfectly, while the estimate of the first change point is concentrated over a few points, but the highest posterior probability is located at 53, which is the true change point. From TABLE 2.2, we can see that the estimates of other parameters are very close to the true values of those parameters, and the 95% empirical confidence intervals are tight and all contain the true values. From the estimates in Table 2.2. we can also see that the first structural change involves a decrease in the variance term with no significant change in other parameters. It seems that the second structural change involves changes in both the intercept and the slope of the time trend. but given the sizes of standard deviations of the intercept terms, the level shift does not appear to be significant.

To test the hypothesis that there are exactly two structural changes. we also estimate a model with no structural change ($m = 0$), a model with one structural change ($m = 1$) and a model with three structural changes ($m = 3$). Using the harmonic mean of the likelihood values from the Gibbs sampling output as the estimate of the marginal likelihood. we obtain:

$$\log_e f(Y|m = 0) = -359.10.$$

$$\log_e f(Y|m = 1) = -294.04.$$

$$\log_e f(Y|m = 2) = -279.25.$$

$$\log_e f(Y|m = 3) = -280.84.$$

Obviously the model with two structural changes and three structural changes are superior, with the $m = 2$ model having a slightly larger marginal likelihood value. However, for the $m = 3$ model, the second and the third change points are estimated at 92 and 93 with posterior probability 0.579 and 0.576. respectively. Generally when the specified number of structural changes are more than the actual number of structural changes, one structural change is estimated at either the beginning of the observations, or the end of the observations, or right next to another structural change point. Therefore, we can conclude that the model with two structural changes is more appropriate.

2.6 Applications to Empirical Data

2.6.1 Structural Changes in the Exchange Rate Regime of Chinese Yuan

The exchange rate regime of Chinese yuan has undergone several changes since the economic reform took place in China in late 1970s.[6] Before January 1. 1986. the exchange rate of yuan was adjusted in accordance with movements in the value of a basket of internationally traded currencies, weighted with reference to their importance in China's external transactions and the trends in their relative values. Published rate between Chinese yuan and U.S. dollar was changed whenever the calculated rate diverged from the previously published rate by a margin of 0.5 cents. After January 1, 1986. China had formally followed an exchange arrangement whereby the exchange rate for yuan was based on developments in the balance of payments and in costs and exchange rates of China's major competitors. However, the exchange rate between Chinese yuan and U.S. dollar remained unchanged from July 1986 to December 1989. On December 15, 1989, yuan was announced to be depreciated by 21.2%. effective December 16. On November 17, 1990, yuan was further depreciated by 9.6%. Since then the official exchange rate for yuan had been adjusted periodically. Finally, the exchange rate for yuan became a single, managed floating rate based demand and supply in the foreign exchange market on January 1, 1994.

According to the Chinese government's announcements, there are three regime changes: one in December 1989, one in November 1990, and one in January 1994. However, there might be another regime change in July 1986. In this subsection the Bayesian approach is employed to determine how many structural changes there are

[6] For more information, see IMF Annual Report on Exchange Arrangements and Exchange Restrictions.

in the Chinese exchange rate regime. For this purpose, we will use monthly exchange rate between Chinese yuan and U.S. dollar from October 1980 to July 1995, taken from the Monthly Bulletin of Statistics of the United Nations. We estimate four different specifications for the logarithm of the Chinese exchange rate: a model with 3 structural changes ($m = 3$), a model with 4 structural changes ($m = 4$), a model with 5 structural changes ($m = 5$) and a model with 6 structural changes ($m = 6$). The autoregressive lag is chosen to be 1 according to diagnostic checks.

To signify our ignorance of the data generating process, the same priors as in the previous section are used. The starting values are chosen in the same way. After running the Gibbs sampler for 500 times, we save the next 2000 draws for inference. Again, using (2.10) we obtain the natural logarithms of the marginal likelihood values for the four different model specifications:

$$\log_e f(\mathbf{Y}|m = 3) = 430.63,$$
$$\log_e f(\mathbf{Y}|m = 4) = 490.28,$$
$$\log_e f(\mathbf{Y}|m = 5) = 538.06,$$
$$\log_e f(\mathbf{Y}|m = 6) = 481.58.$$

Obviously, the model with $m = 5$ is far superior than other specifications. Moreover, when $m = 6$, the fifth and sixth change points are estimated at January 1994 and February 1994 with probability 0.995 and 0.984, respectively. This is another piece of evidence that six structural changes are more than enough.

When $m = 5$, the posterior probabilities of the five change points are plotted in Figure 2.2, together with the time series of the exchange rate. In addition to the three government announced changes in December 1989, November 1990 and January 1994, two more structural changes are detected in the early time periods: one in September 1984 and the other in July 1986. The numerical moments of other parameters are reported in TABLE 2.3, together with their empirical 95% confidence intervals. All the estimates are significant. The first structural change appears to be an innovative outlier and a decrease in the variance term. The second structural change coincides with our visual inspection of the data. The next three structural changes coincide with the Chinese government's announcements perfectly.

2.6.2 *Structural Changes in U.S. Real GDP*

To investigate the evidence of structural changes in US real GDP, in this subsection annual data of US real GDP from 1870 to 1994, taken from Maddison[37], will be studied. The time series is plotted in Figure 2.3. Visually we can see that there is a dip during the Great Depression, then there is a faster growth period during World War II, and after World War II the real GDP seems to have gone back to the pre-Great-Depression trend.

Our visual inspection of the data suggests that there might be two structural changes in US real GDP: one around 1930 and the other at the end of World War II. To decide how many structural changes there are in the time series of US real GDP, we estimate three different specifications: $m = 0$, $m = 1$ and $m = 2$. The autoregressive lag is selected to be 2 according to diagnostic checks. The same prior distributions are used as in previous sections, and the starting values are chosen in the same way. After running the Gibbs sampler for 500 times, we save the next 2000 draws for inference. The natural logarithms of the marginal likelihood values are given by:

$$\log_e f(Y|m = 0) = 186.89,$$
$$\log_e f(Y|m = 1) = 204.08,$$
$$\log_e f(Y|m = 2) = 204.76.$$

Obviously the hypothesis of no structural change can be rejected, but the marginal likelihood value of the $m = 2$ model is only slightly larger than that of the $m = 1$ model. Further investigation of the Gibbs sampling output reveals that when $m = 2$, the estimate of the first change point clusters at the beginning of the series with a posterior probability of 0.895, which implies that the specification $m = 1$ is probably more appropriate.

When $m = 1$, the posterior probability of the only change point is plotted in Figure 2.3 together with the time series. The numerical moments of other parameters are reported in TABLE 2.4. From Figure 2.3 we can see that the structural change happened somewhere between 1947 and 1951, with the highest posterior probability being 0.18 at 1947. From the estimates in TABLE 2.4, the structural change appears to be a decrease in the variance term, with no significant changes in the time trend!

The above results seem to be contradictory with our visual inspection of the data. However, behind our visual inspection of the time series, we implicitly assumed a constant variance over time. In contrast, the variance term w_t is allowed to be subject to structural changes as well in the above models, and this might have significant effects on the estimation of change points. To confirm this conjecture, we then estimate a model with two structural changes in the time trend, with the variance term w_t constant over time. The autoregressive lag is also selected to be 2. The posterior probabilities of the two change points are plotted in Figure 2.4, and the numerical moments are reported in TABLE 2.5.

From Figure 2.4 we can see that the two structural change points are estimated at 1930 and 1945 with very high posterior probabilities. The estimates in TABLE 2.5 are consistent with our visual inspection of the data: the pre-Great-Depression period and the post-World-War-II period share the same time trend, whereas the period between the Great Depression and World War II exhibits a drop in the level and a faster growth rate. However, the natural logarithm of the marginal likelihood value for this model is estimated to be 204.76, which is very close to the value we obtained for the $m = 1$ model with changes in w_t allowed. In short, when we allow the variance term to be subject to structural changes, only one structural change is detected in the US real GDP, which involves a decrease in the variance term after World War II; however, when we do not allow the variance term to change over time, two structural changes are detected in the data, which correspond to the Great Depression and World War II, respectively. There is no evidence that one model dominates the other.

2.7 Conclusion

In this chapter a Bayesian approach to estimating a time series model with multiple structural changes is developed. The structural changes could occur in both the time trend and the variance term, so different types of structural changes are nested in one simple model. Although the posterior distribution of the unknown parameters is nonstandard, we can show that the complete full conditional distributions are of standard forms, with appropriate choice of prior distributions. As a result, Gibbs sampling can be employed to generate sample draws from the posterior distribution of the unknown parameters. Numerical moments of the unknown parameters can be easily obtained from the output of Gibbs sampling. When the number of structural changes is unknown, posterior odds ratio can be used to determine the number of structural changes. The application of this method to a simulated series confirms the validity of this approach. The method is then applied to two empirical data sets: the exchange rate of Chinese yuan and the U.S. real GDP from 1870 to 1994. In addition to three structural changes corresponding to Chinese government's announcements, two more structural changes are detected in the early time periods for the exchange rate of Chinese yuan against US dollar. For the U.S. real GDP, if the variance term is allowed to change over time, only one structural change is detected, which corresponds to a decrease in variance since World War II; in contrast, if the variance term is assumed to be constant, two structural changes are detected, with the first corresponding to the Great Depression which changed the behavior of the time trend, and the second corresponding to the ending of World War II, since when the time trend went back to the pre-Great-Depression structure. However, both models yield similar marginal likelihood values.

TABLE 2.1 – Interpretation of 2 $\log_e K_{ij}$

2 $\log_e Kij$	Evidence Against H_j
0 to 2	Not worth more than a bare mention
2 to 5	Positive
5 to 10	Strong
> 10	Decisive

TABLE 2.2 – Estimates of the Parameters of the Simulated Series

Parameters	Mean	Standard Deviation	95% Interval
α_1	52.17	2.94	(45.67 58.50)
α_2	50.98	3.06	(44.43 57.11)
α_3	46.84	2.92	(40.38 52.80)
β_1	0.95	0.06	(0.82 1.08)
β_2	0.96	0.06	(0.84 1.08)
β_3	0.84	0.05	(0.74 0.94)
σ_1	3.17	0.38	(2.58 3.97)
σ_2	1.19	0.18	(0.92 1.53)
σ_3	1.62	0.08	(1.33 1.67)
ϕ_1	0.61	0.06	(0.49 0.73)
ϕ_2	-0.24	0.05	(-0.33 -0.14)

TABLE 2.3 – Estimates of the Parameters of Chinese Exchange Rate

Parameters	Mean	Standard Deviation	95% Interval
α_1	0.456	0.012	(0.429 0.479)
α_2	0.475	0.058	(0.366 0.571)
α_3	1.264	0.025	(1.218 1.304)
α_4	1.541	0.110	(1.328 1.757)
α_5	1.218	0.034	(1.160 1.272)
α_6	2.577	0.062	(2.461 2.694)
β_1	0.007	0.000	(0.006 0.008)
β_2	0.010	0.001	(0.008 0.012)
β_3	0.000	0.000	(-0.000 0.000)
β_4	0.000	0.001	(-0.002 0.001)
β_5	0.003	0.000	(0.003 0.003)
β_6	-0.003	0.000	(-0.004 -0.002)
σ_1	0.033	0.003	(0.027 0.041)
σ_2	0.023	0.005	(0.017 0.032)
σ_3	0.005	0.000	(0.004 0.006)
σ_4	0.010	0.002	(0.007 0.015)
σ_5	0.011	0.001	(0.009 0.013)
σ_6	0.008	0.001	(0.006 0.011)
ϕ_1	0.037	0.019	(0.006 0.072)

TABLE 2.4 – Estimates of the Parameters of U.S. Real GDP with Changes in w_t

Parameters	Mean	Standard Deviation	95% Interval
α_1	0.597	0.147	(0.312 0.880)
α_2	0.638	0.185	(0.335 0.958)
β_1	0.006	0.002	(0.003 0.009)
β_2	0.006	0.001	(0.003 0.008)
σ_1	0.061	0.005	(0.051 0.073)
σ_2	0.022	0.003	(0.017 0.027)
ϕ_1	1.130	0.091	(0.964 1.294)
ϕ_2	-0.315	0.100	(-0.481 -0.157)

TABLE 2.5 – Estimates of the Parameters of U.S. Real GDP with Constant w_t

Parameters	Mean	Standard Deviation	95% Interval
α_1	0.952	0.196	(0.644 1.281)
α_2	-0.282	0.586	(-0.841 0.423)
α_3	0.951	0.227	(0.626 1.308)
β_1	0.011	0.002	(0.007 0.015)
β_2	0.028	0.007	(0.020 0.037)
β_3	0.010	0.002	(0.007 0.014)
σ	0.044	0.003	(0.038 0.050)
ϕ_1	0.899	0.110	(0.732 1.064)
ϕ_2	-0.204	0.119	(-0.388 -0.035)

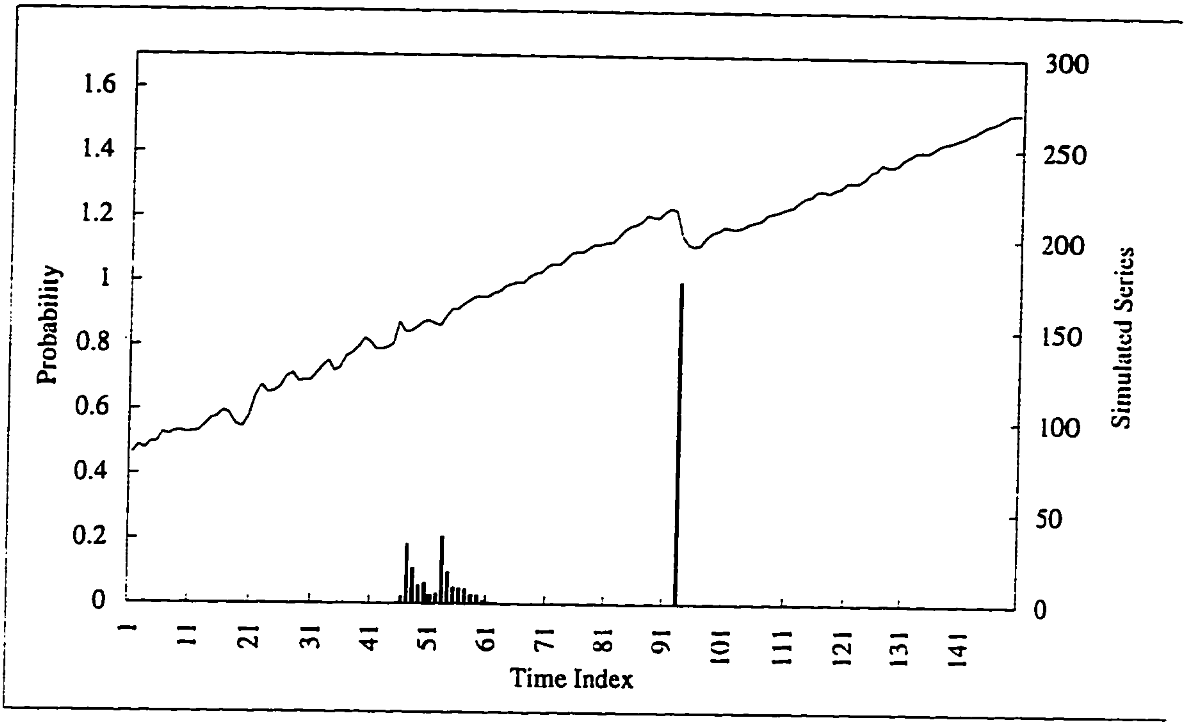

Figure 2.1 – Simulated Series and Estimated Change Points

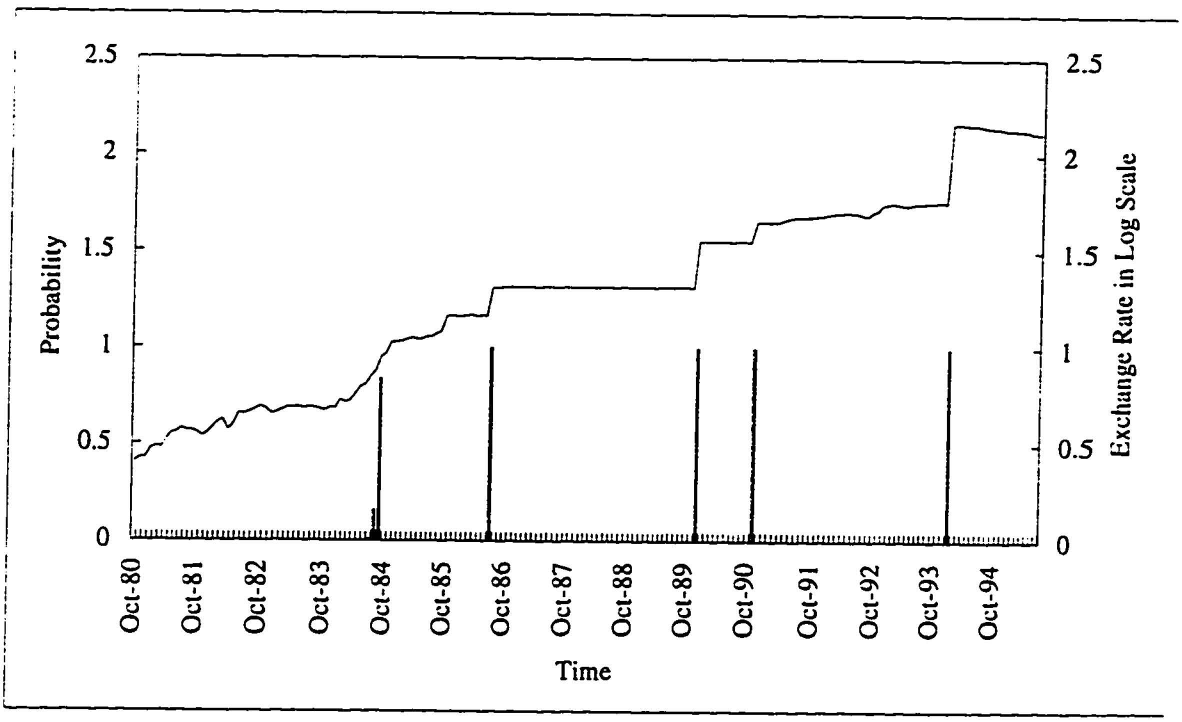

Figure 2.2 – Exchange Rate Between Chinese Yuan and U.S. Dollar

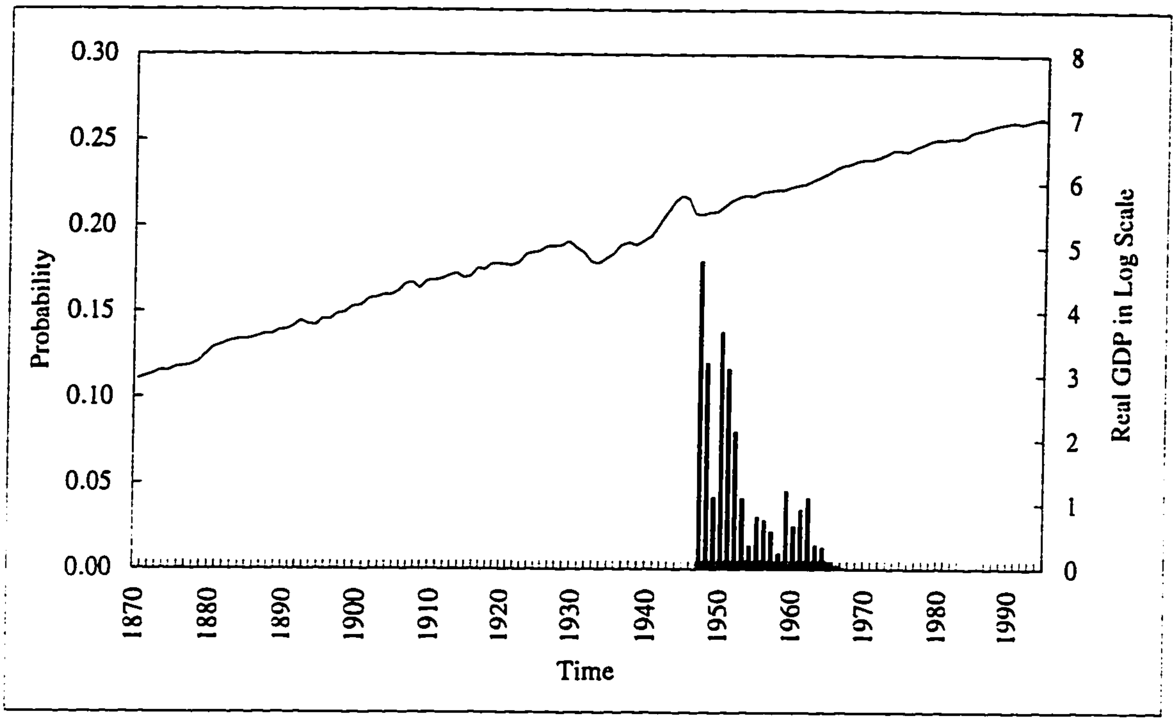

Figure 2.3 – U.S. Real GDP and Change Point (with Changes in w_t)

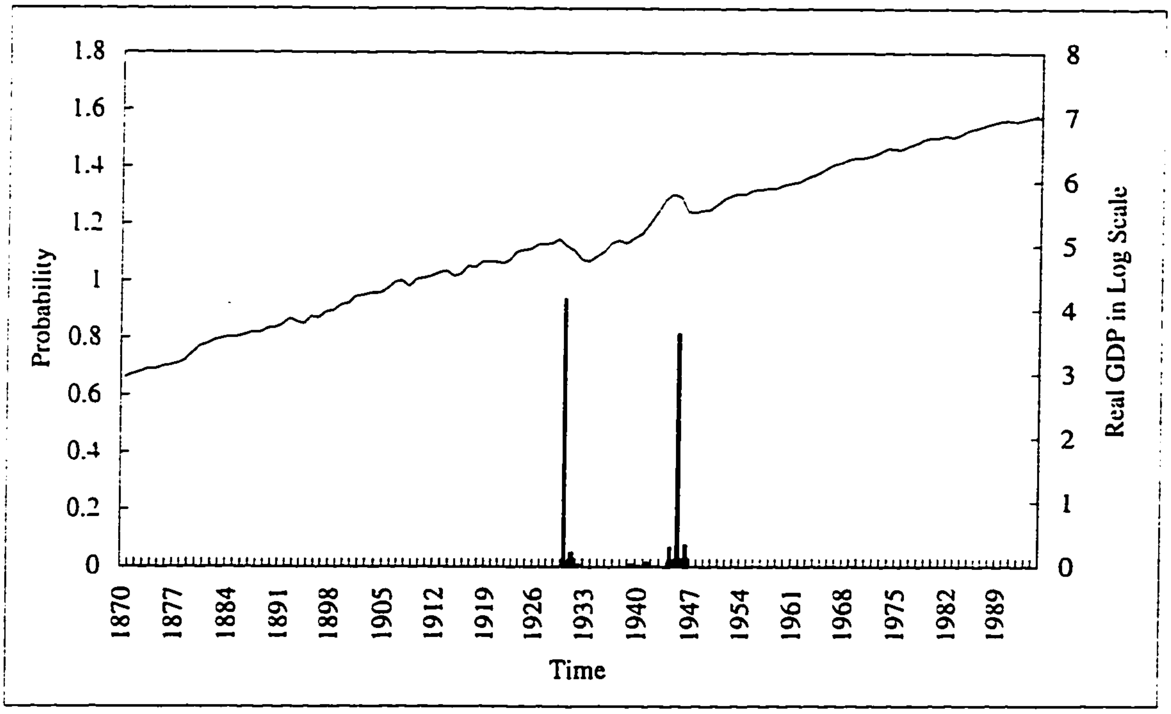

Figure 2.4 – U.S. Real GDP and Change Points (with Constant w_t)

Chapter 3

THE STATISTICAL DISTRIBUTIONS OF HIGH FREQUENCY FOREIGN EXCHANGE RATES

3.1 Introduction

Since Engel and Hamilton[21] suggested the long swings in foreign exchange rates could be modeled as a Markov switching process, more evidence of regime switching has been found in foreign exchange rate data, for example, see Kaminsky[31], Engel[19] and Engel and Hakkio[20]. Among those studies, Engel[19] fit a two state Markov switching model to eighteen quarterly (and monthly) exchange rates and found that a two state Markov switching model fit well in sample but could not outperform the random walk model in out-of-sample forecasting. However, there is no systematic study on the performance of Markov switching model on high frequency daily data. As pointed out by Engel[19], if the exchange rates sampled at a certain frequency follow a two state Markov switching process, that does not necessarily imply exchange rates at different frequencies follow a two state Markov switching process as well. Therefore, there is a need to investigate the performance of Markov switching model on high frequency exchange rates.

For most of the regime switching models considered by the economists, the regimes are assumed to switch according to a first order Markov chain. However, for high frequency data it is natural to assume that the influence of the behavior of the exchange rate will linger for a few days. That implies a high order Markov switching model might fit the data better. Although theoretically any high order Markov switching model can be rewritten, and thus estimated, as a first order Markov switching model, the resulting dimensionality problem makes the approach intractable and impractical. To overcome this problem, in this chapter a parsimonious high order Markov chain proposed by Raftery[51] will be used to model the underlying Markov chain.

Using high order Markov switching models this chapter then explores the evidence of high order Markov switching in high frequency exchange rates data. As a comparison, the chapter starts by examining eight weekly foreign exchange rates. However, no evidence of high order Markov switching is found. It appears that most weekly foreign exchange rates follow a first order Markov switching process. Then Markov switching models of different orders are fitted for daily exchange rates. Not surprisingly, consistent evidence of high order Markov switching is found in most daily exchange rates. In fact, Schwarz Bayesian criterion (SBC) favors the third order Markov switching model for seven out of the eight exchange rates investigated. Therefore it appears that the influence of daily behavior of exchange rates will usually remain for three business days.

There are previous studies that investigate the statistical distributions of daily exchange rates, for example, see Bollerslev[10], Hsieh[28, 29], Baillie and Bollerslev[7] and West and Cho[61]. The general message is that a standard GARCH(1,1) model can successfully remove the conditional heteroskedasticity in daily exchange rates movement, but cannot explain the leptokurtosis very well. However, a GARCH(1,1) model with conditionally t-distributed errors is better at explaining the "fat tails". In addition, with a standard GARCH(1,1) model, Hsieh[29] found that the day-of-the-week effects and holiday effects are insignificant in the mean equation, but significant in the variance equation. In contrast, when using a GARCH(1,1) model with conditionally t-distributed errors, Baillie and Bollerslev[7] reported that most of the variables for the day-of-the-week effects and holiday effects are not significant. That is understandable since t-distributed errors can be treated as following continuous normal mixture distributions.

To examine the forecasting performance of Markov switching model for daily exchange rates, this chapter then compares both in-sample and out-of-sample 1-day ahead, 1-week ahead and 2-week ahead forecasts of the conditional volatilities from the Markov switching models with those from the random walk model and a GARCH(1,1) model with conditionally t-distributed errors. Generally the Markov switching models yield smaller forecast errors than the random walk model. and also have slight advantage over the GARCH(1,1) model. However, the third order Markov switching model does not have obvious advantage over the first order Markov switching model for most exchange rates considered.

This rest of this chapter is organized as follows: Section 3.2 presents the high order Markov switching model using Raftery[51]'s formulation for the hidden Markov chain. In Section 3.3 this model is applied to both weekly and daily foreign exchange rates. It is shown that the weekly exchange rates are well represented by a first order Markov switching model, whereas the daily exchange rates are better represented by a high order Markov switching model. The comparison of the forecasting performance of alternative models is conducted in Section 3.4. Section 3.5 summarizes the main results in this chapter.

3.2 A High Order Markov Switching Model

3.2.1 The Model

In this model we assume that a random variable y_t follows an m-component normal mixture distribution, with each component corresponding to a different regime. Thus. if y_t comes from regime i, y_t is a sample from the normal family $N(\mu_i, \sigma_i)$ and the probability density function of y_t is given by:

$$f(y_t) = \sum_{i=1}^{m} f(y_t|s_t = i) \Pr(s_t = i) = \sum_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp(-\frac{(y_t - \mu_i)^2}{2\sigma_i^2}) \Pr(s_t = i), \quad (3.1)$$

where as usual $\Pr(\cdot)$ denotes the probability mass function, and s_t denotes the unobservable regime at time t. For example, for the empirical studies in the rest of this chapter, y_t will be the percentage change of foreign exchange rates, and we will assume that y_t switches between two regimes.

Let $\mathbf{S}_t = (s_{-\infty}, \cdots, s_{t-1}, s_t)'$. In an ordinary mixture model, the unconditional probability of $s_t = i$ is a constant. In addition, the realizations of $\mathbf{S}_{t-1}$ have no effect on s_t, i.e., $\Pr(s_t = i) = \Pr(s_t = i|\mathbf{S}_{t-1})$. In the first order Markov switching model, the transition of s_t is governed by a first order Markov chain such that the unconditional (ergodic) probability of $s_t = i$ is still a constant, but the conditional probability of $s_t = i$ depends on s_{t-1} only, that is, $\Pr(s_t|S_{t-1}) = \Pr(s_t|s_{t-1}) \neq \Pr(s_t)$, unless in the trivial case where $\mathbf{P} = \mathbf{I}$. In this chapter, we will further relax this restriction by assuming that the transition of s_t follows an r-th order Markov chain such that $\Pr(s_t = i|\mathbf{S}_{t-1}) = \Pr(s_t = i|s_{t-1}, \cdots, s_{t-r})$ with $r > 1$. However. the conventional model for the r-th order Markov chain requires $(m-1)m^r$ parameters. thus this approach becomes impractical and intractable when r and m are large. To overcome this problem. Raftery[51] suggested a parsimonious model which requires only $m(m-1) + r - 1$ parameters for an r-th order Markov chain. In this chapter this formulation will be adopted to model the hidden Markov chain in the Markov switching model. More specifically, it is assumed that

$$\Pr(s_t = i|s_{t-1} = j_1, \cdots, s_{t-r} = j_r) = \sum_{h=1}^{r} \lambda_h \mathbf{P}_{j_h i}, \tag{3.2}$$

where

$$\sum_{h=1}^{r} \lambda_h = 1, \text{ and } 0 \leq \lambda_h \leq 1, \text{ for } h = 1, \cdots, r; \tag{3.3}$$

$$\sum_{j=1}^{m} \mathbf{P}_{ij} = 1, \text{ and } 0 \leq \mathbf{P}_{ij} \leq 1, \text{ for } i, j = 1, \cdots, m. \tag{3.4}$$

Thus the conditional probability of $s_t = i$ is a linear combination of contributions from the previous r regimes. When $r = 1$, this model reduces to a first order Markov chain. and $\mathbf{P} = \{\mathbf{P}_{ij}\}$ is the transition matrix of the first order Markov chain. with the ergodic probabilities $\boldsymbol{\pi}$ given by $\mathbf{P}'\boldsymbol{\pi} = \boldsymbol{\pi}$. When $r > 1$, equation (3.2) represents an

r-th order Markov chain, and each additional order of conditional dependence requires only one extra parameter in λ. Furthermore, Raftery[51] showed that

$$\lim_{t\to\infty} \mathbf{P}(s_t = i | s_r = j_r, \cdots, s_1 = j_1) = \pi_i, \text{ for } i = 1, \cdots, m.$$

so the ergodic probabilities of the r-th order Markov chain can be easily computed as well.

3.2.2 Maximum Likelihood Estimation

Let $\mathbf{Y}_t \equiv (y_1, y_2, \cdots, y_t)'$, then the likelihood function of a sample $\mathbf{Y}_T$ can be written as:[1]

$$f(\mathbf{Y}_T; \boldsymbol{\theta}) = \prod_{t=1}^{T} f(y_t | \mathbf{Y}_{t-1}; \boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ denotes all the unknown parameters. The log likelihood function is given by:

$$L = \log f(\mathbf{Y}_T; \boldsymbol{\theta}) = \sum_{t=1}^{T} \log f(y_t | \mathbf{Y}_{t-1}; \boldsymbol{\theta}), \tag{3.5}$$

where

$$f(y_t | \mathbf{Y}_{t-1}; \boldsymbol{\theta}) = \sum_{s_t=1}^{m} \sum_{s_{t-1}=1}^{m} \cdots \sum_{s_{t-r}=1}^{m} f(y_t | s_t; \boldsymbol{\theta}) \Pr(s_t, s_{t-1}, \cdots, s_{t-r} | \mathbf{Y}_{t-1}; \boldsymbol{\theta}). \tag{3.6}$$

Also note that the conditional probability of $s_t = i$ is given by:

$$\Pr(s_t = i | \mathbf{Y}_{t-1}; \boldsymbol{\theta}) = \sum_{s_{t-1}=1}^{m} \sum_{s_{t-2}=1}^{m} \cdots \sum_{s_{t-r}=1}^{m} \Pr(s_t, s_{t-1}, \cdots, s_{t-r} | \mathbf{Y}_{t-1}; \boldsymbol{\theta}). \tag{3.7}$$

In order to maximize the log likelihood function, from equations (3.5) and (3.6) it is obvious that we only need to find out $\Pr(s_t, s_{t-1}, \cdots, s_{t-r} | \mathbf{Y}_{t-1}; \boldsymbol{\theta})$. Note, however, $\Pr(s_t, s_{t-1}, \cdots, s_{t-r} | \mathbf{Y}_{t-1}; \boldsymbol{\theta})$ can be computed iteratively as follows:

$$\begin{aligned}
\Pr(s_{t+1}, s_t, \cdots, s_{t-r+1} | \mathbf{Y}_t; \boldsymbol{\theta}) &= \Pr(s_{t+1} | s_t, \cdots, s_{t-r+1}; \boldsymbol{\theta}) \Pr(s_t, \cdots, s_{t-r+1} | \mathbf{Y}_t; \boldsymbol{\theta}) \\
&= \Pr(s_{t+1} | s_t, \cdots, s_{t-r+1}; \boldsymbol{\theta}) \sum_{s_{t-r}=1}^{m} \Pr(s_t, \cdots, s_{t-r} | \mathbf{Y}_{t-1}; \boldsymbol{\theta}) \frac{f(y_t | s_t; \boldsymbol{\theta})}{f(y_t | \mathbf{Y}_{t-1}; \boldsymbol{\theta})},
\end{aligned}$$

---

[1] In the following equation, we assume that $\mathbf{Y}_0 \equiv (y_{-r+1}, \cdots, y_0)'$ is known.

where $\Pr(s_{t+1}|s_t, \cdots, s_{t-r+1}; \boldsymbol{\theta})$ can be evaluated from equation (3.2). Thus the parameters in $\boldsymbol{\theta}$ can be estimated by maximum likelihood estimation.[2] During the estimation, the filtered probability of $s_t = i$ given information at time t can be easily calculated as follows:

$$\Pr(s_t|\mathbf{Y}_t; \boldsymbol{\theta}) = \sum_{s_{t+1}=1}^{m} \sum_{s_{t-1}=1}^{m} \cdots \sum_{s_{t-r+1}=1}^{m} \Pr(s_{t+1}, s_t, \cdots, s_{t-r+1}|\mathbf{Y}_t; \boldsymbol{\theta}). \tag{3.8}$$

Finally we note that as in other regime switching models, the likelihood ratio statistic or Wald statistic can be employed to test the equality of means or variances across regimes. Furthermore, the order of the underlying Markov process can be determined by SBC, as suggested by Schwarz[54].[3] Generally, SBC is a good model selection criterion in this framework because it is a consistent estimator of Markov chain order, and it is approximately the same as choosing the model with highest posterior probability. For further reference on this issue, see Katz[33].

3.2.3 Forecasting from a High Order Markov Switching Model

Let $\hat{y}_{t+j|t}$ and $\hat{\sigma}^2_{t+j|t}$ denote the j-step ahead forecasts of y_{t+j} and the conditional variance of y_{t+j} at time t, respectively, for $j > 0$. Using the conditional expectations at time t as the forecasts, i.e., $\hat{y}_{t+j|t} = \mathrm{E}_t[y_{t+j}]$ and $\hat{\sigma}^2_{t+j|t} = \mathrm{E}_t[y_{t+j} - \mathrm{E}_t y_{t+j}]^2$, we can easily show that:

$$\hat{y}_{t+j|t} = \sum_{i=1}^{m} \Pr(s_{t+j} = i|\mathbf{Y}_t; \boldsymbol{\theta})\mu_i, \tag{3.9}$$

$$\hat{\sigma}^2_{t+j|t} = \sum_{i=1}^{m} \Pr(s_{t+j} = i|\mathbf{Y}_t; \boldsymbol{\theta})(\sigma_i^2 + \mu_i^2) - \hat{y}^2_{t+j|t}. \tag{3.10}$$

---

[2] To start the iteration on $f(s_t, \cdots, s_{t-r}|\mathbf{Y}_{t-1}; \boldsymbol{\theta})$, we assume that the first r regimes are determined by the ergodic probabilities $\boldsymbol{\pi}$.

[3] In this chapter, SBC is calculated as $-2L + k\log(T)$, where L is the maximized likelihood value and k is the dimension of $\boldsymbol{\theta}$. Thus, the "best" model will yield the smallest SBC.

Therefore, given all the parameter values, we only need $\Pr(s_{t+j}|\mathbf{Y}_t;\boldsymbol{\theta})$ for forecasting. Note, however, when $j = 1$, $\Pr(s_{t+1}|\mathbf{Y}_t;\boldsymbol{\theta})$ is readily given by equation (3.7); when $j > 1$,

$$\Pr(s_{t+j}|\mathbf{Y}_t;\boldsymbol{\theta}) = \sum_{s_{t+j-1}=1}^{m} \sum_{s_{t+j-2}=1}^{m} \cdots \sum_{s_{t+j-r}=1}^{m} \Pr(s_{t+j}, s_{t+j-1}, \cdots, s_{t+j-r}|\mathbf{Y}_t;\boldsymbol{\theta}),$$

where $\Pr(s_{t+j}, s_{t+j-1}, \cdots, s_{t+j-r}|\mathbf{Y}_t;\boldsymbol{\theta})$ can be computed iteratively as follows:

$$\Pr(s_{t+j}, s_{t+j-1}, \cdots, s_{t+j-r}|\mathbf{Y}_t;\boldsymbol{\theta}) =$$
$$\Pr(s_{t+j}|s_{t+j-1}, \cdots, s_{t+j-r};\boldsymbol{\theta}) \sum_{s_{t+j-r-1}=1}^{m} \Pr(s_{t+j-1}, \cdots, s_{t+j-r}, s_{t+j-r-1}|\mathbf{Y}_t;\boldsymbol{\theta}).$$

3.3 *How Do Regimes of Foreign Exchange Rates Switch Over Time?*

Using the specifications given in equations (3.1) and (3.2) with two regimes, in this section the behavior of eight exchange rates will be investigated: Australian dollar (AD), British pound sterling (BP), Canadian dollar (CD), French franc (FF), German mark (GM). Italian lira (IL), Swiss franc (SF) and Japanese yen (JY). The data are downloaded from Datastream International, and y_t will be used to denote the log difference of actual exchange rates multiplied by 100.

3.3.1 Empirical Evidence in Weekly Foreign Exchange Rates

To investigate if the statistical properties of daily exchange rates differ from those of weekly exchange rates, in this subsection let us first consider weekly foreign exchange rates. The duration of y_t is from January 4, 1985 to December 27, 1996, with 626 observations in total. Two alternative models will be estimated: a first order Markov switching model ($r = 1$) and a high order Markov switching model with $r = 2$. However, since the changes in exchange rates are usually very small, we might expect the means for two regimes to be equal. Actually, the estimated means are almost identical for some exchange rates. Thus for each exchange rate, we will adopt a two-step approach. First, we estimate both a model with equal means and a model with different means for the two regimes, and compute the p-value using the likelihood ratio statistic for testing the null hypothesis that two means are equal, then decide if the two means are equal based on the overall magnitude of the p-values. Second, we estimate a model with $r = 1$ and a model with $r = 2$ for each exchange rate, with the specification of the mean terms determined from the first step. During the estimation, the constraints given in (3.3) and (3.4) are imposed by estimating λ_h for $h = 1, \cdots, r-1$, $\mathbf{P}_{ij}$ for $i = 1, \cdots, m$, $j = 1, \cdots, m-1$ and letting $\lambda_r = 1 - \sum_{h=1}^{r-1} \lambda_h$ and $\mathbf{P}_{im} = 1 - \sum_{j=1}^{m-1} \mathbf{P}_{ij}$ for $i = 1, \cdots, m$.

The results from the first step reveal that the means in two regimes are different for Australian dollar, Canadian dollar and Japanese yen, and the means are equal for other currencies. The results from the second step are reported in TABLE 3.1. First, from TABLE 3.1 we can see that the estimates of $\mathbf{P}_{11}$ and $\mathbf{P}_{21}$ are significantly different for all eight exchange rates, which confirms that Markov switching does exist in exchange rates. Second, we can see that for all the exchange rates, the second order Markov switching model does not significantly change the estimates of the mean parameters, variance parameters and transition probabilities. For French franc (FF), German mark (GM) and Italian lira (IL), the estimate of weighting parameter λ_1 is very close to 1, and for other currencies, the estimate of λ_1 has a large standard deviation. Overall, changing the order of the underlying Markov chain from 1 to 2 does not increase the maximized likelihood values significantly for all eight exchange rates. As a result, SBC selects the first order Markov switching model over the second order Markov switching model unanimously. In short, for weekly exchange rates, there is no evidence of high order Markov switching.

3.3.2 Empirical Evidence in Daily Foreign Exchange Rates

In this subsection we investigate the evidence of Markov switching in daily exchange rates. The duration of y_t is from January 1, 1985 to December 31, 1996, with 3131 observations in total. However, we will only use the data up to December 31, 1990, with 1565 observations in total, to estimate the Markov switching models, saving the rest of the data for the study of out-of-sample forecasting performance.

For daily exchange rates, we will follow the same approach taken with respect to weekly exchange rates. First some preliminary estimations suggest that the mean terms in two regimes are different for Australian dollar and Canadian dollar, and equal for other currencies. Then Markov switching models of different orders are estimated for each exchange rate, with the order of the underlying Markov chain from 1 to 5, thus comparing the transition properties of exchange rate regimes over a time period of a week. The resulting maximized likelihood values and values of SBC are reported in TABLE 3.2.

Contrary to what we have found for weekly exchange rates, SBC selects the third order Markov switching model for seven out of the eight exchange rates, with Canadian dollar being the only exception. For Canadian dollar, SBC favors the first order Markov switching. The estimates of the parameters from the third order Markov switching model are reported together with the estimates from the first order Markov switching model in TABLE 3.4.[4]

From TABLE 3.4, we can see that changing the order of the underlying Markov chain from 1 to 3 does not significantly change the estimates of the means and variances. However, some of the transition probabilities are changed significantly. When $r = 3$, either λ_1 or λ_2 is estimated to be 0 for six out of the eight exchange rates, with Canadian and Japanese yen being two exceptions. This is also supportive evidence

---

[4] Since SBC favors the first order Markov switching model for Canadian dollar, only the estimates from the first order Markov switching model are reported in TABLE 3.4.

of high order Markov switching in daily exchange rates.

Finally, note that the estimates of the two variances for most European exchange rates are very close, with the low variance at about 0.20 and the high variance at around 1.0. Especially French franc and German mark have very close estimates for all the parameters.

The evidence found in this section confirms that Markov switching does exist in daily foreign exchange rates data. Moreover, for most western currencies, the Markov switching in the exchange rates appears to be of high order. However, because of the so called "day-of-the-week" effects often found in financial asset returns, one might suspect that the high volatility regime and low volatility regime just correspond to specific days in a week. To investigate this possibility, we can look at the filtered probabilities of high volatility regime for the eight exchange rates, from March 4, 1985 to December 31, 1990, which are presented in Figures 3.1-3.8. Obviously, there is no distinct pattern in the filtered probabilities. However, the transition of regimes appears to be very similar for European exchange rates, with British pound exhibiting occasional deviations from the others.

3.4 *Comparison of Forecasting Performance of Alternative Models*

There are several previous studies which aim at comparing the forecasting performance of alternative models for financial asset returns. Pagan and Schwert[46] used monthly U.S. stock returns data from 1834 to 1925 to compare alternative models of conditional stock volatility, such as GARCH type models, the first order two-state Markov switching model and nonparametric kernel estimators. They found that in terms of in-sample forecasts, nonparametric kernel estimators have the best properties, but in out-of-sample predictions, GARCH type models are better than the Markov switching model and nonparametric estimators. Engel[19] studied eighteen quarterly (and monthly) exchange rates from 1973 to 1991, and found that the first order two-state Markov switching model fit the data well in sample, but could not outperform the random walk model and the forward exchange rates in terms of out-of-sample forecasting. Using five weekly foreign rates from March 1973 to September 1989, West and Cho[61] compared the out-of-sample forecasting performance of the random walk model with mean zero, a GARCH(1,1) model, an IGARCH(1,1) model, two autoregressive models and a nonparametric model for conditional variance, and find that GARCH models tend to have more accurate forecasts for a one-week horizon.

However, there is no systematic study on the comparison of forecasting performance of alternative models for daily exchange rate data. Ideally we would like to compare the forecasts of both returns and conditional variances from different models. But, the daily returns on foreign exchange rates are usually close to zero, and models such as the random walk model, simple GARCH models and Markov switching models with equal means all forecast a constant return. Hence, in this section we concentrate on comparing the forecasts of conditional variance for daily exchange rates, using the random walk model, a GARCH(1,1) model with conditionally t-distributed errors, a first order Markov switching model and the third order Markov switching model selected by SBC in the previous section.

The GARCH(1,1) model we consider is given by:

$$y_t = \mu + u_t, \tag{3.11}$$

where u_t follows a t-distribution conditional on past information, with v degrees of freedom ($v > 2$). The conditional variance of u_t is given by the GARCH equation:

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 h_{t-1}, \tag{3.12}$$

where $\alpha_i > 0$, for $i = 0, 1, 2$, and $\alpha_1 + \alpha_2 < 1$. For the eight daily exchange rates we investigated in the previous section, the estimates of the parameters for this GARCH(1,1) model are reported in TABLE 3.3.

3.4.1 Comparison of Forecasts Using MSPE and MAPE

First we will compare the forecasts from different models using mean squared prediction error (MSPE) and mean absolute prediction error (MAPE) as the criteria. As in Section 3.2.3, let $\hat{\sigma}^2_{t+j|t}$ denote the j-step ahead forecast of σ^2_{t+j} at time t. MSPE is given by:

$$MSPE = \frac{1}{T}\sum_{t=1}^{T}(\sigma^2_{t+j} - \hat{\sigma}^2_{t+j|t})^2.$$

However, the true value of σ^2_{t+j} is unknown. Thus in stead of directly comparing the forecasts of the conditional variance of y_{t+j}, we compare the forecasts of y^2_{t+j}. Again. using the conditional expectations as the forecasts. we can show that for the GARCH(1,1) model,

$$\hat{y}^2_{t+j|t} = \mathrm{E}_t(y^2_{t+j}) = \mathrm{E}_t(h_{t+j}) + \mu^2.$$

and for Markov switching models,

$$\hat{y}^2_{t+j|t} = \mathrm{E}_t(y^2_{t+j}) = \sum_{i=1}^{m}\Pr(s_{t+j} = i|\mathbf{Y}_t;\boldsymbol{\theta})(\sigma^2_i + \mu^2_i).$$

MSPE can thus be computed as

$$MSPE = \frac{1}{T}\sum_{i=1}^{T}(y^2_{t+j} - \hat{y}^2_{t+j|t})^2.$$

Note that the expression of MSPE in the last equation involves the fourth moment of y_t, the existence of which is only guaranteed by the condition $3\alpha_1^2 + 2\alpha_1\alpha_2 + \alpha_2^2 < 1$ for the GARCH(1,1) model.[5] For some exchange rates we investigated, this condition is violated by the estimates in TABLE 3.3. Therefore, following Hamilton and Susmel[27], MAPE is also reported, which can be computed as:

$$MAPE = \frac{1}{T}\sum_{i=1}^{T}|y^2_{t+j} - \hat{y}^2_{t+j|t}|.$$

[5] See Bollerslev[9].

Now let us look at the in-sample comparisons for 1 day ahead, 1 week ahead and 2 weeks ahead forecasts in TABLE 3.5 and TABLE 3.6. When MSPE is used as the criterion, generally Markov switching models give smaller MSPE than the random walk model, but the advantage of Markov switching models disappears for 2 weeks ahead forecasts. However, the third order Markov switching model does not have obvious advantage over the first order Markov switching model except for Swiss franc. MSPE from the GARCH model is slightly larger than that from the Markov switching models in most cases, and sometimes it is even bigger than that from the corresponding random walk model.[6] Similar results are found when MAPE is used as the criterion.

The comparisons of out-of-sample forecasts are reported in TABLE 3.7 and TABLE 3.8. When MSPE is used as the criterion, the GARCH model and Markov switching models yield smaller MSPE than the random walk model, with the GARCH model having slight advantage over the Markov switching models. Again, MSPE from the third order Markov switching model is very close to that from the first order Markov switching model. When MAPE is used as the criterion, the GARCH model and Markov switching models again yield smaller MAPE than the random walk model, but this time the Markov switching model has slight advantage over the GARCH model. Again, the third order Markov switching model does not have obvious advantage over the first order Markov switching model.

3.4.2 Test of Equivalence of Forecasts

In the previous subsection, 1-day ahead, 1-week ahead and 2-week ahead forecasts from the random walk model, the GARCH model and Markov switching models of different orders are compared using MSPE and MAPE as the criteria. Although the Markov switching models have a slight advantage over the GARCH model and the

---

[6] Hamilton and Susmel[27] found similar results for U.S. weekly stock returns.

random walk model, in some cases the resulting MSPE and MAPE are very close for all the models considered. To get a more reliable comparison in this kind of situation, Diebold and Mariano[38] proposed a test of equivalence of forecast accuracy based on some asymptotic results for covariance stationary processes. More specifically, let y_{it} be the forecast of y_t from model i, and let the associated forecast errors be e_{it}, i.e., $e_{it} = y_{it} - y_t$. Also, let $g(\cdot)$ denote the loss function associated with the forecast errors. Different loss functions can be used to assess the forecast accuracy. For instance, MSPE uses an L_2 loss function, i.e., $g(x) = x^2$, whereas MAPE uses an L_1 loss function, i.e., $g(x) = |x|$. Let $d_{ij,t} = g(e_{it}) - g(e_{jt})$ be the loss differential between model i and model j, then the null hypothesis of equal forecast accuracy is equivalent to the null hypothesis that the $\mathrm{E}[d_{ij,t}] = \mu = 0$. However, if the loss differential series is covariance stationary and of short memory, the time average of the loss differential series converges to a normal random variable, i.e.,

$$\sqrt{T}(\bar{d_{ij}} - \mu) \xrightarrow{d} N(0, \Sigma_0).$$

where $\bar{d_{ij}} = \sum_{t=1}^{T} d_{ij,t}/T$, and Σ_0 is the spectral density of the loss differential series at frequency zero. Thus, an immediate test statistic for the equal accuracy hypothesis is the familiar z-statistic,

$$z_{ij} = \bar{d_{ij}}/\sqrt{\Sigma_0/T},$$

which should follow a standard normal distribution if the null hypothesis is true.

To obtain a consistent estimate of Σ_0, Diebold and Mariano[38] suggested using a rectangular lag window with bandwidth $k-1$ for k-step ahead forecasts.[7] Let the random walk model be model 1, the GARCH-t model be model 2, the first order Markov switching model be model 3, and the third order Markov switching model be model 4. The z statistics are reported in TABLES 3.9-3.12.

[7] As a comparison, the Bartlett lag window used in Section 2.4 was also used to estimate Σ_0, which gave similar results.

Since the z statistics only tell us if the forecasts from two models are significantly different, it does not give any information as regard to which one is more accurate. However, we can use the information in MSPE and/or MAPE to gauge the performance of different models when the z statistics are significant.

TABLE 3.9 reports the z statistics for in-sample forecasts when an L_2 loss function is used. Generally we cannot reject that the forecasts from the GARCH(1,1) model is equivalent to those from the random walk model at 5% level, but the forecasts from the Markov switching models are significantly different from those from the random walk model in most cases, with Canadian dollar and Japanese yen as the exceptions. For the GARCH(1,1) model and Markov switching models, the forecasts appear to be equivalent, though for Swiss franc the forecasts from the first order Markov switching model are significantly inferior to those from the GARCH(1,1) model and the third order Markov switching model.

TABLE 3.10 reports the z statistics for out-of-sample forecasts when an L_1 loss function is used. For most exchange rates, the forecasts from the GARCH(1,1) model are comparable with those from the random walk model, though for Japanese yen the forecasts from the GARCH(1,1) model are significant inferior to those from the random walk model. The forecasts from the Markov switching models are consistently better than those from the random walk model. The forecasts from the Markov switching models are different from those from the GARCH(1,1) model in most cases, though the third order Markov switching model is not significantly different from the first order Markov switching model, with Swiss franc being the only exception.

TABLE 3.11 reports the z statistics for out-of-sample forecasts when an L_2 loss function is used. For Australian dollar, British pound, German mark and Swiss franc, the forecasts from the GARCH(1,1) model appear to be better than those from the random walk model, whereas for other currencies the forecasts from the GARCH(1,1) model are not significantly different from those from the random walk model. The Markov switching models consistently generate better forecasts than the random walk model, with Canadian dollar being the only exception. For Australian dollar, British pound, Japanese yen and Swiss franc, the third order Markov switching model appears to generate better forecasts than the first order Markov switching model and the GARCH(1,1) model; for other currencies, the Markov switching models and the GARCH(1,1) model has comparable forecasts.

TABLE 3.12 reports the z statistics for out-of-sample forecasts when an L_1 loss function is used. Again for Australian dollar, British pound and Japanese yen the forecasts from the GARCH(1,1) model are significantly different from those from the random walk model;[8] for other currencies, the GARCH(1,1) model generates comparable forecasts with the random walk model. The Markov switching models again consistently generate better forecasts from the random walk model, with Canadian dollar being the only exception.

3.5 Conclusion

The statistical distributions of financial asset returns, including the foreign exchange rates, are among the most extensively studied topics in financial economics. However, it is usually found that most of the existing models do not perform better than the random walk model. In this chapter, a high order Markov switching model is used to investigate the statistical distributions of high frequency foreign exchange rate data. In general the empirical evidence suggests that the Markov switching models perform better than the random walk model, and also have slight advantage over the GARCH(1,1) model with conditionally t-distributed errors. However, forecasts from the high order Markov switching model are not significantly different from those from the first order Markov switching model, even though SBC prefers the high order Markov switching model for most western exchange rates.

[8] Note that for Japanese yen the forecasts from the GARCH(1,1) model are actually worse.

TABLE 3.1 – Estimates for Weekly Exchange Rates

		μ_1	μ_2	σ_1^2	σ_2^2	λ_1	P_{11}	P_{21}	L (SBC)
AD	r = 1	-0.20	0.72	0.65	5.09	-	0.85	0.53	-994.23
		(0.05)	(0.38)	(0.09)	(0.97)	-	(0.09)	(0.36)	2027.09
	r = 2	-0.21	0.67	0.62	4.87	0.44	0.86	0.42	-992.76
		(0.05)	(0.24)	(0.09)	(0.85)	(0.25)	(0.04)	(0.11)	2030.60
BP	r = 1	0.10	-	1.09	3.46	-	0.98	0.02	-1129.29
		(0.06)	-	(0.14)	(0.52)	-	(0.01)	(0.02)	2290.77
	r = 2	0.10	-	1.08	3.46	0.75	0.97	0.03	-1129.28
		(0.06)	-	(0.15)	(0.56)	(0.73)	(0.02)	(0.03)	2297.19
CD	r = 1	-0.13	0.06	0.07	0.48	-	0.86	0.06	-555.96
		(0.03)	(0.04)	(0.01)	(0.04)	-	(0.04)	(0.03)	1150.56
	r = 2	-0.13	0.06	0.07	0.48	1.00	0.86	0.06	-555.96
		(0.03)	(0.04)	(0.01)	(0.05)	(0.50)	(0.06)	(0.04)	1157.00
FF	r = 1	-0.11	-	1.46	6.39	-	0.97	0.13	-1135.84
		(0.06)	-	(0.16)	(1.36)	-	(0.02)	(0.05)	2303.88
	r = 2	-0.11	-	1.46	6.39	1.00	0.97	0.13	-1135.84
		(0.06)	-	(0.16)	(1.37)	(0.01)	(0.02)	(0.05)	2310.32
GM	r = 1	-0.12	-	1.45	5.01	-	0.95	0.11	-1166.87
		(0.06)	-	(0.22)	(1.00)	-	(0.02)	(0.06)	2365.94
	r = 2	-0.12	-	1.45	5.01	1.00	0.95	0.11	-1166.87
		(0.06)	-	(0.22)	(1.00)	(0.03)	(0.02)	(0.06)	2372.38
IL	r = 1	-0.08	-	1.06	3.97	-	0.96	0.07	-1110.94
		(0.05)	-	(0.12)	(0.58)	-	(0.02)	(0.03)	2254.07
	r = 2	-0.08	-	1.06	3.97	1.00	0.96	0.07	-1110.94
		(0.05)	-	(0.22)	(0.73)	(0.04)	(0.02)	(0.03)	2260.52
JY	r = 1	-0.34	0.21	3.27	0.48	-	0.79	0.29	-1109.67
		(0.13)	(0.07)	(0.36)	(0.15)	-	(0.07)	(0.08)	2257.97
	r = 2	-0.34	0.21	3.27	0.48	1.00	0.79	0.29	-1109.67
		(0.13)	(0.07)	(0.36)	(0.15)	(0.59)	(0.07)	(0.08)	2264.41
SF	r = 1	-0.08	-	2.11	6.57	-	0.95	0.13	-1242.26
		(0.07)	-	(0.29)	(1.53)	-	(0.03)	(0.08)	2516.72
	r = 2	-0.07	-	2.02	6.47	0.40	0.94	0.16	-1242.15
		(0.08)	-	(0.33)	(1.45)	(0.78)	(0.04)	(0.09)	2522.93

Note: The standard errors of the estimates are reported in parentheses. SBCs are reported below the maximized likelihood values.

TABLE 3.2 – Comparison of SBCs for Daily Exchange Rates Models

		r = 1	r = 2	r = 3	r = 4	r = 5
AD	L	-1502.37	-1499.07	-1490.96	-1490.00	-1490.96
	SBC	3048.87	3049.64	3040.77	3048.01	3055.48
BP	L	-1625.10	-1619.75	-1613.64	-1613.64	-1610.75
	SBC	3286.98	3283.64	3278.78	3286.13	3287.71
CD	L	-87.21	-87.21	-85.63	-84.83	-84.83
	SBC	218.55	225.90	230.11	235.87	243.22
FF	L	-1491.97	-1488.82	-1482.09	-1482.09	-1479.10
	SBC	3020.72	3021.78	3015.67	3023.02	3024.40
GM	L	-1555.57	-1549.22	-1542.40	-1542.40	-1534.25
	SBC	3147.92	3142.57	3136.28	3143.64	3143.70
IL	L	-1443.54	-1441.16	-1434.54	-1434.54	-1430.88
	SBC	2923.87	2926.45	2920.56	2927.92	2927.96
JY	L	-1436.31	-1432.13	-1427.67	-1427.67	-1426.19
	SBC	2909.39	2908.39	2906.83	2914.19	2918.58
SF	L	-1772.09	-1756.95	-1750.00	-1748.50	-1746.06
	SBC	3580.95	3558.02	3551.50	3555.84	3558.32

TABLE 3.3 – Estimates from GARCHT model for Daily Exchange Rates

	L (SBC)	μ	α_0	α_1	α_2	ν
AD	-1484.86	-0.06	0.04	0.15	0.81	3.33
	3006.50	(0.01)	(0.02)	(0.04)	(0.05)	(0.30)
BP	-1606.84	0.04	0.02	0.07	0.91	4.74
	3250.45	(0.02)	(0.01)	(0.02)	(0.02)	(0.62)
CD	-64.11	-0.02	0.01	0.16	0.80	4.80
	165.00	(0.01)	(0.00)	(0.03)	(0.04)	(0.56)
FF	-1487.76	-0.03	0.02	0.08	0.87	5.12
	3012.29	(0.01)	(0.01)	(0.02)	(0.03)	(0.69)
GM	-1543.40	-0.03	0.02	0.07	0.90	4.59
	3123.58	(0.02)	(0.01)	(0.02)	(0.02)	(0.56)
IL	-1435.65	-0.03	0.02	0.08	0.87	5.15
	2908.07	(0.01)	(0.01)	(0.02)	(0.03)	(0.69)
JY	-1422.38	-0.01	0.04	0.12	0.83	3.39
	2881.54	(0.01)	(0.01)	(0.03)	(0.04)	(0.32)
SF	-1754.05	-0.02	0.02	0.06	0.91	5.30
	3544.87	(0.02)	(0.01)	(0.01)	(0.02)	(0.74)

Note: The standard errors of the estimates are reported in the parentheses.
SBCs are reported below the maximized likelihood values.

TABLE 3.4 – Estimates from Markov Switching Models for Daily Exchange Rates

		μ_1	μ_2	σ_1^2	σ_2^2	λ_1	λ_2	P_{11}	P_{21}
AD	r = 1	-0.10	0.22	0.15	1.17	-	-	0.89	0.24
		(0.01)	(0.06)	(0.01)	(0.10)	-	-	(0.02)	(0.04)
	r = 3	-0.10	0.23	0.15	1.23	0.50	0.00	0.90	0.22
		(0.01)	(0.06)	(0.01)	(0.11)	(0.09)	(0.05)	(0.02)	(0.04)
BP	r = 1	0.04	-	0.23	0.89	-	-	0.96	0.05
		(0.02)	-	(0.02)	(0.08)	-	-	(0.01)	(0.02)
	r = 3	0.05	-	0.21	0.97	0.00	0.33	0.95	0.08
		(0.02)	-	(0.02)	(0.08)	(0.02)	(0.09)	(0.02)	(0.02)
CD	r = 1	-0.03	0.05	0.03	0.20	-	-	0.95	0.13
		(0.01)	(0.02)	(0.00)	(0.02)	-	-	(0.01)	(0.03)
FF	r = 1	-0.03	-	0.23	0.87	-	-	0.98	0.04
		(0.01)	-	(0.02)	(0.09)	-	-	(0.01)	(0.02)
	r = 3	-0.03	-	0.20	0.94	0.00	0.40	0.96	0.10
		(0.01)	-	(0.02)	(0.09)	(0.03)	(0.09)	(0.01)	(0.03)
GM	r = 1	-0.04	-	0.25	0.86	-	-	0.97	0.05
		(0.02)	-	(0.02)	(0.08)	-	-	(0.01)	(0.02)
	r = 3	-0.04	-	0.20	0.94	0.00	0.48	0.93	0.11
		(0.01)	-	(0.02)	(0.09)	(0.03)	(0.09)	(0.02)	(0.03)
IL	r = 1	-0.04	-	0.22	0.85	-	-	0.98	0.05
		(0.01)	-	(0.02)	(0.08)	-	-	(0.01)	(0.02)
	r = 3	-0.03	-	0.18	0.89	0.00	0.35	0.95	0.10
		(0.01)	-	(0.02)	(0.09)	(0.00)	(0.09)	(0.01)	(0.03)
JY	r = 1	-0.01	-	0.10	0.73	-	-	0.79	0.20
		(0.01)	-	(0.02)	(0.06)	-	-	(0.05)	(0.05)
	r = 3	-0.02	-	0.11	0.78	0.19	0.33	0.85	0.18
		(0.01)	-	(0.02)	(0.07)	(0.18)	(0.17)	(0.03)	(0.04)
SF	r = 1	-0.02	-	0.16	0.92	-	-	0.55	0.34
		(0.02)	-	(0.03)	(0.08)	-	-	(0.08)	(0.08)
	r = 3	-0.02	-	0.26	1.06	0.00	0.33	0.93	0.10
		(0.02)	-	(0.03)	(0.10)	(0.00)	(0.11)	(0.02)	(0.03)

Note: The standard errors of the estimates are reported in parentheses.

TABLE 3.5 – Comparison of In-Sample MSPE

		RW	GARCH	MS-1	MS-3		RW	GARCH	MS-1	MS-3
1-Day	AD	1.12	1.13	1.07	1.07	GM	0.69	0.67	0.67	0.66
1-Week		1.12	1.17	1.11	1.10		0.69	0.68	0.66	0.66
2-Week		1.12	1.21	1.11	1.11		0.69	0.70	0.67	0.68
1-Day	DP	0.62	0.61	0.60	0.60	IL	0.53	0.51	0.51	0.50
1-Week		0.62	0.62	0.60	0.60		0.53	0.51	0.51	0.51
2-Week		0.62	0.64	0.61	0.61		0.53	0.53	0.52	0.52
1-Day	CD	0.04	0.04	0.03	-	JP	0.89	0.90	0.89	0.88
1-Week		0.04	0.04	0.04	-		0.89	0.92	0.89	0.88
2-Week		0.04	0.04	0.04	-		0.89	0.93	0.89	0.89
1-Day	FF	0.63	0.60	0.59	0.59	SF	1.02	1.00	1.02	0.99
1-Week		0.63	0.61	0.60	0.60		1.02	1.01	1.02	0.99
2-Week		0.63	0.63	0.61	0.61		1.02	1.03	1.02	1.01

TABLE 3.6 – Comparison of In-Sample MAPE

		RW	GARCH	MS-1	MS-3		RW	GARCH	MS-1	MS-3
1-Day	AD	0.59	0.56	0.52	0.52	GM	0.51	0.49	0.48	0.48
1-Week		0.59	0.62	0.55	0.54		0.51	0.50	0.48	0.48
2-Week		0.59	0.67	0.55	0.55		0.51	0.52	0.49	0.49
1-Day	BP	0.53	0.50	0.49	0.49	IL	0.45	0.42	0.42	0.42
1-Week		0.53	0.51	0.49	0.49		0.45	0.43	0.42	0.42
2-Week		0.53	0.53	0.50	0.50		0.45	0.44	0.43	0.43
1-Day	CD	0.09	0.09	0.09	-	JY	0.52	0.55	0.51	0.50
1-Week		0.09	0.09	0.09	-		0.52	0.58	0.51	0.50
2-Week		0.09	0.10	0.09	-		0.52	0.60	0.51	0.51
1-Day	FF	0.48	0.46	0.45	0.45	SF	0.64	0.62	0.63	0.60
1-Week		0.48	0.47	0.45	0.45		0.64	0.62	0.63	0.61
2-Week		0.48	0.48	0.46	0.46		0.64	0.64	0.63	0.62

TABLE 3.7 – Comparison of Out-of-Sample MSPE

		RW	GARCH	MS-1	MS-3		RW	GARCH	MS-1	MS-3
1-Day	AD	0.33	0.24	0.26	0.25	GM	1.01	0.97	0.97	0.97
1-Week		0.33	0.27	0.28	0.26		1.01	0.97	0.98	0.97
2-Week		0.33	0.33	0.29	0.28		1.01	0.97	0.98	0.98
1-Day	BP	0.83	0.76	0.77	0.76	IL	3.02	2.83	2.94	2.94
1-Week		0.83	0.77	0.78	0.77		3.02	2.97	2.97	2.97
2-Week		0.83	0.77	0.79	0.78		3.02	2.92	3.00	3.00
1-Day	CD	0.03	0.03	0.03	-	JY	0.98	0.98	0.96	0.96
1-Week		0.03	0.03	0.03	-		0.98	0.98	0.97	0.96
2-Week		0.03	0.03	0.03	-		0.98	0.99	0.98	0.97
1-Day	FF	1.30	1.19	1.25	1.24	SF	1.56	1.52	1.55	1.52
1-Week		1.30	1.21	1.25	1.25		1.56	1.50	1.56	1.52
2-Week		1.30	1.25	1.27	1.27		1.56	1.51	1.56	1.53

TABLE 3.8 – Comparison of Out-of-Sample MAPE

		RW	GARCH	MS-1	MS-3		RW	GARCH	MS-1	MS-3
1-Day	AD	0.48	0.34	0.36	0.35	GM	0.56	0.55	0.53	0.54
1-Week		0.48	0.40	0.42	0.39		0.56	0.56	0.54	0.54
2-Week		0.48	0.48	0.43	0.42		0.56	0.57	0.54	0.54
1-Day	BP	0.57	0.49	0.49	0.49	IL	0.58	0.57	0.55	0.55
1-Week		0.57	0.51	0.50	0.49		0.58	0.58	0.55	0.55
2-Week		0.57	0.52	0.52	0.51		0.58	0.58	0.56	0.56
1-Day	CD	0.09	0.09	0.09	-	JY	0.49	0.52	0.48	0.47
1-Week		0.09	0.09	0.09	-		0.49	0.54	0.48	0.48
2-Week		0.09	0.09	0.09	-		0.49	0.56	0.48	0.48
1-Day	FF	0.56	0.54	0.53	0.53	SF	0.67	0.66	0.66	0.64
1-Week		0.56	0.54	0.53	0.53		0.67	0.66	0.66	0.64
2-Week		0.56	0.55	0.53	0.53		0.67	0.67	0.66	0.65

TABLE 3.9 – In-Sample z-Statistics with $g(x) = x^2$

		z_{23}	z_{24}	z_{34}	z_{12}	z_{13}	z_{14}
AD	1-Day	-1.95	-2.53*	0.16	-0.26	3.48*	4.52*
	1-Week	-1.99	-2.38*	1.03	-1.46	3.55*	2.91*
	2-Week	-3.33*	-3.39*	-0.75	-3.06*	2.22*	1.57
BP	1-Day	-1.38	-3.10*	1.87	0.88	1.9	2.68*
	1-Week	-1.84	-1.93	-0.11	-0.23	3.13*	2.60*
	2-Week	-1.43	-1.45	-0.33	-1.01	2.05*	1.45
CD	1-Day	-2.61*	-	-	-0.62	2.64*	-
	1-Week	-1.91	-	-	-1.14	1.19	-
	2-Week	-0.91	-	-	-0.67	0.7	-
FF	1-Day	-0.72	-1.36	0.74	2.44*	3.32*	3.88*
	1-Week	-1.18	-1.24	-0.20	1.25	3.79*	3.93*
	2-Week	-1.78	-1.65	-1.65	-0.53	2.81*	2.45*
GM	1-Day	-0.75	-1.85	1.55	1.74	2.54*	3.48*
	1-Week	-2.25	-2.14*	-0.64	0.58	4.10*	4.43*
	2-Week	-1.88	-1.82	-0.71	-0.86	2.89*	2.98*
IL	1-Day	0.28	-0.65	1.32	2.33*	2.40*	3.39*
	1-Week	-0.40	-0.59	0.69	1.56	3.20*	3.50*
	2-Week	-1.63	-1.58	-0.40	-0.19	2.77*	2.76*
JY	1-Day	-0.71	-1.30	1.59	-0.66	0.03	1.16
	1-Week	-2.13*	-2.44*	0.72	-2.09*	1.01	0.91
	2-Week	-1.75	-1.78	-0.03	-1.83	-0.50	-0.40
SF	1-Day	1.41	-2.06*	-2.90*	1.66	1.16	3.16*
	1-Week	0.40	-2.13*	-3.09*	0.60	1.84	3.49*
	2-Week	-1.08	-2.06*	1.72	-0.94	1.48	2.05*

Note: The numbers with asterisks are significant at 5% level.

TABLE 3.10 – In-sample z-Statistics with g(x) = |x|

		z_{23}	z_{24}	z_{34}	z_{12}	z_{13}	z_{14}
AD	1-Day	-4.51*	-5.42*	-0.35	2.94*	12.20*	12.75*
	1-Week	-4.71*	-5.79*	1.28	-1.96	16.52*	8.76*
	2-Week	-7.32*	-8.05*	0.23	-4.96*	18.59*	11.25*
BP	1-Day	-2.78*	-3.70*	0.13	5.64*	7.41*	7.74*
	1-Week	-2.51*	-2.88*	0.41	1.66	5.54*	4.76*
	2-Week	-2.39*	-2.60*	0.06	-0.14	5.53*	3.79*
CD	1-Day	-2.41*	-	-	1.11	4.47*	-
	1-Week	-1.77	-	-	-0.84	1.55	-
	2-Week	-2.55*	-	-	-1.96	1.50	-
FF	1-Day	-2.20*	-3.43*	1.09	4.86*	6.41*	7.43*
	1-Week	-3.25*	-3.29*	-0.13	1.87	5.24*	5.18*
	2-Week	-2.79*	-2.85*	0.27	0.17	4.16*	3.95*
GM	1-Day	-3.71*	-4.64*	0.42	3.87*	6.49*	7.60*
	1-Week	-4.18*	-3.73*	-1.84	0.97	5.70*	5.96*
	2-Week	-3.46*	-3.08*	-2.86*	-0.64	4.60*	4.53*
IL	1-Day	0.16	-1.10	1.63	5.33*	4.98*	6.50*
	1-Week	-1.62	-2.28*	1.87	2.33*	4.05*	4.63*
	2-Week	-1.96	-2.29*	-2.05*	0.62	3.82*	3.84*
JY	1-Day	-8.17*	-9.91*	0.71	-6.23*	3.29*	3.74*
	1-Week	-8.03*	-9.31*	0.54	-6.91*	12.42*	3.63*
	2-Week	-8.08*	-8.56*	0.53	-7.43*	15.55*	5.88*
SF	1-Day	2.02*	-3.89*	4.95*	4.41*	11.52*	7.27*
	1-Week	0.42	-3.23*	3.58*	1.71	20.11*	5.28*
	2-Week	-1.04	-2.95*	2.20*	-0.02	16.86*	4.03*

Note: The numbers with asterisks are significant at 5% level.

TABLE 3.11 – Out-of-Sample z-Statistics with $g(x) = x^2$

		z_{23}	z_{24}	z_{34}	z_{12}	z_{13}	z_{14}
AD	1-Day	3.11*	1.95	3.61*	15.68*	13.35*	15.89*
	1-Week	1.89	-2.04*	8.58*	8.97*	20.08*	15.50*
	2-Week	-5.64*	-7.93*	7.63*	-0.33	20.28*	16.41*
BP	1-Day	0.89	0.47	1.73	4.03*	6.57*	6.89*
	1-Week	1.05	0.26	3.37*	4.07*	6.83*	7.02*
	2-Week	3.31*	1.70	4.65*	6.34*	7.06*	6.95*
CD	1-Day	-0.36	-	-	0.11	0.87	-
	1-Week	0.56	-	-	0.77	1.65	-
	2-Week	-0.89	-	-	-0.65	2.09*	-
FF	1-Day	1.22	1.16	0.98	1.86	3.41*	3.71*
	1-Week	1.42	1.29	1.98	2.12*	2.55*	2.89*
	2-Week	0.82	0.82	0.24	1.71	2.95*	3.11*
GM	1-Day	0.02	0.21	-0.56	2.28*	3.61*	3.71*
	1-Week	1.00	0.65	1.45	2.54*	3.05*	3.79*
	2-Week	0.73	0.81	-0.51	2.43*	3.96*	4.24*
IL	1-Day	0.76	0.82	-0.65	1.23	3.06*	3.37*
	1-Week	-0.06	0.03	-0.59	0.75	2.16*	2.76*
	2-Week	1.32	1.31	0.25	1.55	2.68*	2.57*
JY	1-Day	-0.82	-1.40	1.84	-0.10	2.08*	3.27*
	1-Week	-0.28	-0.86	2.94*	-0.14	2.68*	3.50*
	2-Week	-0.73	-1.05	2.84*	-0.72	0.67	3.09*
SF	1-Day	1.33	-0.09	2.52*	1.70	3.31*	3.06*
	1-Week	2.18*	1.29	2.66*	2.30*	0.46	2.85*
	2-Week	2.10*	1.46	2.78*	2.23*	0.39	3.00*

Note: The numbers with asterisks are significant at 5% level.

TABLE 3.12 – Out-of-Sample z-Statistics with g(x) = |x|

		z_{23}	z_{24}	z_{34}	z_{12}	z_{13}	z_{14}
AD	1-Day	8.14*	6.08*	7.04*	33.05*	28.17*	32.07*
	1-Week	3.91*	-1.99	12.06*	12.27*	32.49*	22.23*
	2-Week	-6.34*	-10.34*	9.14*	0.75	35.59*	23.67*
BP	1-Day	-0.66	-1.15	0.74	11.96*	16.70*	17.05*
	1-Week	-0.22	-2.21*	5.04*	5.62*	9.83*	9.29*
	2-Week	0.47	-1.38	4.29*	3.93*	7.94*	7.00*
CD	1-Day	1.27	-	-	1.65	0.83	-
	1-Week	-0.07	-	-	0.73	2.30	-
	2-Week	-1.31	-	-	-0.79	2.74	-
FF	1-Day	-0.77	-1.36	1.87	2.42*	4.87*	5.91*
	1-Week	-1.60	-2.29*	2.61*	1.54	3.67*	4.18*
	2-Week	-2.54*	-3.13*	2.13*	0.76	3.72*	3.87*
GM	1-Day	-5.27*	-4.48*	-2.83*	1.23	5.92*	5.45*
	1-Week	-4.54*	-4.12*	-1.29	0.14	4.10*	4.00*
	2-Week	-4.93*	-4.32*	-3.08*	-0.73	4.31*	3.90*
IL	1-Day	-1.67	-2.14*	2.24*	0.86	5.10*	6.65*
	1-Week	-1.93	-2.24*	3.04*	-0.17	3.93*	4.57*
	2-Week	-1.66	-1.90	2.20*	-0.34	3.31*	3.39*
JY	1-Day	-7.38*	-9.22*	1.59	-4.31*	6.31*	7.10*
	1-Week	-6.11*	-7.66*	3.01*	-4.87*	16.02*	6.50*
	2-Week	-9.26*	-10.72*	3.64*	-8.12*	17.67*	9.34*
SF	1-Day	-0.12	-4.45*	4.03*	2.08*	13.98*	6.78*
	1-Week	0.52	-2.90*	3.25*	1.72	20.22*	4.94*
	2-Week	-0.43	-3.49*	2.76*	0.77	17.77*	4.60*

Note: The numbers with asterisks are significant at 5% level.

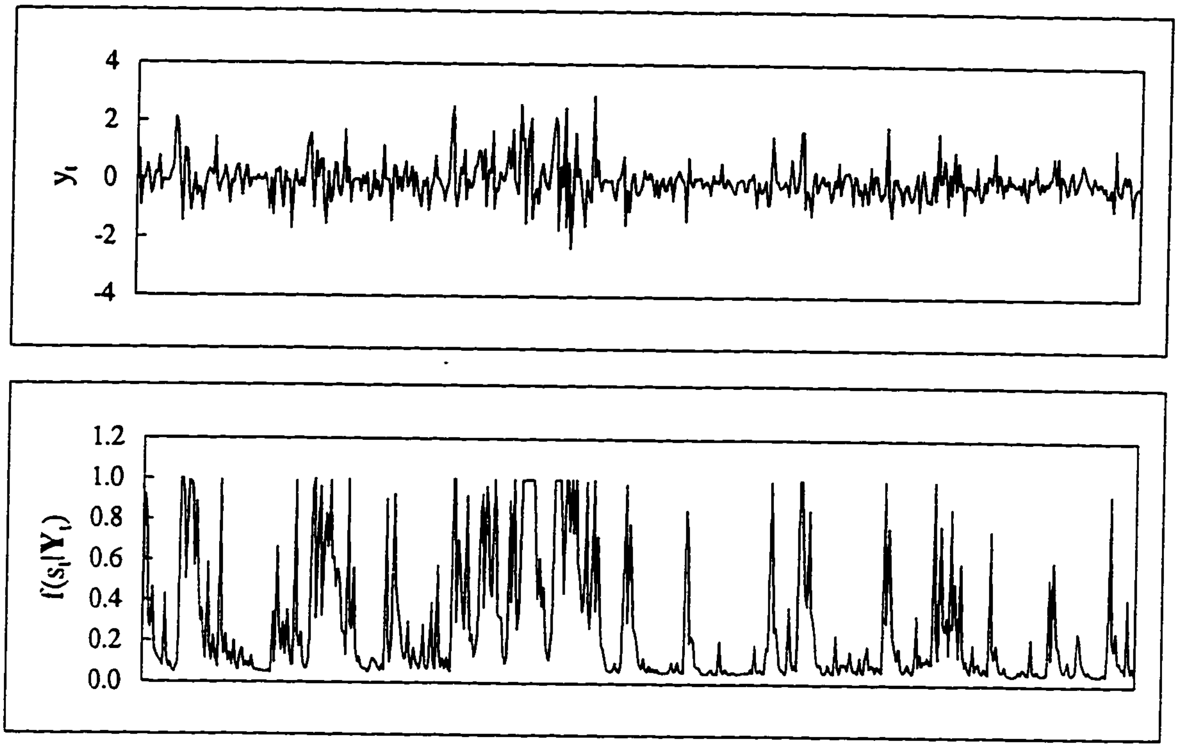

Figure 3.1 – Filtered Probabilities of High Volatility Regime (AD)

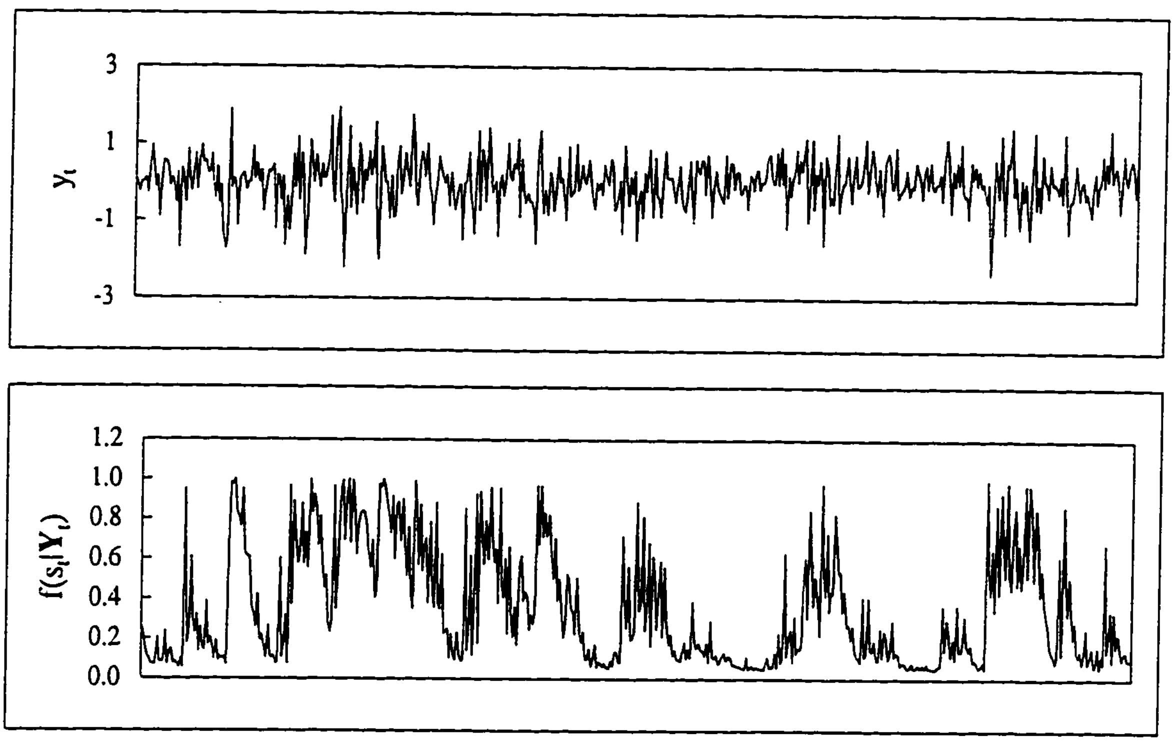

Figure 3.2 – Filtered Probabilities of High Volatility Regime (BP)

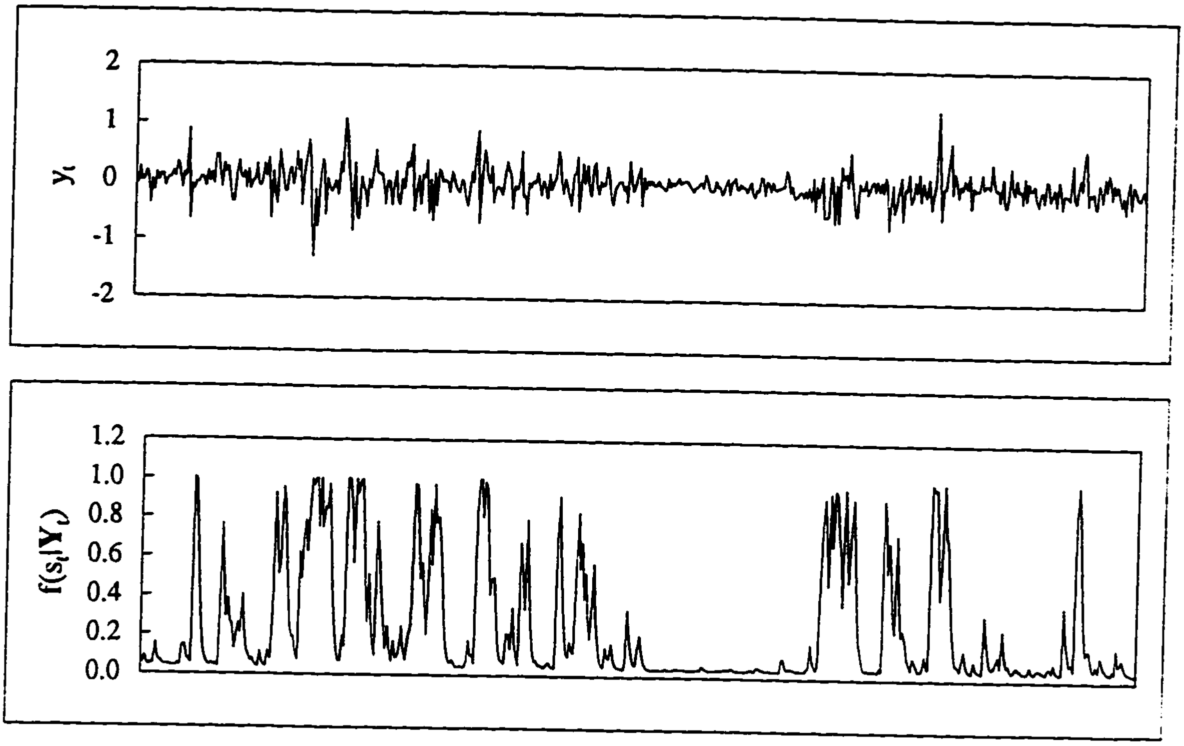

Figure 3.3 – Filtered Probabilities of High Volatility Regime (CD)

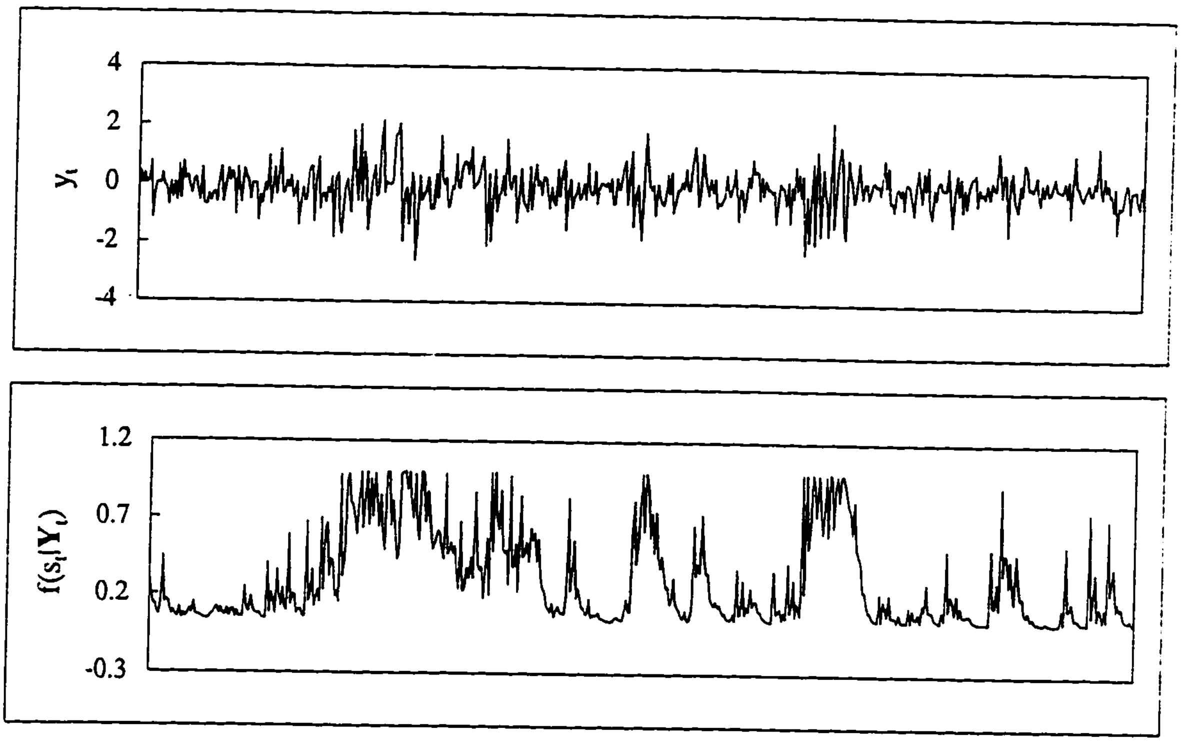

Figure 3.4 – Filtered Probabilities of High Volatility Regime (FF)

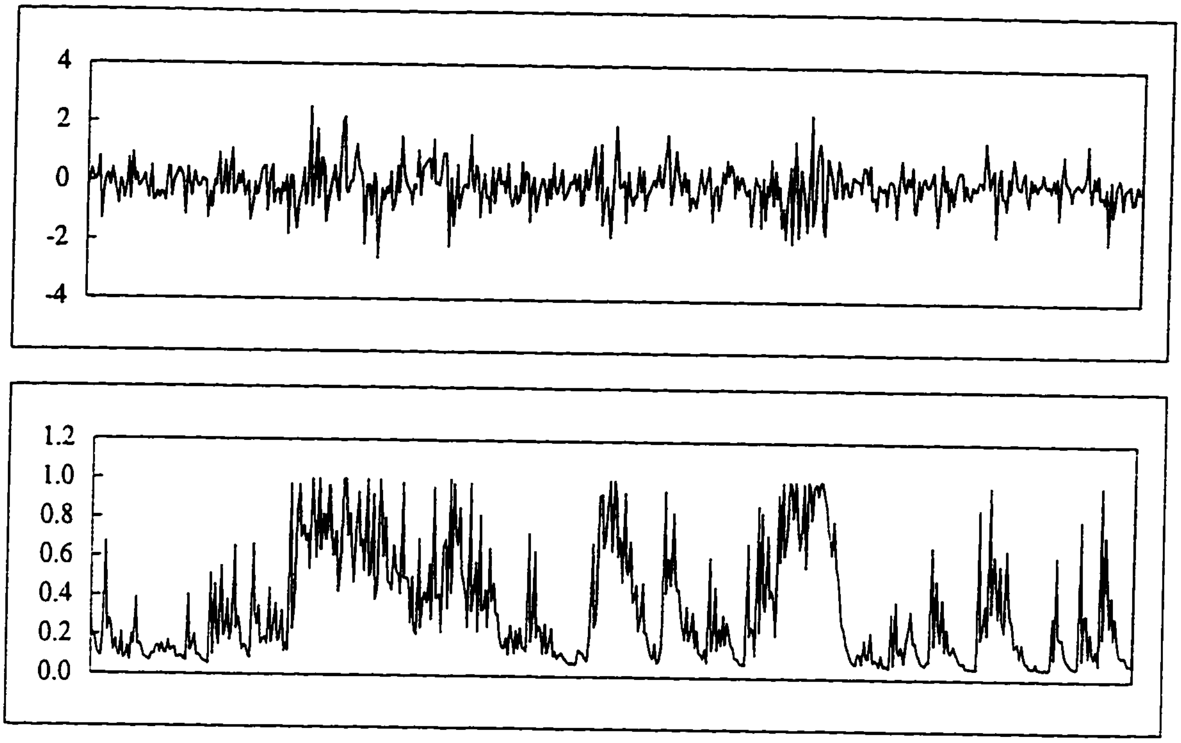

Figure 3.5 – Filtered Probabilities of High Volatility Regime (GM)

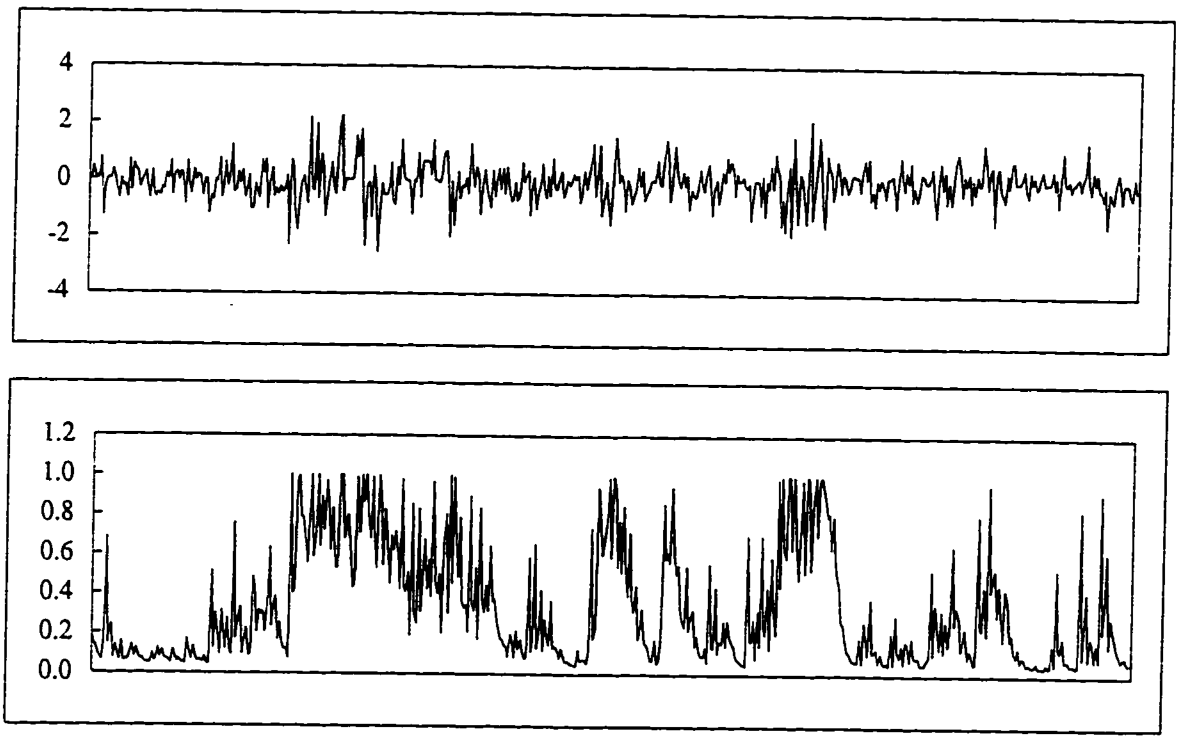

Figure 3.6 – Filtered Probabilities of High Volatility Regime (IL)

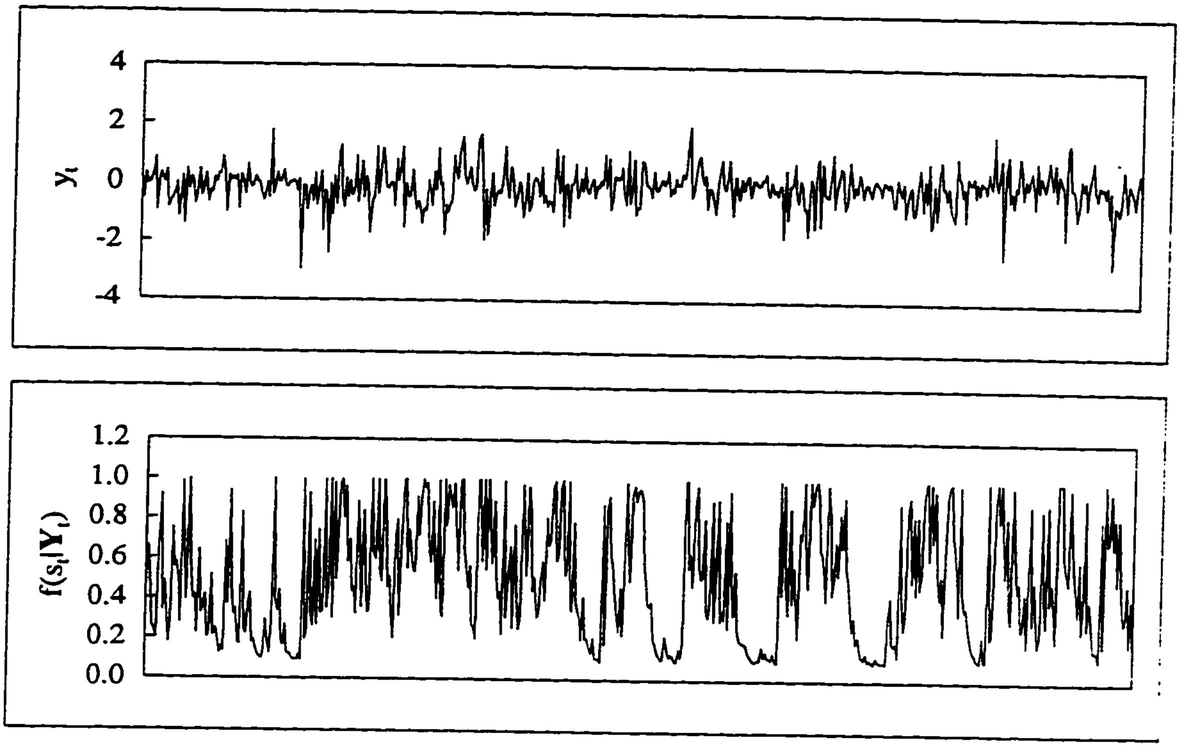

Figure 3.7 – Filtered Probabilities of High Volatility Regime (JY)

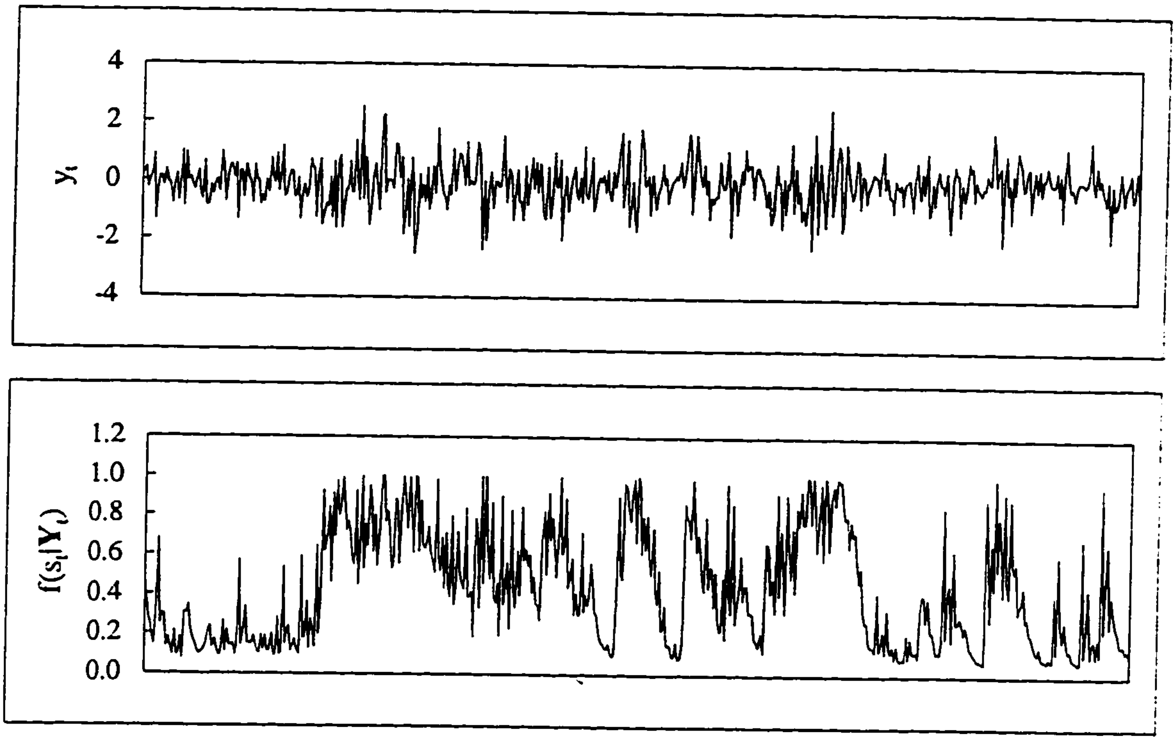

Figure 3.8 – Filtered Probabilities of High Volatility Regime (SF)

BIBLIOGRAPHY

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, Inc., 1984.

[2] T. W. Anderson and H. Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20:46–63, 1949.

[3] Donald W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858, 1991.

[4] Donald W. K. Andrews. Tests for parameter instability and structural changes with unknown change point. *Econometrica*, 61(4):821–856, 1993.

[5] Jushan Bai. Estimating multiple breaks one at a time. Working Paper, Department of Economics, Massachusetts Institute of Technology, 1995.

[6] Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. Working Paper, Department of Economics, Massachusetts Institute of Technology, 1995.

[7] Richard T. Baillie and Tim Bollerslev. The message in daily exchange rates: A conditional-variance tale. *Journal of Business and Economics Statistics*, 7(3):297–305, 1989.

[8] Anindya Banerjee, Robin L. Lumsdaine, and James H. Stock. Recursive and sequential tests of the unit-root and trend-break hypotheses: Theory and inter-

national evidence. *Journal of Business and Economic Statistics*, 10(3):271–287. 1992.

[9] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[10] Tim Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69:542–547, 1987.

[11] John Bound, David A. Jaeger, and Regina M. Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association*, 90(430):443–459, 1995.

[12] George E. P. Box and G. C. Tiao. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70:70–79, 1975.

[13] Lyle D. Broemeling and Hiroki Tsurumi. *Econometrics and Structural Change*. Marcel Dekker, Inc., 1987.

[14] Bradley P. Carlin, Alan E. Gelfand, and Adrian F. M. Smith. Hierarchical Bayesian analysis of changepoint problems. *Applied Statistician*, 41(2):389–405. 1992.

[15] Siddhartha Chib and Edward Greenberg. Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory*, 12:409–431, 1996.

[16] Russell Davidson and James G. MacKinnon. *Estimation and Inference in Econometrics*. Oxford University Press, 1993.

[17] Jean-Marie Dufour. Some impossibility theorems in econometrics with applications to instrumental variables, dynamic models, and cointegration. Forthcoming in *Econometrica*, 1996.

[18] Jean-Marie Dufour and Jan F. Kiviet. Exact tests for structural changes in first-order dynamic models. *Journal of Econometrics*, 70(1):39–69, 1996.

[19] Charles Engel. Can the Markov switching model forecast exchange rates? *Journal of International Economics*, 36:151–165, 1994.

[20] Charles Engel and Craig S. Hakkio. The distribution of exchange rates in the EMS. International Journal of Finance and Economics. 1996.

[21] Charles Engel and James Hamilton. Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review*, 80(4):689–713. 1990.

[22] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987–1008, 1982.

[23] A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

[24] S. German and D. German. Stochastic relaxation. Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern and Machine Intelligence*, 6:721–741. 1984.

[25] Charles J. Geyer. Practical Markov chain Monte Carlo. *Statistical Science*, 7(4):473–483, 1992.

[26] Alastair R. Hall, Glenn D. Rudebusch, and David W. Wilcox. Judging instrument revelance in instrumental variables estimation. *International Economic Review*. 37(2):283–298, 1996.

[27] James D. Hamilton and Raul Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64:307–333. 1994.

[28] David A. Hsieh. The statistical properties of daily foreign exchange rates: 1974–1983. *Journal of International Economics*, 24:129–145, 1988.

[29] David A Hsieh. Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business and Economic Statistics*, 7(3):307–317, 1989.

[30] George G. Judge, W. E. Griffiths, R. Carter Hill, Helmut Lütkepohl, and Tsoung-Chao Lee. *The Theory And Practice of Econometrics*. John Wiley and Sons. 1985.

[31] Graciela Kaminsky. Is there a Peso problem? Evidence from the dollar/pound exchange rate, 1976–1987. *American Economic Review*, 83(3):450–472, 1993.

[32] Robert E. Kass and Adrian E. Raftery. Bayes factor. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

[33] R. W. Katz. On some criteria for estimating the order of a Markov chain. *Technometrics*, 23:243–249, 1981.

[34] J. Liu, W. Wong, and A. Kong. Correlation structure and the convergence of the Gibbs sampler, I. Discussion Paper 299, University of Chicago, Department of Statistics, 1991.

[35] J. Liu, W. Wong, and A. Kong. Correlation structure and the convergence of the Gibbs sampler, II: Applications to various scans. Discussion Paper 304, University of Chicago, Department of Statistics, 1991.

[36] G. S. Maddala and J. Jeong. On the exact small sample distribution of the IV estimator. *Econometrica*, 60:181–183, 1992.

[37] Angus Maddison. *Monitoring the World Economy: 1820–1992*. Development Centre of OECD, 1995.

[38] Francis X. Mariano and Roberto S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263, 1995.

[39] Charles R. Nelson and Richard Startz. The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *Journal of Business*, 63:S125–S140, 1990.

[40] Charles R. Nelson and Richard Startz. Some further results on the exact small sample properties of the instrumental variables estimator. *Econometrica*, 58:967–976, 1990.

[41] Charles R. Nelson, Richard Startz, and Eric Zivot. Valid confidence intervals and inference in the presence of weak instruments. Working Paper, Department of Economics, University of Washington, 1996.

[42] Whitney K. Newey and Kenneth D. West. Hypothesis testing with efficient method of moments estimators. *International Economic Review*, 28:777–787, 1987.

[43] Whitney K. Newey and Kenneth D. West. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61:631–653, 1994.

[44] M. A. Newton and Adrian E. Raftery. Approximate Bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society*, 56:3–48, 1994.

[45] John M. H. Olmsted. *Solid Analytic Geometry*. New York, D. Appleton-Century Company, Inc., 1947.

[46] Adrian R. Pagan and G. William Schwert. Alternative models for conditional stock volatility. *Journal of Econometrics*, 45:267–290, 1990.

[47] Pierre Perron. The great crash, the oil price shock and the unit root hypothesis. *Econometrica*, 57:1361–1401, 1989.

[48] Pierre Perron and Timothy J. Vogelsang. Nonstationary and level shifts with an application to purchasing power parity. *Journal of Business and Economic Statistics*, 10(3):301–320, 1992.

[49] David B. Phillips and Adrian F. M. Smith. Bayesian model comparison via jump diffusion. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo*, chapter 13. Chapman & Hall, 1996.

[50] Peter C. B. Phillips. Partially identified econometric models. *Eonometric Theory*, 5:181–240, 1989.

[51] Adrian E. Raftery. A model for high-order Markov chains. *Journal of the Royal Statistical Society, Series B*, 47(3):528–539, 1985.

[52] Adrian E. Raftery. Change point and change curve modeling in stochastic processes and spatial statistics. *Journal of Applied Statistical Science*, 1(4), 1994.

[53] C. R. Rao. *Linear Statistical Inference and Its Applications*. John Wiley & Sons, Inc., 1973.

[54] Gideon Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

[55] G. William Schwert. Why does stock market volatility change over time? *Journal of Finance*, 44(5):1115–1153, 1989.

[56] Douglas Staiger and James H. Stock. Instrumental variables regression with weak instruments. forthcoming in *Econometrica*, 1996.

[57] Martin A. Tanner. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions.* Springer-Verlag, 1993.

[58] Martin A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–550, 1987.

[59] Ruey S. Tsay. Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, 7:1–20, 1988.

[60] George C. G. Wei and Martin A. Tanner. Calculating the content and boundary of the highest posterior density region via data augmentation. *Biometrika*, 77(3):649–652, 1990.

[61] Kenneth D. West and Dongchul Cho. The predictive ability of several models of exchange rate volatility. *Journal of Econometrics*, 69:367–391, 1995.

[62] Arnold Zellner. *An Introduction to Bayesian Inference in Econometrics.* John Wiley & Sons, Inc., 1971.

[63] Eric Zivot and Donald W. K. Andrews. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics*, 10(3):251–270, 1992.

[64] Eric Zivot and Peter C. B. Phillips. A Bayesian analysis of trend determination. *Econometric Reviews*, 13(3):291–336, 1994.

Appendix A

PROOF OF THEOREM 1

(a) Consider the LM_{GMM} expression for $\mathcal{GMM}_0$ as given in Section 1.3:

$$\mathcal{GMM}_0 = (\mathbf{u}^{\perp\prime}\mathbf{u}^{\perp}/T)^{-1}\mathbf{u}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{Y}^{\perp}(\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{Y}^{\perp})^{-1}\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{u}^{\perp}$$
$$\xrightarrow{d} \mathbf{Z}_u'(\boldsymbol{\lambda} + \mathbf{Z}_V)[(\boldsymbol{\lambda} + \mathbf{Z}_V)'(\boldsymbol{\lambda} + \mathbf{Z}_V)]^{-1}(\boldsymbol{\lambda} + \mathbf{Z}_V)'\mathbf{Z}_u,$$

where the convergence follows from parts (a)-(c) of Lemma 1.

(b) After some algebraic rearrangement, the likelihood ratio statistic in (1.7) may be rewritten as:

$$LR_{LIML} = -T\ln(1 - \mathbf{u}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{u}^{\perp}/\mathbf{u}^{\perp\prime}\mathbf{u}^{\perp}) - T\ln(1 + (\hat{k}_{LIML} - 1))$$
$$= T\mathbf{u}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{u}^{\perp}/\mathbf{u}^{\perp\prime}\mathbf{u}^{\perp} - T(\hat{k}_{LIML} - 1) + o_p(1),$$

The result follows from Lemma 1 part (d) and SS's Theorem 2.

(c) When $k_2 = n$, $\boldsymbol{\lambda} + \mathbf{Z}_V$ is a nonsingular square matrix and $\hat{k}_{LIML} = 1$ so the limiting expressions for $\mathcal{GMM}_0$ and LR_{LIML} reduce to $\mathbf{Z}_u'\mathbf{Z}_u \sim \chi^2(k_2)$. When $k_2 > n$, $(\boldsymbol{\lambda} + \mathbf{Z}_V)[(\boldsymbol{\lambda} + \mathbf{Z}_V)'(\boldsymbol{\lambda} + \mathbf{Z}_V)]^{-1}(\boldsymbol{\lambda} + \mathbf{Z}_V)' = \mathbf{P}_{\lambda + Z_V}$, which is the projection matrix onto the space spanned by $\boldsymbol{\lambda} + \mathbf{Z}_V$, so $\mathbf{Z}_u'\mathbf{P}_{\lambda+Z_V}\mathbf{Z}_u < \mathbf{Z}_u'\mathbf{Z}_u \sim \chi^2(k_2)$. In addition, ξ^* is positive since, from the proof of SS's Theorem 2, it is the vector of eigenvalues from the positive definite matrix $\Upsilon'\Xi_0^*\Upsilon$ where $\Upsilon = \mathrm{diag}(\sigma_{uu}^{\frac{1}{2}}, \Sigma_{VV}^{\frac{1}{2}})$ and Ξ_0^* is defined in the statement of the theorem. It follows trivially that $\mathbf{Z}_u'\mathbf{Z}_u - \xi^* < \mathbf{Z}_u'\mathbf{Z}_u \sim \chi^2(k_2)$.

Appendix B

PROOF OF PROPOSITION 1

(a) Since the AR statistic follows an F distribution under the assumption of normality, the $(1-\alpha)\cdot 100\%$ confidence interval for β is given by:

$$\frac{(\mathbf{y}^{\perp}-\mathbf{Y}^{\perp}\beta)'\mathbf{P}_{Z^{\perp}}(\mathbf{y}^{\perp}-\mathbf{Y}^{\perp}\beta)}{(\mathbf{y}^{\perp}-\mathbf{Y}^{\perp}\beta)'\mathbf{M}_{Z^{\perp}}(\mathbf{y}^{\perp}-\mathbf{Y}^{\perp}\beta)}\frac{T-k}{k_2} \leq F_{1-\alpha}(k_2, T-k),$$

where $F_{1-\alpha}(k_2, T-k)$ is the $(1-\alpha)$ quantile of F distribution with k_2 and $T-k$ degrees of freedom. From the last inequality it follows that

$$\frac{(\mathbf{y}^{\perp}-\mathbf{Y}^{\perp}\beta)'(\mathbf{y}^{\perp}-\mathbf{Y}^{\perp}\beta)}{(\mathbf{y}^{\perp}-\mathbf{Y}^{\perp}\beta)'\mathbf{M}_{Z^{\perp}}(\mathbf{y}^{\perp}-\mathbf{Y}^{\perp}\beta)} \leq C_{AR},$$

where $C_{AR} = 1 + F_{1-\alpha}(k_2, T-k)\frac{k_2}{T-k}$. This can be rewritten more compactly as:

$$\begin{pmatrix} 1 & -\beta \end{pmatrix} Q_{AR} \begin{pmatrix} 1 \\ -\beta \end{pmatrix} \leq 0,$$

which is a quadratic form and

$$Q_{AR} = \begin{pmatrix} \mathbf{y}^{\perp\prime}\mathbf{y}^{\perp} & \mathbf{y}^{\perp\prime}\mathbf{Y}^{\perp} \\ \mathbf{Y}^{\perp\prime}\mathbf{y}^{\perp} & \mathbf{Y}^{\perp\prime}\mathbf{Y}^{\perp} \end{pmatrix} - C_{AR}\begin{pmatrix} \mathbf{y}^{\perp\prime}\mathbf{M}_{Z^{\perp}}\mathbf{y}^{\perp} & \mathbf{y}^{\perp\prime}\mathbf{M}_{Z^{\perp}}\mathbf{Y}^{\perp} \\ \mathbf{Y}^{\perp\prime}\mathbf{M}_{Z^{\perp}}\mathbf{y}^{\perp} & \mathbf{Y}^{\perp\prime}\mathbf{M}_{Z^{\perp}}\mathbf{Y}^{\perp} \end{pmatrix}.$$

Therefore the $(1-\alpha)\cdot 100\%$ confidence interval for β will be unbounded when the $(2,2)$ element of matrix Q_{AR} is negative, i.e., when

$$\frac{\mathbf{Y}^{\perp\prime}\mathbf{Y}^{\perp}}{\mathbf{Y}^{\perp\prime}\mathbf{M}_{Z^{\perp}}\mathbf{Y}^{\perp}} < C_{AR}.$$

The above condition can be rewritten as

$$\frac{\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^{\perp}}\mathbf{Y}^{\perp}}{\mathbf{Y}^{\perp\prime}\mathbf{M}_{Z^{\perp}}\mathbf{Y}^{\perp}}\frac{T-k}{k_2} < F_{1-\alpha}(k_2, T-k).$$

This condition is equivalent to the statement that the first stage F statistic is insignificant at level α, with the first stage regression given in equation (1.4).

(b) When $\mathbf{\Pi} = 0$,

$$\Pr\{\frac{\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{Y}^\perp}{\mathbf{Y}^{\perp\prime}\mathbf{M}_{Z^\perp}\mathbf{Y}^\perp}\frac{T-k}{k_2} < F_{1-\alpha}(k_2, T-k)\} = 1-\alpha.$$

This implies that the confidence interval for β is unbounded with probability $(1-\alpha)$ at significance level α, using the result in (a).

(c) Using the expression for LR_{LIML} we know that the $(1-\alpha)\cdot 100\%$ (conservative) confidence interval for β is the set of values of β that satisfies the following inequality:

$$T\ln\frac{(\mathbf{y}^\perp - \mathbf{Y}^\perp\beta)'(\mathbf{y}^\perp - \mathbf{Y}^\perp\beta)}{(\mathbf{y}^\perp - \mathbf{Y}^\perp\beta)'\mathbf{M}_{Z^\perp}(\mathbf{y}^\perp - \mathbf{Y}^\perp\beta)} - T\ln\hat{k} \leq \chi^2_{1-\alpha}(k_2),$$

where $\chi^2_{1-\alpha}(K_2)$ is the $(1-\alpha)$ quantile of a $\chi^2(k_2)$ distribution. The last inequality can be rewritten as

$$\frac{(\mathbf{y}^\perp - \mathbf{Y}^\perp\beta)'(\mathbf{y}^\perp - \mathbf{Y}^\perp\beta)}{(\mathbf{y}^\perp - \mathbf{Y}^\perp\beta)'\mathbf{M}_{Z^\perp}(\mathbf{y}^\perp - \mathbf{Y}^\perp\beta)} \leq C_{LR},$$

where $C_{LR} = \hat{k}\exp\{\chi^2_{1-\alpha}(k_2)/T\}$. We can further rewrite the inequality in a quadratic form:

$$\begin{pmatrix} 1 & -\beta \end{pmatrix} Q_{LR} \begin{pmatrix} 1 \\ -\beta \end{pmatrix} \leq 0,$$

where

$$Q_{LR} = \begin{pmatrix} \mathbf{y}^{\perp\prime}\mathbf{y}^\perp & \mathbf{y}^{\perp\prime}\mathbf{Y}^\perp \\ \mathbf{Y}^{\perp\prime}\mathbf{y}^\perp & \mathbf{Y}^{\perp\prime}\mathbf{Y}^\perp \end{pmatrix} - C_{LR}\begin{pmatrix} \mathbf{y}^{\perp\prime}\mathbf{M}_{Z^\perp}\mathbf{y}^\perp & \mathbf{y}^{\perp\prime}\mathbf{M}_{Z^\perp}\mathbf{Y}^\perp \\ \mathbf{Y}^{\perp\prime}\mathbf{M}_{Z^\perp}\mathbf{y}^\perp & \mathbf{Y}^{\perp\prime}\mathbf{M}_{Z^\perp}\mathbf{Y}^\perp \end{pmatrix}.$$

Therefore the confidence interval obtained by solving this inequality will be unbounded if the $(2,2)$ element of Q_{LR} is negative. i.e., if

$$\frac{\mathbf{Y}^{\perp\prime}\mathbf{Y}^\perp}{\mathbf{Y}^{\perp\prime}\mathbf{M}_{Z^\perp}\mathbf{Y}^\perp} < C_{LR}.$$

This inequality can be rewritten as

$$\frac{\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{Y}^\perp}{\mathbf{Y}^{\perp\prime}\mathbf{M}_{Z^\perp}\mathbf{Y}^\perp}\frac{T-k_1-k_2}{k_2} < (\hat{k}\exp\{\chi^2_{1-\alpha}(k_2)/T\}-1)\frac{T-k_1-k_2}{k_2}.$$

Note that $\hat{k}$ is approximately equal to 1, and $\exp\{\chi^2_{1-\alpha}(k_2)/T\} \approx 1+\chi^2_{1-\alpha}(k_2)/T$, so the right-hand-side of the inequality is approximately equal to $\chi^2_{1-\alpha}(k_2)/k_2$. From part (a) of this proof, the left-hand-side of the inequality is the first stage F statistic. We know that asymptotically the F statistic converges to $\chi^2_{1-\alpha}(k_2)/k_2$, thus, the $(1-\alpha)\cdot 100\%$ confidence interval obtained by inverting LR_{LIML} will be unbounded when the first stage F statistic is insignificant at level α, using the asymptotic distribution.

(d) The $(1-\alpha)\cdot 100\%$ (conservative) confidence interval for β can be obtained by inverting $\mathcal{GMM}_0$, i.e., solving the following inequality:

$$\frac{(\mathbf{y}^\perp-\mathbf{Y}^\perp\beta)'\mathbf{P}_{Z^\perp}(\mathbf{y}^\perp-\mathbf{Y}^\perp\beta)}{(\mathbf{y}^\perp-\mathbf{Y}^\perp\beta)'(\mathbf{y}^\perp-\mathbf{Y}^\perp\beta)/T} \le \chi^2_{1-\alpha}(k_2),$$

which can be rewritten compactly in the quadratic form:

$$\begin{pmatrix}1 & -\beta\end{pmatrix} Q_{LM}\begin{pmatrix}1\\ -\beta\end{pmatrix} \le 0,$$

where

$$Q_{LM}=\begin{pmatrix}\mathbf{y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{y}^\perp & \mathbf{y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{Y}^\perp\\ \mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{y}^\perp & \mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{Y}^\perp\end{pmatrix} - C_{LM}\begin{pmatrix}\mathbf{y}^{\perp\prime}\mathbf{y}^\perp & \mathbf{y}^{\perp\prime}\mathbf{Y}^\perp\\ \mathbf{Y}^{\perp\prime}\mathbf{y}^\perp & \mathbf{Y}^{\perp\prime}\mathbf{Y}^\perp\end{pmatrix}.$$

and

$$C_{LM}=\chi^2_{1-\alpha}(k_2)/T.$$

Therefore the confidence interval will be unbounded when the $(2,2)$ element of Q_{LM} is negative, i.e., when

$$\frac{\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{Y}^\perp}{\mathbf{Y}^{\perp\prime}\mathbf{Y}^\perp} < C_{LM} \quad\text{or}\quad T\frac{\mathbf{Y}^{\perp\prime}\mathbf{P}_{Z^\perp}\mathbf{Y}^\perp}{\mathbf{Y}^{\perp\prime}\mathbf{Y}^\perp} < \chi^2_{1-\alpha}(k_2).$$

Note that the left-hand-side of the inequality is T times the uncentered R^2 from the first stage regression, i.e., the first stage LM statistic for testing $H_0 : \mathbf{\Pi} = 0$, which follows a $\chi^2(k_2)$ distribution asymptotically. Therefore, the $(1-\alpha)\cdot 100\%$ confidence interval obtained by inverting $\mathcal{GMM}_0$ will be unbounded if the first stage LM statistic is insignificant at level α.

(e) The proof is similar to the proof in (b), if we use the asymptotic (bounding) distribution for LR_{LIML} and $\mathcal{GMM}_0$.

VITA

My name is Jiahui Wang. I was born in August of 1969, in Guannan, Jiangsu Province, the Peoples Republic of China. I was admitted to the Department of World Economy at Fu Dan University in Shanghai in 1986, and received my Bachelors degree in economics in 1990. Afterwards I was admitted to the Institute of World Economy at Fu Dan University with honor for graduate study. In 1991 I was admitted to the Department of Economics at Washington State University in Pullman, Washington, and continued my graduate study there until I was admitted to the Ph.D. program in the Department of Economics at the University of Washington in 1993. In June of 1997 I received my Doctoral degree in economics from the University of Washington in Seattle, Washington.