

Abstract

KALAYLIOĞLU, ZEYNEP İŞİL. Frequentist and Bayesian Unit Root Tests in Stochastic Volatility Models. (Under the direction of Dr. Sujit K. Ghosh and Dr. Sastry G. Pantula.)

In stochastic volatility models, the unit root test on the time series of the unobserved log-volatilities may be performed by applying the commonly used frequentist unit root tests. For instance, augmented Dickey Fuller tests based on the log-squared mean corrected returns can be used. The log-squared mean corrected returns have the same second order properties as that of an autoregressive moving average process. However, we observed that the moving average parameter of the resulting process (based on the log-squared mean corrected returns) is typically close to the autoregressive parameter. For this reason, the unit root tests applied to stochastic volatility models tend to reject the unit root in finite samples. We propose a method for performing the frequentist unit root tests in stochastic volatility models based on the finite sampling distribution of the well known test statistics.

In addition to the frequentist testing procedures, Bayesian unit root tests can be used to test for a unit root in stochastic volatility models as well. A Bayesian test based on the Bayes factor has been suggested by So and Li (1999). In this approach, they work with the mean corrected returns instead of the log-squared mean corrected returns. They treat the unobserved log-volatilities as missing observations. The prior densities they use for the autoregressive parameter are continuous densities defined on an interval that does not include the value being tested. Such prior densities for the autoregressive parameter are not suitable where one's main concern is to test for a unit root in log-volatilities. We introduce a new prior density for this parameter that puts a positive mass on the point being tested. We also consider continuous prior densities defined on an interval that includes the point one. These prior densities allow us to use the posterior interval of the autoregressive parameter as a testing criterion. The advantage of our method is that it is simple and useful.

The performance of these tests are demonstrated by a simulation study. We illustrate the testing procedures by applying them to four sets of exchange rates.

**FREQUENTIST AND BAYESIAN UNIT ROOT TESTS IN
STOCHASTIC VOLATILITY MODELS**

by

ZEYNEP I. KALAYLIOĞLU

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

STATISTICS

in the

GRADUATE SCHOOL

at

NC STATE UNIVERSITY

2002

Dr. Sastry G. Pantula, Co-Chair

Dr. Sujit K. Ghosh, Co-Chair

Dr. David A. Dickey

Dr. Alastair R. Hall

To my dear mother and the memory of my dear father

Biography

Zeynep Işıl Kalaylıođlu was born on June 24, 1974, in Ankara, Turkey. She received a B.Sc. in Statistics in 1995 at Middle East Technical University in Ankara. She worked as a research and teaching assistant there for two years. A scholarship was bestowed upon her to get a Ph.D. in Statistics abroad by the Turkish Higher Education Council/Middle East Technical University. She joined the graduate program in Statistics at North Carolina State University in August, 1997. She got her M.S degree with a minor in Computational Engineering and Sciences in May, 2000. She defended her dissertation in January, 2002.

Acknowledgements

I would like to start with thanking Dr. Pantula for his guidance and great patience throughout my research. I feel lucky to have been exposed to his knowledge and way of thinking. I have learned from him to approach complicated problems by thinking the simple cases first and to pay attention to details. I would like to express my gratitude to Dr. Ghosh who taught me the Bayesian way of thinking. I can not thank him enough for his encouragement whenever I lost the sight of the light during this journey. I appreciate the comments and suggestions I received from my committee members Dr. Dickey and Dr. Hall.

I would like to extend my gratitudes to Middle East Technical University for giving me the opportunity to enhance my knowledge and perspective both in Statistics and in life by financially supporting my education in the U.S. I thank Dr. Taylan A. Ula and Dr. Sevtap A. Selcuk for their friendship.

I am indepted to Billy for his patience, tolerance, understanding, and always being there for me during my research. I will never forget that.

I don't think the words are enough to express my gratitude to my mom who supported me to go abroad for my graduate education inspite of the sorrow she was carrying inside. After I came here, she always encouraged me not to give up with her comforting talks on the phone while indeed her tears were sliding down to her heart. I am so grateful to my dear father for everything he did for me. I am not sad that he is not going to be there while I am getting my diploma. Because I know that he has been watching me from a distance in the blue sky and has been very proud of me.

Contents

List of Tables	vii
List of Figures	x
1 Introduction	1
1.1 Stochastic Volatility Models (SVM)	1
1.2 Parameter Estimation in SVM	3
1.3 Frequentist and Bayesian Unit Root Tests in SVM	9
1.4 Motivation for Testing the Unit Root Hypothesis in SVM	12
2 Frequentist Unit-Root Tests in Stochastic Volatility Models	18
2.1 Introduction	18
2.2 Augmented Dickey-Fuller Unit-Root Test in SVM	30
2.3 Said-Dickey (One Step Gauss-Newton) Unit-Root Test in SVM	37
2.4 Instrumental Variable Approach for a Unit-Root Test in SVM	40
2.5 Choosing the Critical Points	66
2.6 Monte Carlo Power Study	68
3 Bayesian Unit-Root Tests in Stochastic Volatility Models	82
3.1 Introduction	82
3.2 Choice of Prior Densities and Introducing Mixed Prior Density for ϕ	84
3.3 Posterior Inference in SVM	87
3.3.1 Gibbs Sampling in SVM	87
3.3.2 Gibbs Sampling in SVM via BUGS	91
3.4 Bayesian Unit-Root Test Based on the Bayes Factor	94
3.5 Bayesian Unit-Root Test Based on the Posterior Interval	99
3.6 Monte Carlo Power Study	100
3.7 Some Other Simulation Results	107
3.7.1 Some distributional properties of the parameters after having observed the data	107

3.7.2	Sensitivity of the posterior inference about ϕ to the prior distribution of ϕ	114
3.7.3	Proportion of the posterior intervals that contain the true value of ϕ when the true value of ϕ is 0.98 or 0.95	116
3.8	Discussion	117
4	Application	119
4.1	Application of Frequentist Unit Root Tests	122
4.2	Application of Bayesian Posterior Interval Unit Root Test	128
5	Conclusions	131
	Bibliography	137
A	Properties of the Joint Posterior Density of the Parameters in SVM	141
B	Smoothing the Percentiles	146
C	The Sampling Distribution of $\hat{\theta}$	163

List of Tables

1.1	Posterior Means of σ_η and ϕ	14
1.2	Estimates of ϕ and σ_η^2	16
2.1	Data Arrangement of WS Regression	32
2.2	Percentiles of Dickey-Fuller Pivotal Test Statistic. <i>Reported from second parts of Table 10.A.2. and Table 10.A.4 of Fuller(1996)</i>	33
2.3	Empirical 5th Percentiles for $\tau_{\mu,ADF/OLS,n,\theta}$. <i>Order of the Augmented Dickey-Fuller model fit is $p_{min} = \min(p_{AIC} + 2, [10n^{1/4}])$</i>	34
2.4	Empirical 10th Percentiles for $\tau_{\mu,ADF/OLS,n,\theta}$. <i>Order of the Augmented Dickey-Fuller model fit is $p_{min} = \min(p_{AIC} + 2, [10n^{1/4}])$</i>	34
2.5	Empirical 5th Percentiles for $\tau_{\mu,ADF/OLS,n,\theta}$. <i>Order of the Augmented Dickey-Fuller model fit is $p_{full} = [10n^{1/4}]$</i>	35
2.6	Empirical 10th Percentiles for $\tau_{\mu,ADF/OLS,n,\theta}$. <i>Order of the Augmented Dickey-Fuller model fit is $p_{full} = [10n^{1/4}]$</i>	35
2.7	Empirical 5th Percentiles for $\tau_{\mu,ADF/WS,n,\theta}$. <i>Order of the Augmented Dickey-Fuller model fit is $p_{full} = [10n^{1/4}]$</i>	36
2.8	Empirical 10th Percentiles for $\tau_{\mu,ADF/WS,n,\theta}$. <i>Order of the Augmented Dickey-Fuller model fit is $p_{full} = [10n^{1/4}]$</i>	36
2.9	Empirical 5th Percentiles for $\tau_{\mu,SDD,n,\theta}$	39
2.10	Empirical 10th Percentiles for $\tau_{\mu,SDD,n,\theta}$	39
2.11	Percentiles of Dickey-Fuller $n(\hat{\phi}_{ols} - 1)$ Test Statistic. <i>Reported from second part of Table 10.A.1 of Fuller(1996)</i>	59
2.12	Empirical 5th Percentiles for $(n(\hat{\phi}_{IV} - 1))_{n,\theta}$	60
2.13	Empirical 10th Percentiles for $(n(\hat{\phi}_{IV} - 1))_{n,\theta}$	61
2.14	Empirical 5th Percentiles for $\tau_{\mu,IV,Hall,n,\theta}$	61
2.15	Empirical 10th Percentiles for $\tau_{\mu,IV,Hall,n,\theta}$	62
2.16	Empirical 5th Percentiles for $\tau_{\mu,IV,SVM,n,\theta}^{rst}$	62
2.17	Empirical 10th Percentiles for $\tau_{\mu,IV,SVM,n,\theta}^{rst}$	63
2.18	Empirical 5th Percentiles for $(n(\hat{\phi}_{IV} - 1))_{n,\theta}^{SVM}$	63
2.19	Empirical 10th Percentiles for $(n(\hat{\phi}_{IV} - 1))_{n,\theta}^{SVM}$	64
2.20	Empirical 5th Percentiles for $\tau_{\mu,IV,Hall,n,\theta}^{SVM}$	64

2.21	Empirical 10th Percentiles for $\tau_{\mu,IV,Hall,n,\theta}^{SVM}$	65
2.22	Empirical 5th Percentiles for $\tau_{\mu,IV,SVM,n,\theta}^{SVM}$	65
2.23	Empirical 10th Percentiles for $\tau_{\mu,IV,SVM,n,\theta}^{SVM}$	66
2.24	Average of p over 1000 Monte Carlo Iterations. <i>Data are generated from SVM as in (2.1).</i>	70
2.25	Empirical Significance Levels and Powers for 5% Level Test Criteria. ($n=100$ and 1,000 Monte Carlo replications)	76
2.26	Empirical Significance Levels and Powers for 5% Level Test Criteria. ($n=500$ and 1,000 Monte Carlo replications)	77
2.27	Empirical Significance Levels and Powers for 5% Level Test Criteria. ($n=1000$ and 1,000 Monte Carlo replications)	78
2.28	Empirical Significance Levels and Powers for 10% Level Test Criteria. ($n=100$ and 1,000 Monte Carlo replications)	79
2.29	Empirical Significance Levels and Powers for 10% Level Test Criteria. ($n=500$ and 1,000 Monte Carlo replications)	80
2.30	Empirical Significance Levels and Powers for 10% Level Test Criteria. ($n=1000$ and 1,000 Monte Carlo replications)	81
3.1	Average Values of $\log_{10}(PO)$ and $\log_{10}(BF)$ and the Proportion of Correct Decisions (PCD). (<i>Sample Size $n=100$, 500 replications</i>). . .	102
3.2	Proportion of Correct Decisions (PCD) of the Bayesian Unit Root Test based on The Posterior Interval. (<i>Sample Size $n=100$, 500 replications</i>). . .	103
3.3	Number of Correct Decisions. (<i>100 replications</i>).	105
3.4	Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. <i>Sample Size $n=100$, Flat Priors on μ and σ_{η}^2</i>	108
3.5	Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. <i>Sample Size $n=500$, Flat Priors on μ and σ_{η}^2</i>	109
3.6	Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. <i>Sample Size $n=1000$, Flat Priors on μ and σ_{η}^2</i>	110
3.7	Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. <i>Sample Size $n=100$, Informative Priors on μ and σ_{η}^2</i>	111
3.8	Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. <i>Sample Size $n=500$, Informative Priors on μ and σ_{η}^2</i>	112
3.9	Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. <i>Sample Size $n=1000$, Informative Priors on μ and σ_{η}^2</i>	113

3.10	Sensitivity of Priors Using Wilcoxon's Rank Sum Test. ($H_0 : E[\phi \underline{x}, \phi \sim U(0, 1)] = E[\phi \underline{x}, \phi \sim \text{Mixed Uniform with } p = 0.95]$)	115
3.11	Sensitivity of Priors Using Wilcoxon's Rank Sum Test. ($H_0 : E[\phi \underline{x}, \phi \sim U(0, 1.95)] = E[\phi \underline{x}, \phi \sim \text{Mixed Uniform with } p = 0.95]$)	115
3.12	The number of the posterior intervals that include the true value of ϕ . <i>100 Monte Carlo replications.</i>	116
4.1	p-values for the Ljung-Box Test Statistic	121
4.2	Different estimates of ϕ and σ^2 of the ARMA(1,1) model that corresponds to the SVM.	122
4.3	Post. Inference on ϕ . $\phi \sim \text{Mixed } U(0, 1)$ with $p = 0.95$	129
4.4	Post. Inference on ϕ and p . $\phi \sim \text{Mixed } U(0, 1)$ with $p \sim \text{Beta}(0.5, 0.5)$	129
4.5	Post. Inference on ϕ . $\phi \sim U(0, 1.95)$	130
5.1	Total Error Rates. <i>Sample Size</i> $n = 100$, <i>Nominal Level for the Frequentist Tests:</i> 0.05	134
5.2	Total Error Rates. <i>Sample Size</i> $n = 500$, <i>Nominal Level for the Frequentist Tests:</i> 0.05	135
5.3	Total Error Rates. <i>Sample Size</i> $n = 1000$, <i>Nominal Level for the Frequentist Tests:</i> 0.05	136
B.1	Estimates of the smoothing parameters $(\lambda_1, \lambda_2, \lambda_3)$. ($n=100$)	147
B.2	Estimates of the smoothing parameters $(\lambda_1, \lambda_2, \lambda_3)$. ($n=500$)	148
B.3	Estimates of the smoothing parameters $(\lambda_1, \lambda_2, \lambda_3)$. ($n=1000$)	149
C.1	Sampling Distribution of $\hat{\theta}_{durbin}$. (<i>1,000 Monte Carlo Replications</i>)	163
C.2	Sampling Distribution of $\hat{\theta}_{adj.durbin}$. (<i>1,000 Monte Carlo Replications</i>)	164
C.3	Sampling Distribution of $\hat{\theta}_{one\ step\ gn}$. (<i>1,000 Monte Carlo Replications</i>)	164

List of Figures

1.1	θ versus σ_η^2	13
2.1	Mean Corrected Returns versus Time. <i>Seed Set #1</i>	20
2.2	Mean Corrected Returns versus Time. <i>Seed Set #1</i>	21
2.3	Mean Corrected Returns versus Time. <i>Seed Set #1</i>	22
2.4	Mean Corrected Returns versus Time. <i>Seed Set #1</i>	23
2.5	Mean Corrected Returns versus Time. <i>Seed Set #2</i>	24
2.6	Mean Corrected Returns versus Time. <i>Seed Set #2</i>	25
2.7	Mean Corrected Returns versus Time. <i>Seed Set #2</i>	26
2.8	Mean Corrected Returns versus Time. <i>Seed Set #2</i>	27
3.1	DAG representation of SVM with mixed density on ϕ	92
4.1	Plot of the mean corrected returns r_{it} versus time	120
B.1	Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}(p_{min}),n = 100$. <i>Top figure</i> : level of the test=0.05, <i>Bottom figure</i> : level of the test=0.10	150
B.2	Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}(p_{min}),n = 500$. <i>Top figure</i> : level of the test=0.05, <i>Bottom figure</i> : level of the test=0.10	151
B.3	Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}(p_{min}),n = 1000$. <i>Top figure</i> : level of the test=0.05, <i>Bottom figure</i> : level of the test=0.10	152
B.4	Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}(p_{full}),n = 100$. <i>Top figure</i> : level of the test=0.05, <i>Bottom figure</i> : level of the test=0.10	153
B.5	Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}(p_{full}),n = 500$. <i>Top figure</i> : level of the test=0.05, <i>Bottom figure</i> : level of the test=0.10	154
B.6	Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}(p_{full}),n = 1000$. <i>Top figure</i> : level of the test=0.05, <i>Bottom figure</i> : level of the test=0.10	155
B.7	Prediction Curve for $\tau_{\mu,ADF/WS,n,\theta}(p_{full}),n = 100$. <i>Top figure</i> : level of the test=0.05, <i>Bottom figure</i> : level of the test=0.10	156
B.8	Prediction Curve for $\tau_{\mu,ADF/WS,n,\theta}(p_{full}),n = 500$. <i>Top figure</i> : level of the test=0.05, <i>Bottom figure</i> : level of the test=0.10	157

B.9	Prediction Curve for $\tau_{\mu,ADF/WS,n,\theta}, (p_{full}), n = 1000$. <i>Top figure:</i> level of the test=0.05, <i>Bottom figure:</i> level of the test=0.10	158
B.10	Prediction Curve for $\tau_{\mu,SDD,n,\theta}, n = 100$. <i>Top figure:</i> level of the test=0.05, <i>Bottom figure:</i> level of the test=0.10	159
B.11	Prediction Curve for $\tau_{\mu,SDD,n,\theta}, n = 500$. <i>Top figure:</i> level of the test=0.05, <i>Bottom figure:</i> level of the test=0.10	160
B.12	Prediction Curve for $\tau_{\mu,SDD,n,\theta}, n = 1000$. <i>Top figure:</i> level of the test=0.05, <i>Bottom figure:</i> level of the test=0.10	161
B.13	Prediction Curve for $\tau_{\mu,IV,Hall,n,\theta}, n = 100$. <i>Top figure:</i> level of the test=0.05, <i>Bottom figure:</i> level of the test=0.10	162

Chapter 1

Introduction

1.1 Stochastic Volatility Models (SVM)

In finance it is very common to see that the mean corrected return on holding an asset has time dependent conditional variance (i.e. volatility) σ_t^2 . Explaining this variation is a fundamental problem in finance. Let r_t denote a series of interest such as a mean corrected stock return. It is very common to model r_t as $r_t = \sigma_t u_t$ where the u_t 's are independent and identically distributed as $N(0,1)$. Here σ_t is the square root of the volatility and hence conditional standard deviation of the mean corrected stock returns. As an example of return data r_t , first: let p_t denote the Turkish Lira/U.S. Dollar exchange rate at time t . Let n be the sample size. Then mean corrected return, r_t , may be computed as $r_t = [(\log p_t - \log p_{t-1}) - \frac{1}{n} \sum_{i=1}^n (\log p_i - \log p_{i-1})]$. A person must make a decision based on the distribution of the return. At this point not only the mean of the distribution of the return but also some measure of the dispersion of that distribution is of primary importance. Conditional variance (volatility) of the return is a measure of its dispersion. It is a measure of instability of return. In other words, it is a measure of uncertainty. Uncertainty is especially important in finance and business world since it affects one's decision and one's decision affects the stock prices. Interest in modeling unobserved volatility, i.e. conditional variance of returns,

which changes over time dates back to the 1970's. Researchers have been interested in modeling the time dependent feature of unobserved volatility. A model that is commonly used is

$$\begin{aligned} r_t &= \sigma_t u_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 \end{aligned}$$

The likelihood function of $\{r_t\}$ is easy to maximize in this setup. This model is also called an Autoregressive Conditional Heteroscedasticity (ARCH) model of order 1. In this model, conditional variance function of mean corrected return data is expressed in terms of past data, r_{t-1} , and unknown parameters α_0 and α_1 .

Generalized ARCH (GARCH) models have been popular to model the volatility over time due to the same reason as stated above. The GARCH model which is used most commonly can be expressed as

$$\begin{aligned} r_t &= \sigma_t u_t \\ \sigma_t^2 &= \gamma + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned}$$

where $\gamma > 0$ and $\alpha + \beta < 1$. This model is called a GARCH(1,1) model. In this setup, volatility is modeled as depending on the past volatility and the square of the past return. In ARCH and GARCH models, given the past observations volatility is deterministic.

An alternative approach is to use Stochastic Volatility Models (SVM). In the SVM, the logarithm of the unobserved volatility, $\log \sigma_t^2$, is modeled as a stochastic process. Let us denote $\log \sigma_t^2$ by h_t . It is more convenient to model h_t than to model σ_t^2 as the range of h_t is the entire real line. A common model for h_t is an autoregressive model of order 1, AR(1), with nonzero mean μ and error term η_t . Hence a SVM is written as

$$\begin{aligned} r_t &= e^{h_t/2} u_t, u_t \sim NID(0, 1) \\ (h_t - \mu) &= \phi(h_{t-1} - \mu) + \eta_t, \eta_t \sim NID(0, \sigma_\eta^2) \end{aligned} \tag{1.1}$$

The two error terms, u_t and η_t , are usually assumed to be uncorrelated. Parameters of the SVM consist of the parameters of the AR(1) model for the log volatility, namely, the mean of the log volatility μ , an autoregressive parameter ϕ , and the error variance σ_η^2 . In the literature, there are also SV models with slightly modified assumptions. For instance, the nature of the problem at hand might motivate the practitioner to assume that the two random variables in the SVM, u_t and η_t , are correlated. Another modification might be to include a multiplying constant in the equation for the mean corrected return data which can be thought of as the model instantaneous volatility. In this research, we focus on the unit-root test in (1.1) assuming that the error terms u_t and η_t are uncorrelated.

1.2 Parameter Estimation in SVM

Jacquier, Polson, and Rossi (1994) focus on estimating the parameters of SVM which is described in (1.1) with $\begin{pmatrix} u_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}\right)$. They write the stochastic model for the log volatility with an intercept term; $h_t = \alpha + \phi h_{t-1} + \eta_t$. Here the parameters are α , ϕ , and σ_η^2 . They assume that $|\phi| < 1$. They view the SVM as a three-level hierarchical Bayesian model and compute Bayes estimators. In the first level of hierarchy, the conditional distribution of the data is specified given the unobserved volatility. At the second level, the conditional distribution of the unobserved volatility is specified given the hyperparameters, i.e. α , ϕ , and σ_η^2 . Finally at the last level, they specify a conjugate prior for $(\alpha, \phi, \sigma_\eta)$; $f(\alpha, \phi, \sigma_\eta) = f(\alpha, \phi | \sigma_\eta) f(\sigma_\eta)$ where $\alpha, \phi | \sigma_\eta \sim MN((\bar{\alpha}, \bar{\phi})', \sigma_\eta^2 A^{-1})$ and $\sigma_\eta \sim IG(\eta_0, s_0^2)$ with η_0 and s_0^2 to be known. Viewing the unobserved volatility as missing data, they use a Markov Chain Monte Carlo (MCMC) algorithm that generates samples from the stationary distribution which is the posterior distribution of the parameters given the observed returns. We denote the unobserved log-volatilities, observed returns,

and the parameters of SVM by $h = (h_1, \dots, h_n)$, $r = (r_1, \dots, r_n)$, and $w = (\alpha, \phi, \sigma_\eta^2)$. respectively. In addition, h_{-t} is the vector h excluding the element h_t . The steps of the algorithm that they used are as follows;

1. Initialize h and w
2. Sample the vector w from $f(w|h, r)$.
3. Sample h_t from $f(h_t|h_{-t}, r, \mu, \phi, \sigma_\eta^2)$ for $t=1, \dots, n$ by using a Metropolis accept-reject algorithm.
4. Go to step 2.

They estimate each parameter by the average of the points in the corresponding Markov Chain which is constructed by applying the algorithm above. They also find the maximum likelihood estimates of the parameters. First, they express the SVM as a linear state space model by taking the logarithm of the squares of the observations:

$$\begin{aligned} \log r_t^2 &= h_t + \log u_t^2 \\ (h_t - \mu) &= \phi(h_{t-1} - \mu) + \eta_t \end{aligned} \quad (1.2)$$

They define the likelihood function using the predictive error decomposition (see Harvey, Ruiz, and Shephard (1994));

$$\log L(\mu, \phi, \sigma_\eta^2) \propto \frac{-1}{2} \sum_{t=1}^n \log f_t - \frac{1}{2} \sum_{t=1}^n \frac{e_t^2}{f_t} \quad (1.3)$$

where e_t is the one-step ahead prediction error for the best linear estimator of $\log r_t^2$ and f_t is the prediction error variance. They denote the best linear estimator of $\log r_t^2$ by $\overline{\log r_t^2}$. Hence $e_t = \log r_t^2 - \overline{\log r_t^2} = \log r_t^2 - \overline{h_{t/t-1}} - (-1.27)$ where $\overline{h_{t/t-1}}$ is the prediction of h_t based on the information available at time $t-1$ and (-1.27) is the expectation of $\log u_t^2$. In addition, prediction error variance f_t is expressed as $f_t = \Omega_{t/t-1} + \pi^2/2$ where $\Omega_{t/t-1}$ is the variance of the prediction of h_t and $\pi^2/2$ is the variance of $\log u_t^2$. They compute the prediction of unobserved log-volatility, $\overline{h_{t/t-1}}$, and variance of the prediction, $\Omega_{t/t-1}$, by employing a Kalman filtering algorithm assuming that the noise term $\log u_t^2$ has a Normal distribution. They call

the maximum likelihood estimators obtained by maximizing this likelihood function Quasi Maximum Likelihood Estimators. The third estimation procedure they applied to estimate the parameters of an SVM is the method of moments estimates. They use the first four moments of $|r_t|$, 10 lags of $E[r_t^2 r_{t-k}^2]$, and 10 lags of $E[|r_t| |r_{t-k}|]$ to compute the method of moments estimates. They compare the mean and the root mean squared errors of Bayes estimators obtained via MCMC method to the method of moments and quasi-maximum likelihood estimators in SVM via a simulation study. Their simulation results show that Bayes estimators perform better than the other two estimators.

Kim, Shephard, and Chib (1998) fit SVM as in (1.1) to the data set they analyse. They assume that u_t and η_t are uncorrelated. Also, they assume a stationary AR(1) process for the log-volatility. The data consist of the daily observations of weekday close exchange rates for the U.K. Sterling/U.S. Dollar from 1/10/81 to 28/6/85. The sample size is $n=946$. They employ three different MCMC techniques to estimate the parameters. The first one of the MCMC algorithms they employed uses the Model 1.1. In this algorithm, volatilities are sampled one at a time. The algorithm is as follows;

1. Initialize h , ϕ , σ_η^2 , and μ .
2. Sample h_t from $f(h_t | h_{-t}, r, \phi, \sigma_\eta^2, \mu)$ for $t = 1, \dots, n$.
3. Sample σ_η^2 from $f(\sigma_\eta^2 | h, r, \phi, \mu)$.
4. Sample ϕ from $f(\phi | h, r, \sigma_\eta^2, \mu)$.
5. Sample μ from $f(\mu | h, r, \sigma_\eta^2, \phi)$.
6. Go to step 2.

They indicate that this algorithm is not efficient as the volatilities are sampled one at a time. In order to improve the algorithm, the volatilities should be sampled all at once. They suggest another algorithm to do this and they call this one the Improved MCMC algorithm. First, they express the Model 1.1 as a linear state space form by a log-square transformation. They take into account that the mean adjusted return r_t

can take the value 0, and so they use $\log(r_t^2 + c)$ rather than $\log r_t^2$ in the state space representation of the model. Here c is a subjectively chosen constant. Throughout their study, they set $c = 0.001$. They note that $\{\log r_t^2\}$ and $\{\log(r_t^2 + c)\}$, of course, contain the same information. They approximate the exact density of $\log u_t^2$ by a mixture of normal densities. That is $f(\log u_t^2) = \sum_{i=1}^K q_i f_N(\log u_t^2 | m_i - 1.27, v_i^2)$. In this expression, K is the number of the mixing normal densities, f_N is a normal density with component probability q_i , mean $m_i - 1.27$, and variance v_i^2 . They note that this mixture density can also be written in terms of a component indicator variable s_t . That is, $(\log u_t^2 | s_t = i) \sim N(m_i - 1.27, v_i^2)$ with $Pr(s_t = i) = q_i$. They select the number of mixing densities as well as the mean, variance, and the weight (component probability) of each density so as to have the approximating density lie within a small distance of the true density. Note that in the following development, $s = (s_1, \dots, s_n)$, $h = (h_1, \dots, h_n)$, $y^* = (\log(r_1^2 + c), \dots, \log(r_n^2 + c))$. Steps of the first MCMC algorithm that they suggest are as follows;

1. Initialize s , ϕ , σ_η^2 , and μ .
2. Sample h from $f(h | y^*, s, \phi, \sigma_\eta^2, \mu)$.
3. Sample s from $f(s | y^*, h)$.
4. Sample ϕ , σ_η^2 , and μ from their full conditional densities $f(\phi | \sigma_\eta^2, \mu, h)$, $f(\sigma_\eta^2 | \mu, \phi, h)$, and $f(\mu | \phi, \sigma_\eta^2, h)$ respectively.
5. Go to step 2.

Writing the model in a linear state space form with an error term whose distribution is approximated by a mixture of normal densities enables the use of Kalman filter runs in Step 2. See the Appendix part of Kim et al (1998) for more details.

They note that the output of this chain suggests that 20,000 draws from this sampler is sufficient to do the inference. They indicate that although this algorithm is a big improvement over the MCMC algorithm of Jacquier et al., convergence of the chain is still slow because of the high correlation between the volatilities and the parameters. They propose another algorithm for which they call Integration Sampler.

The structure of this algorithm is as follows;

1. Initialize $(s, \phi, \sigma_\eta, \mu)$.
2. Sample (ϕ, σ_η^2) from $f(\phi, \sigma_\eta^2)$ by using Metropolis-Hasting accept-reject probability.
3. Sample (h, μ) from $f(h, \mu | y^*, s, \phi, \sigma_\eta^2)$
4. Sample s from $f(s | y^*, h)$.
5. Go to step 2.

They note that the output of this sampler suggests that 2000 draws from this sampler is sufficient for inferential purposes. For this data set, they compare the two MCMC algorithms based on the simulation inefficiencies. Simulation inefficiency is one of the diagnostic statistics which measures how well the chain mixes. It is estimated as the variance of the sample mean from the MCMC sampling scheme divided by the variance of the sample mean from a hypothetical sampler which draws independent random variables from the posterior. The variance of the sample mean from the MCMC sampling is estimated by the time series methods to account for the serial correlation in the draws. They conclude that the second algorithm performs better. They state that this conclusion seems to hold up for some other exchange rate series.

Sandmann and Koopman (1998) introduce a nonzero correlation between the u_t and η_t , a multiplying constant, β , in the equation for the mean corrected return, and zero mean for log-volatility. That is, $r_t = \beta e^{\frac{h_t}{2}} u_t$ where $h_t = \phi h_{t-1} + \eta_t$ and $\begin{pmatrix} u_t \\ \eta_t \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{pmatrix} \right)$. Here, r_t is the mean adjusted return on an asset. They work with the linearized model, $\log r_t^2 = 2 \log \beta + h_t + \log u_t^2$ as well. Harvey et al. (1994) proved that as long as u_t and η_t are jointly symmetric, $\text{Cov}(\log u_t^2, \eta_t) = 0$ even if u_t and η_t have nonzero correlation. Sandmann and Koopman decompose the exact likelihood function of $\log r_t^2$ into two components one of which is the Gaussian log-likelihood function and the other one is a correction factor for the departures from the Gaussian assumption. Gaussian log-likelihood component is based on the model with disturbances $\log u_t^2$ distributed as $N(0, H_t)$ where they select the variance

H_t by equalizing the derivative of the log density of $N(0, H_t)$ to the log density of $\log \chi^2$ with 1 degree of freedom. The Gaussian log-likelihood is constructed by using Kalman filter. They replace the second component, i.e. the correction factor, by its unbiased estimator. The unbiased estimate is calculated by simulation. They call the estimators obtained by numerical optimization of the sum of these two components Monte Carlo Maximum Likelihood (MCL) Estimators. In addition to this estimation technique, they also find Quasi Maximum Likelihood (QML) Estimators obtained by optimizing the Gaussian log-likelihood function which is given in equation (1.3). Also, they use the method of Jacquier et al. to obtain Bayesian estimators of the parameters. They perform a simulation experiment to compare MCL estimators with QML and MCMC estimators using Mean Squared Errors. They find QML inefficient but not as inefficient as reported in Jacquier et al. They indicate that this raises a question regarding the QML estimation of Jacquier et al. in terms of poor starting values, a different convergence criteria, or inefficient implementation of the algorithm. The results of their experiment also demonstrate that MCL estimators have similar mean squared errors as MCMC estimators across most parameter values. They conclude that their estimators perform well in small samples and hence serve as an alternative to the MCMC estimators of Jacquier et al.

There are several examples in the literature where the SVM is applied in financial time series data such as stock returns and exchange rates to estimate the volatilities and the change in volatility by using the various estimation methods outlined above. For instance, Chib, Nardari, and Shephard (1998) apply a SVM to the data that consist of daily continuously compounded returns on the Standard and Poor's 500 index from January 2, 1980 through December 29, 1987. The total number of observations is 2022. They estimate the parameter ϕ to be 0.981. In other real data examples that we present in the next section, we notice that the estimates of ϕ are close to 1. This raises the question of a unit-root in the model for log-volatility.

1.3 Frequentist and Bayesian Unit Root Tests in SVM

When there is a unit root in the log-volatility process, shocks to volatility don't decay rapidly. That is, the effect of past shocks on current volatility (conditional variance of returns) remain persistent for long periods. Persistence of shocks to volatility has important implications in economics and finance. High persistence of shocks to volatility increases the fluctuation in the volatility (change in the return over time). As a result, the business environment becomes more uncertain. That causes the market to plunge. See Pyndick (1984, 1986), Poterba and Summers (1986), Chou (1988), Bollerslev and Engle (1993) who studied the volatility and its relationship with market fluctuations. Tests for a unit-root in SV models continue to receive a lot of attention.

Harvey, Ruiz, and Shephard (1994) fit a SVM to four exchange rates. They assume u_t and η_t are generated independently. They define the SVM as in (1.1). They find that QML estimates of the parameter ϕ in each of the four series are respectively 0.9912, 0.9646, 0.9948, and 0.9575. Corresponding standard errors are respectively 0.007, 0.021, 0.005, and 0.002. They test the null hypothesis of a unit root by applying Augmented Dickey-Fuller (ADF) unit-root test (see Said and Dickey, 1984) to each of the four series in the form of $\log r_t^2$. The ADF regression they fit has nine lags and a constant term. They use Dickey-Fuller critical point for 1% level and 500 observations. The hypothesis of a unit-root is strongly rejected at the 1% level for all the series. Based on these examples, they conclude that reliability of the classical ADF test is questionable in SVM. They make this argument based on this example not based on a simulation study.

So and Li (1999) propose a Bayesian unit-root test procedure to test for the unit root in SVM which is defined in (1.1). They assume that u_t and η_t are independent. They assign noninformative normal and noninformative inverted gamma priors for the

mean μ and the error variance σ_η^2 of the log-volatility respectively. They consider four different prior densities for the persistence parameter ϕ . The first one is the simplest prior that has been used in the Bayesian literature. This is the uniform prior defined between 0 and 1. This is the only noninformative prior they consider for ϕ . Another one they choose as a prior density for ϕ is the truncated Normal density on the interval $(0, 1)$. The other informative priors they select are the Beta priors. They treat the unit-root testing as comparing the model under the null hypothesis to the model under the alternative hypothesis. They select the model based on the value of the posterior odds (PO). If one labels the model under the null as M_0 and the model under the alternative as M_1 , then PO is the ratio of $p(M_0|\underline{r})$ to $p(M_1|\underline{r})$ where \underline{r} represents the mean corrected returns i.e. the data. This expression is equal to the product of the Bayes Factor (BF), $p(\underline{r}|M_0)/p(\underline{r}|M_1)$, and the prior odds, $p(M_0)/p(M_1)$. Usually the prior odds is chosen to be 1 to indicate prior ignorance. In this case, PO is equal to the BF. The numerator and the denominator of BF are the marginal likelihood of the data under the null model and the alternative model respectively. To evaluate $p(\underline{r}|M_i)$, the marginal likelihood under the model M_i , they use the ratio $\frac{P_i(\underline{r}|\beta_i)f(\beta_i)}{P_i(\beta_i|\underline{r})}$ where β_i is the vector of the parameters under the model M_i . Here, $P_i(\underline{r}|\beta_i)$ is the likelihood function of the mean corrected returns and $P_i(\beta_i|\underline{r})$ is the joint posterior density of the parameters, and $f(\beta_i)$ is the joint prior density of the parameters under the model M_i . This identity is evaluated at the posterior mean of β_i , denoted by β_i^* , as advised by Chib (1995). In order to calculate these densities, they implement the Gibbs Sampling algorithm. Full conditional densities of μ and σ_η^2 are standard known distributions therefore samples from these distributions are drawn directly. The full conditional density of ϕ is not a standard density so sampling from this density is done by an accept-reject algorithm. To simulate \underline{h} all at once, they follow the method of Kim et al. (1998) as explained earlier.

So and Li obtain the posterior ordinate evaluated at the posterior means, $P_i(\beta_i^*|\underline{r})$, by following the procedure of Chib (1995). Note that under the alternative, the

posterior ordinate evaluated at posterior means is

$$P_i(\mu^*, \phi^*, \sigma_\eta^{2*} | \mathbf{r}) = p(\mu^* | \phi^*, \sigma_\eta^{2*}, \mathbf{r}) p(\phi^* | \sigma_\eta^{2*}, \mathbf{r}) p(\sigma_\eta^{2*} | \mathbf{r})$$

where

$$\begin{aligned} p(\mu^* | \phi^*, \sigma_\eta^{2*}, \mathbf{r}) &= \int p(\mu^* | \phi^*, \sigma_\eta^{2*}, \mathbf{h}, \mathbf{r}) p(\mathbf{h} | \phi^*, \sigma_\eta^{2*}, \mathbf{r}) d\mathbf{h} \\ p(\phi^* | \sigma_\eta^{2*}, \mathbf{r}) &= \int p(\phi^* | \mu, \sigma_\eta^{2*}, \mathbf{h}, \mathbf{r}) p(\mu, \mathbf{h} | \phi^*, \sigma_\eta^{2*}, \mathbf{r}) d\mathbf{h} d\mu \\ p(\sigma_\eta^{2*} | \mathbf{r}) &= \int p(\sigma_\eta^{2*} | \mu, \phi, \mathbf{h}, \mathbf{r}) p(\mathbf{h}, \mu, \phi | \mathbf{r}) d\mathbf{h} d\mu d\phi \end{aligned}$$

Appropriate Monte Carlo estimates of $p(\mu^* | \phi^*, \sigma_\eta^{2*}, \mathbf{r})$, $p(\phi^* | \sigma_\eta^{2*}, \mathbf{r})$, and $p(\sigma_\eta^{2*} | \mathbf{r})$ are $\frac{1}{G} \sum_{g=1}^G p(\mu^* | \phi^*, \sigma_\eta^{2*}, \mathbf{h}^{(g)}, \mathbf{r})$, $\frac{1}{G} \sum_{g=1}^G p(\phi^* | \mu^{(g)}, \sigma_\eta^{2*}, \mathbf{h}^{(g)}, \mathbf{r})$, and $\frac{1}{G} \sum_{g=1}^G p(\sigma_\eta^{2*} | \mu^{(g)}, \phi^{(g)}, \mathbf{h}^{(g)}, \mathbf{r})$. Here, G is the number of Gibbs Sampling iterations and $\mathbf{h}^{(g)}, \mu^{(g)}, \phi^{(g)}$ are the log-volatilities and parameters sampled in each iteration of Gibbs Sampling. These Monte Carlo estimates are consistent estimators of these ordinates. Each of these Monte Carlo estimate is basically the average of the joint full conditional density of the corresponding parameter.

In order to calculate the likelihood function $P_i(\mathbf{r} | \mu, \sigma_\eta^2, \phi)$, So and Li use the simulation based method of Kim et al. (1998).

The rejection criterion of So and Li is to reject the null hypothesis if PO is less than or equal to 1. They conducted experiments to determine the proportion of correct decisions with their Bayesian unit-root test for various combinations of sample size n , autoregressive parameter ϕ , and the prior distribution of ϕ . Simulation results show that one can get reliable results from So and Li's Bayesian unit-root test. Those results will be presented later in Section 3.6.

1.4 Motivation for Testing the Unit Root Hypothesis in SVM

As mentioned in the previous section, a SVM can be transformed to an ARMA(1,1) model. As in Harvey et al., consider

$$\log r_t^2 = h_t + \log u_t^2$$

Since h_t is an AR(1) process and $\log u_t^2$ is a sequence of iid random variables, $\log r_t^2$ will have the same correlation structure as an ARMA(1,1) process with parameters $c, \phi, \theta, \sigma_\epsilon^2$ given by

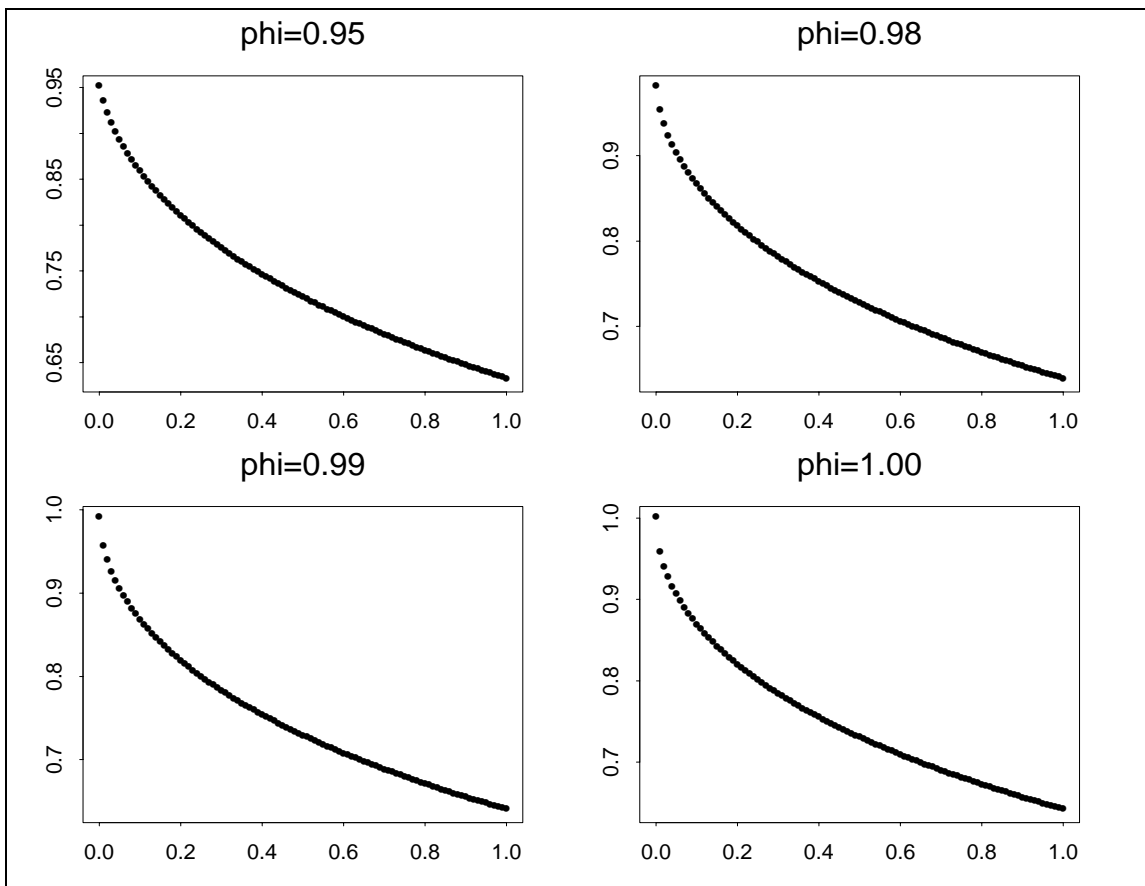
$$\log r_t^2 = c + \phi \log r_{t-1}^2 - \theta v_{t-1} + v_t$$

where v_t is a $(0, \sigma^2)$ white noise process. Note that the coefficient of $\log r_{t-1}^2$ is the same as ϕ in the AR(1) model for h_t . Hence, one may test the unit root hypothesis using the methods for unit root tests in ARMA(1,1) models. Note that the other parameters c, θ , and σ^2 of the ARMA(1,1) model are related to μ, ϕ , and σ_η^2 . Let a and σ_ϵ^2 be respectively the mean and variance of $\log u_t^2$. Here, $a = E[\log u_t^2] \cong -1.27$ and $\sigma_\epsilon^2 = \frac{\pi^2}{2}$. Then, the relationships are

$$\begin{aligned} c &= (a + \mu)(1 - \phi) \\ \theta &= \phi \frac{\sigma_\epsilon^2}{\sigma^2} = \phi \frac{\pi^2}{2\sigma^2} \\ \sigma^2 &= \frac{(\sigma_\eta^2 + (1 + \phi^2)\sigma_\epsilon^2) + \sqrt{(\sigma_\eta^2 + (1 + \phi^2)\sigma_\epsilon^2)^2 - 4\phi^2(\sigma_\epsilon^2)^2}}{2} \end{aligned} \tag{1.4}$$

The plots in Figure (1.1) are useful in terms of displaying how the true value of θ of the transformed model gets influenced by the true values of ϕ and σ_η^2 of the original model. In each plot, σ_η^2 is the horizontal axis whereas θ is the vertical axis.

What one can observe from the plots is that, if the true value of the error variance σ_η^2 is roughly less than 0.2, the true value of the moving average parameter θ is close

Figure 1.1: θ versus σ_η^2

to the true value of the autoregressive parameter ϕ . This situation causes the unit-root null hypothesis being rejected most of the time although the parameter ϕ is 1. Schwert (1989) shows by an extensive simulation study that ADF unit-root test statistic for the ARMA models where the MA parameter is close to the AR parameter has a different finite sampling distribution than the one which is tabulated in Fuller (1996). This provides us a motivation to develop new unit root tests for SV Models as we see that in most real data sets σ_η^2 is less than 0.2. Below, we present some of many such cases in Examples 1.2.1 and 1.2.2. Therefore, one has to be cautious in performing various statistical unit root test procedures in SVM such as ADF.

Example 1.2.1. Jacquier et al. fit SVM separately to each of the four series they have. Those series are equal-weighted (EW) and value-weighted (VW) indices of NYSE, and three decile portfolios corresponding to the 1, 5, and 10 deciles of stock, denoted by D1, D5, D10 respectively. They use Bayes estimators obtained via a MCMC algorithm to estimate the parameters of the SVM, namely μ , ϕ , σ_η^2 . The Bayes estimators, i.e. the posterior means, of the parameters are shown in the following table.

Table 1.1: Posterior Means of σ_η and ϕ

	σ_η	ϕ
EW	0.39 (0.025)	0.91 (0.015)
VW	0.23 (0.026)	0.95 (0.013)
D1	0.32 (0.032)	0.93 (0.016)
D5	0.32 (0.095)	0.91 (0.046)
D10	0.29 (0.056)	0.93 (0.022)

In this table, the values in the parentheses are the posterior standard deviations of σ_η and ϕ . From the table above, one can see that the estimates of σ_η^2 are all less

than 0.2 which implies that true value of θ may be close to that of ϕ in the ARMA model.

Example 1.2.2. Mahieu and Schotman (1998) analyze the weekly observations on bilateral exchange rates among US Dollar, Pound Sterling, Japanese Yen, and Deutschmark: Yen/Dollar, Deutschmark/Dollar, Pound/Dollar, Deutschmark/Yen, Pound/Yen and Deutschmark/Pound. The sample period is January 3, 1973 until February 9, 1994 resulting in 1102 observations. The data are obtained from DATA-STREAM and are sampled on Wednesdays. Mahieu and Schotman define the return as the difference between the logarithm of the exchange rate at time t and the logarithm of the exchange rate at time $(t-1)$. They fit a SVM to the mean corrected returns. They assume that u_t and η_t are independent. They have a scale parameter as a multiplier in their model for the mean corrected return and they have the mean parameter, μ fixed at 0 in the model for log-volatility in order to provide identifiability. They work with the log-squared model. They use various types of estimation procedures depending on how the true distribution of $\log u_t^2$ is approximated. They use QML estimation procedure when they approximate the true density by a Normal density with fixed error variance (estimation procedure denoted by QML1) and unrestricted error variance (estimation procedure denoted by QML2). In the second case, they have one additional unknown parameter. They use Simulated EM algorithms when they approximate the true density by three mixing Normal densities with fixed six elements (estimation procedure denoted by SIEM1) and with flexible six elements (estimation procedure denoted by SIEM2), where these six elements are nothing but the mean and variance of the approximating normal densities. Estimation results are given in the following table. The first half of the table shows the estimates and the MSE of the estimates of ϕ , whereas the second half shows the ones for σ_η^2 .

From this table as well, we can see that it is very usual for σ_η^2 to have a value less than roughly 0.2. Another thing that one should notice from these examples is that the estimates of ϕ are close to 1. An overall conclusion from the real data examples

Table 1.2: Estimates of ϕ and σ_{η}^2

	JP/US	GE/US	UK/US	GE/JP	UK/JP	GE/UK
QML1	0.976 (0.015)	0.967 (0.027)	0.960 (0.015)	0.985 (0.009)	0.653 (0.391)	0.884 (0.061)
QML2	0.983 (0.010)	0.978 (0.016)	0.963 (0.014)	0.985 (0.010)	0.981 (0.015)	0.895 (0.05)
SIEM1	0.878 (0.016)	0.928 (0.013)	0.952 (0.009)	0.921 (0.014)	0.584 90.042)	0.768 (0.027)
SIEM2	0.979 (0.007)	0.975 (0.007)	0.967 (0.008)	0.954 (0.011)	0.957 (0.011)	0.930 (0.014)
QML1	0.05 (0.081)	0.04 (0.097)	0.09 (0.055)	0.008 (0.027)	0.71 (0.612)	0.18 (0.142)
QML2	0.03 (0.051)	0.02 (0.062)	0.09 (0.055)	0.008 (0.031)	0.01 (0.057)	0.15 (0.125)
SIEM1	0.31 (0.024)	0.08 (0.018)	0.10 (0.017)	0.08 (0.018)	0.68 (0.036)	0.42 (0.028)
SIEM2	0.04 (0.013)	0.02 (0.012)	0.06 (0.013)	0.04 (0.015)	0.04 (0.016)	0.08 (0.018)

that we have looked at here and from many others that are in the literature is that these are very common features of the financial time series data.

In this research, we adjust the most commonly used frequentist unit-root testing procedures in the literature to be safely used for unit-root testing in SVM. In addition to frequentist approaches, we make use of the Bayesian tools to perform the unit-root test in SVM.

The rest of the thesis consists of four parts. The first part is the study of the frequentist unit-root tests in SVM, the second part is the study of Bayesian techniques for unit-root testing in SVM, the third part is an application of the tests discussed here to a real data set, and the last part contains some comparisons between frequentist and Bayesian tests. Basically these parts are discussed in Chapter 2, Chapter 3, Chapter 4, and Chapter 5 respectively. In Chapter 2, we study several frequentist test criteria and present adjustments to these criteria for testing the unit-root in SVM. We simulate the empirical percentiles of these test statistics and present the

results from a power study. Chapter 3 discusses the Bayesian unit-root test based on the posterior interval of ϕ as well as the test based on the Bayes Factor. We present the results of a power study based on the Bayesian criteria. In Chapter 4, we apply the frequentist and Bayesian unit root tests discussed in this thesis to the four types of close exchange rates from October 1981 to June 1985. In Chapter 5, we compare the frequentist and Bayesian unit root tests by looking at the sum of Type 1 and Type 2 errors. We also include some conclusions and some topics of future research.

Chapter 2

Frequentist Unit-Root Tests in Stochastic Volatility Models

2.1 Introduction

Consider the SVM as defined in the previous chapter;

$$\begin{aligned} r_t &= e^{h_t/2} u_t \\ (h_t - \mu) &= \phi(h_{t-1} - \mu) + \eta_t, \end{aligned} \tag{2.1}$$

for $t = 1, \dots, n$, where $\{\eta_t\}$ is a sequence of uncorrelated random variables having a $N(0, \sigma_\eta^2)$ distribution and $\{u_t\}$ is a sequence of uncorrelated $N(0, 1)$ random variables. $\{u_t\}$ and $\{\eta_t\}$ are assumed to be independent. In this structure, $\{r_t\}$ is the time series of observed data such as mean corrected stock returns whereas $\{h_t\}$ is the time series of the natural logarithm of the unobserved volatility. The conditional variance of the stock return at time t is the volatility at time t which is unobserved. The parameter space consists of μ , ϕ , and σ_η^2 . If the process for h_t has a unit root, random shocks to volatility have effects that remain for a long time. Thus, ϕ is usually called the *persistence parameter* as it can be thought of as a measure of influence (persistence) of the random shocks to volatility.

Note that the return series r_t is nonstationary if the log-volatility process h_t has a unit root. Assume that the log-volatility process h_t has a unit-root. Then h_t is equal to $h_0 + \sum_{i=1}^t \eta_i$. Hence, $\text{Var}[r_t] = \text{E}[e^{h_t}] = \text{E}[e^{h_0 + \sum_{i=1}^t \eta_i}] = \text{E}[e^{h_0}] \text{E}[e^{\sum_{i=1}^t \eta_i}] = \text{E}[e^{h_0}] e^{t\sigma_\eta^2/2}$ assuming that h_0 is independent of $\{\eta_t\}$. That is, the fluctuation in return is exploding over time. As a result of this situation, the business environment becomes much more uncertain. Figures (2.1)-(2.8) display the influence of the persistence parameter ϕ on the mean corrected return series for different sample sizes. These figures are constructed by using data sets simulated from model 2.1. We used two different set of seeds to generate the data. Figures (2.1) to (2.4) are constructed by using one set of seeds, Seed Set #1, whereas the figures from (2.5) to (2.8) are constructed from the second set of seeds, Seed Set #2.

In this research our main concern is to provide new unit-root testing procedures and improve the existing ones to test $H_0 : \phi = 1$ in SVM.

In this chapter, we consider some commonly used frequentist unit-root tests and develop new tests in the context of SVM. We consider Bayesian unit root tests in the next chapter.

In this section, we make use of the ARMA(1,1) structure of the log-squared mean corrected return data and apply commonly used frequentist unit-root tests to SVM. When we square the observations r_t and take the natural logarithm of the squared observations, the resulting series becomes a linear combination of the autoregressive process h_t and error term $\log u_t^2$. That is, $\log r_t^2 = h_t + \log u_t^2$. Adding and subtracting $a = \text{E}[\log u_t^2] \cong -1.27$, a known quantity, on the right hand side of the equation, we get $\log r_t^2 = a + h_t + (\log u_t^2 - a)$. Let $y_t = \log r_t^2$, $\mu^* = a + \mu$, and $e_t = \log u_t^2 - a$. Here $\mu = \text{E}[h_t]$. Also notice $\text{E}[e_t] = \text{E}[\log u_t^2 - \text{E}[\log u_t^2]] = 0$ and $\sigma_e^2 = \text{Var}[e_t] = \text{Var}[\log u_t^2] = \pi^2/2 = 4.93$. See Abramowitz and Stegun (1972) for the variance of a $\log \chi_{(1)}^2$ random variable. Then

$$y_t = \mu^* + (h_t - \mu) + e_t \tag{2.2}$$

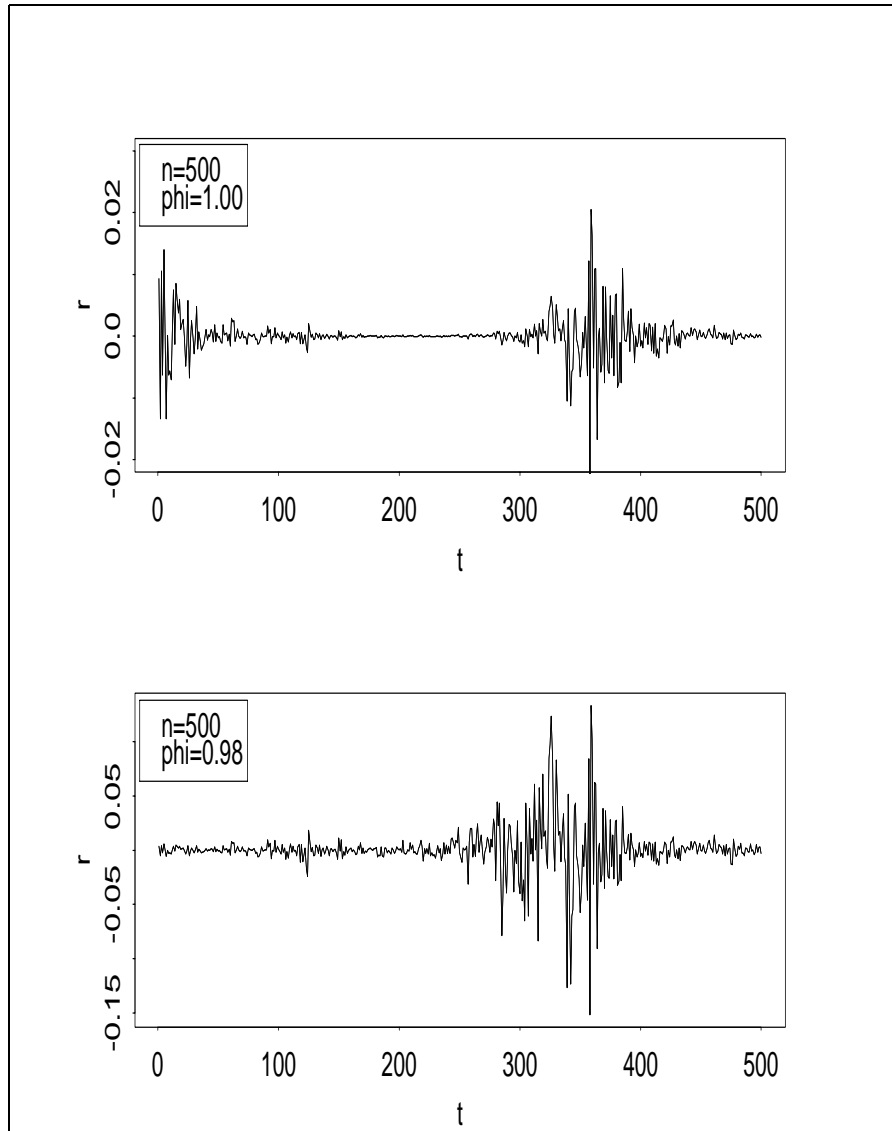


Figure 2.1: Mean Corrected Returns versus Time. *Seed Set #1.*

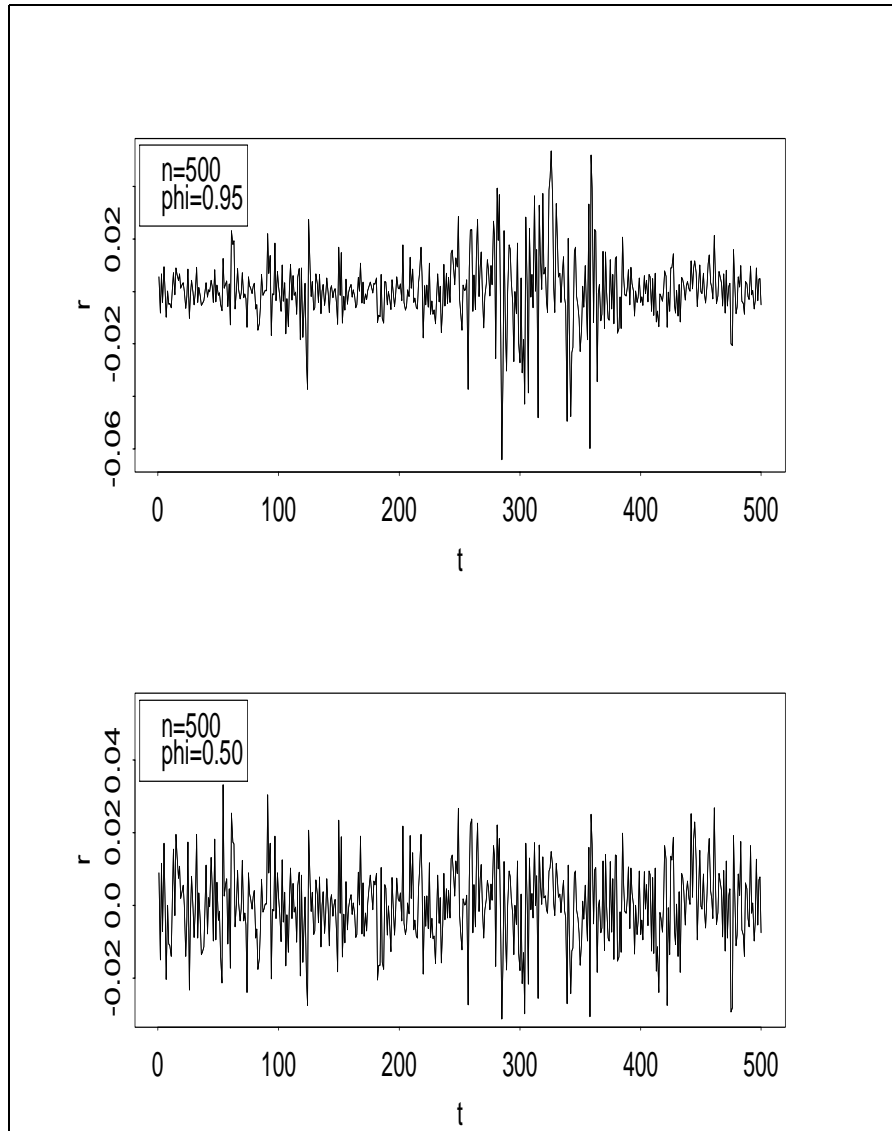


Figure 2.2: Mean Corrected Returns versus Time. *Seed Set #1.*

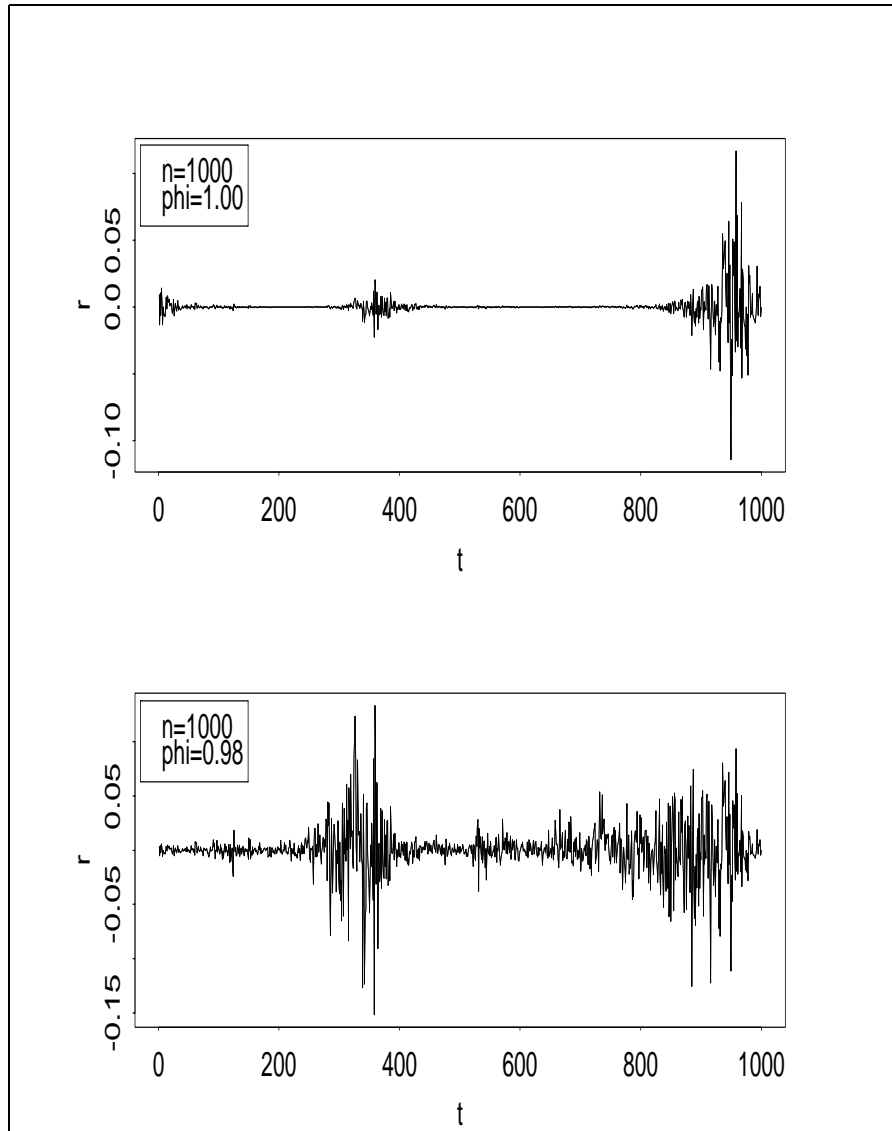


Figure 2.3: Mean Corrected Returns versus Time. *Seed Set #1.*

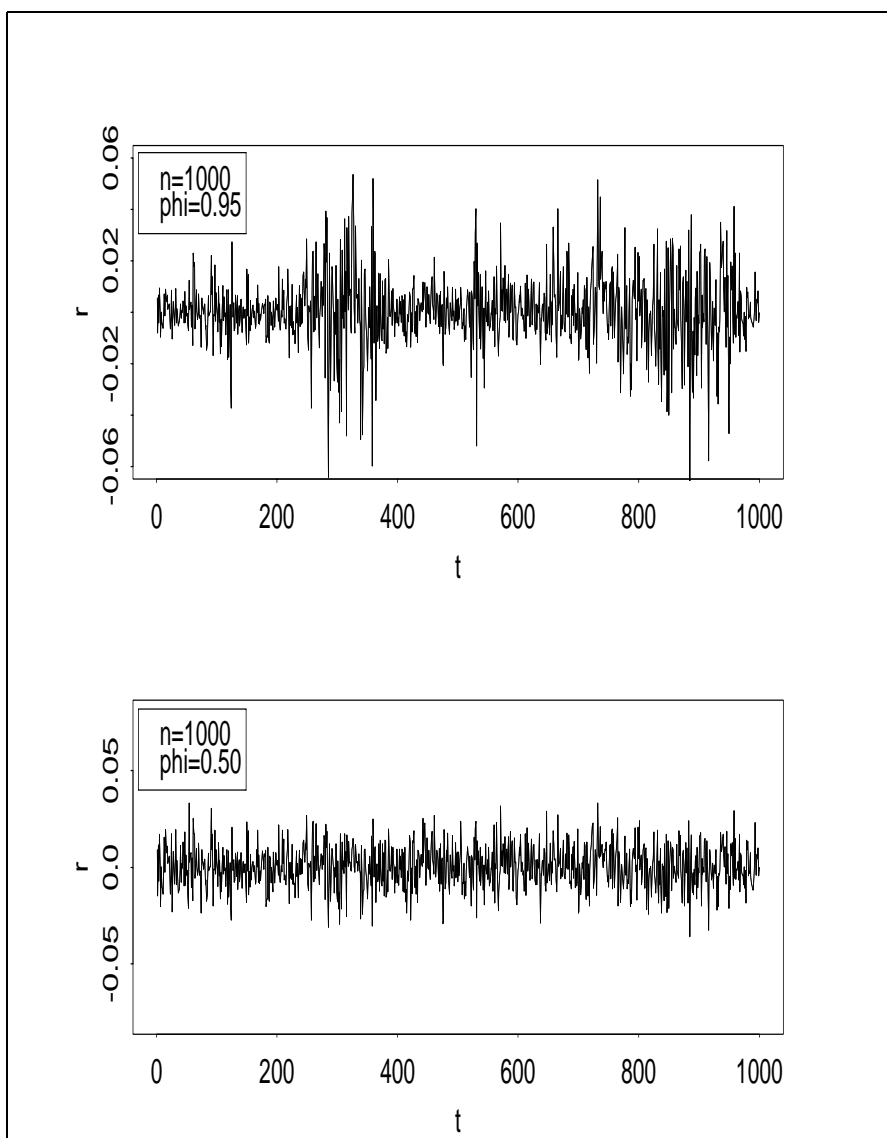


Figure 2.4: Mean Corrected Returns versus Time. *Seed Set #1*.

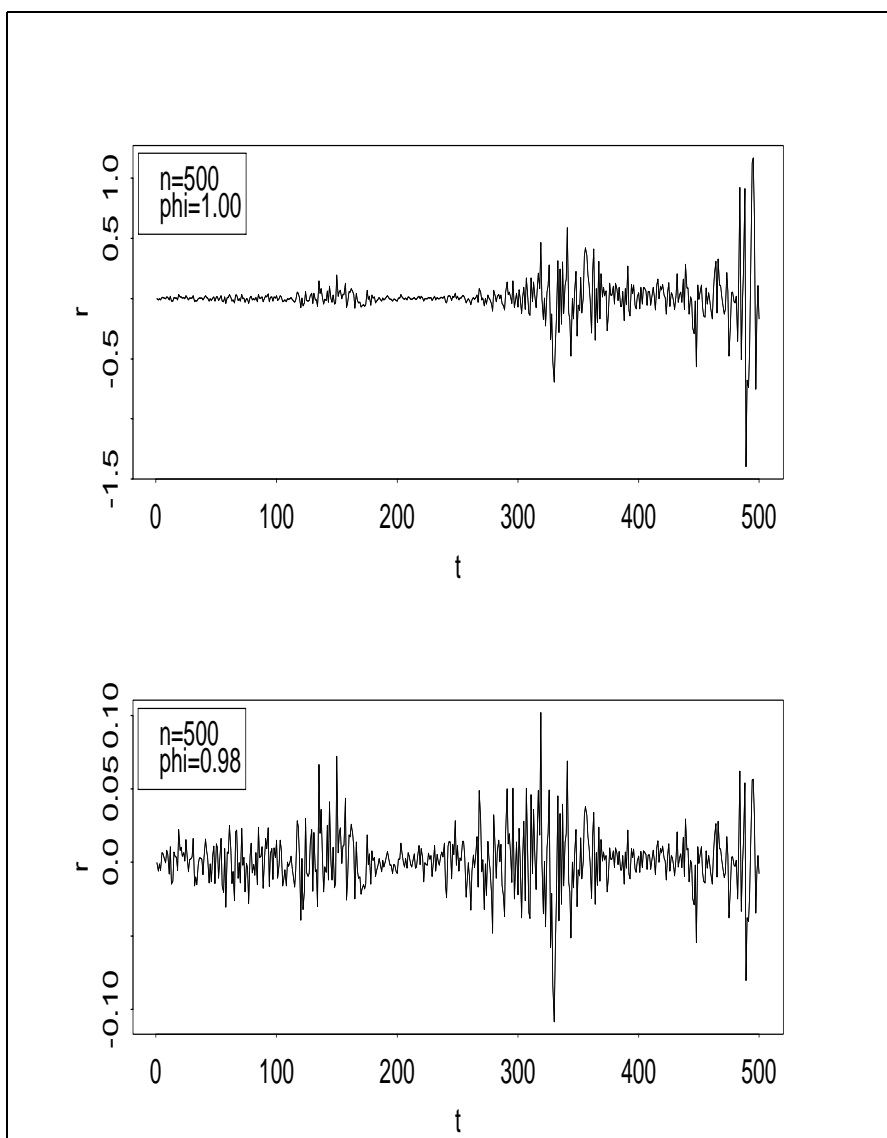


Figure 2.5: Mean Corrected Returns versus Time. *Seed Set #2.*

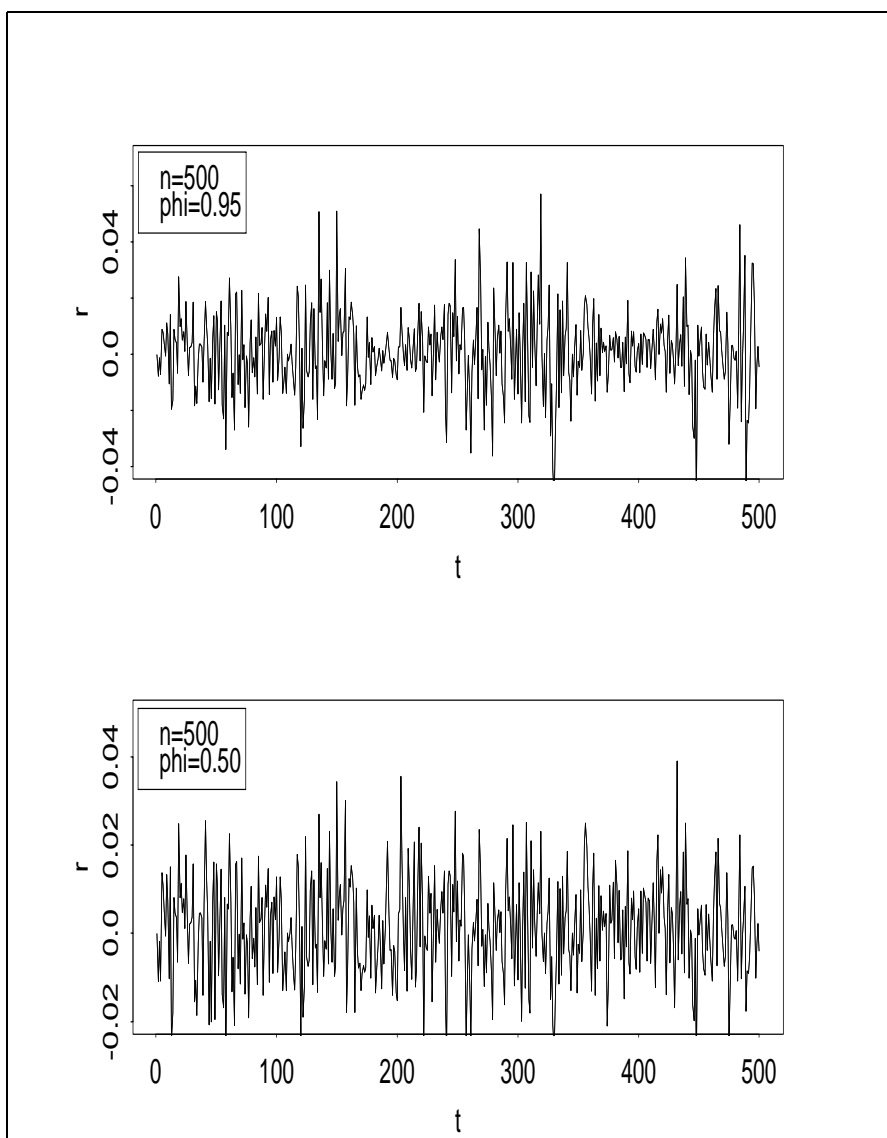


Figure 2.6: Mean Corrected Returns versus Time. *Seed Set #2.*

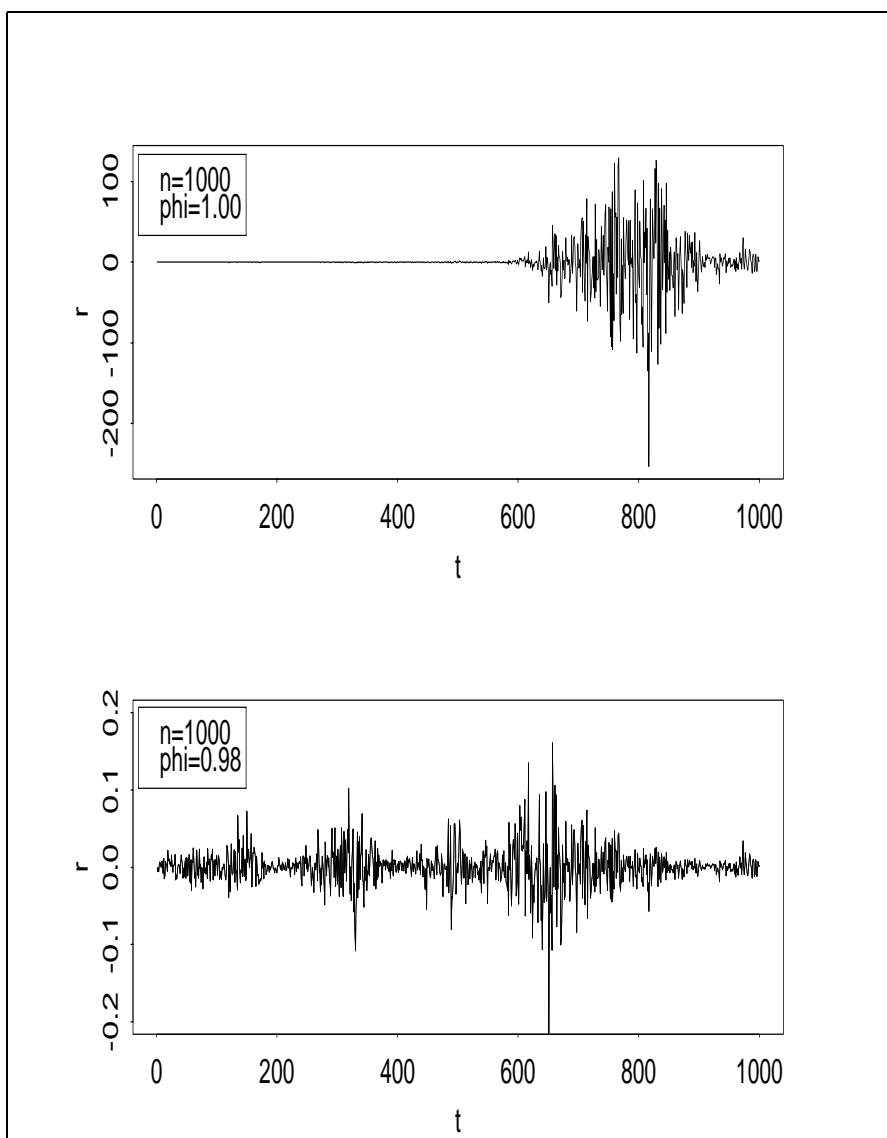


Figure 2.7: Mean Corrected Returns versus Time. *Seed Set #2.*

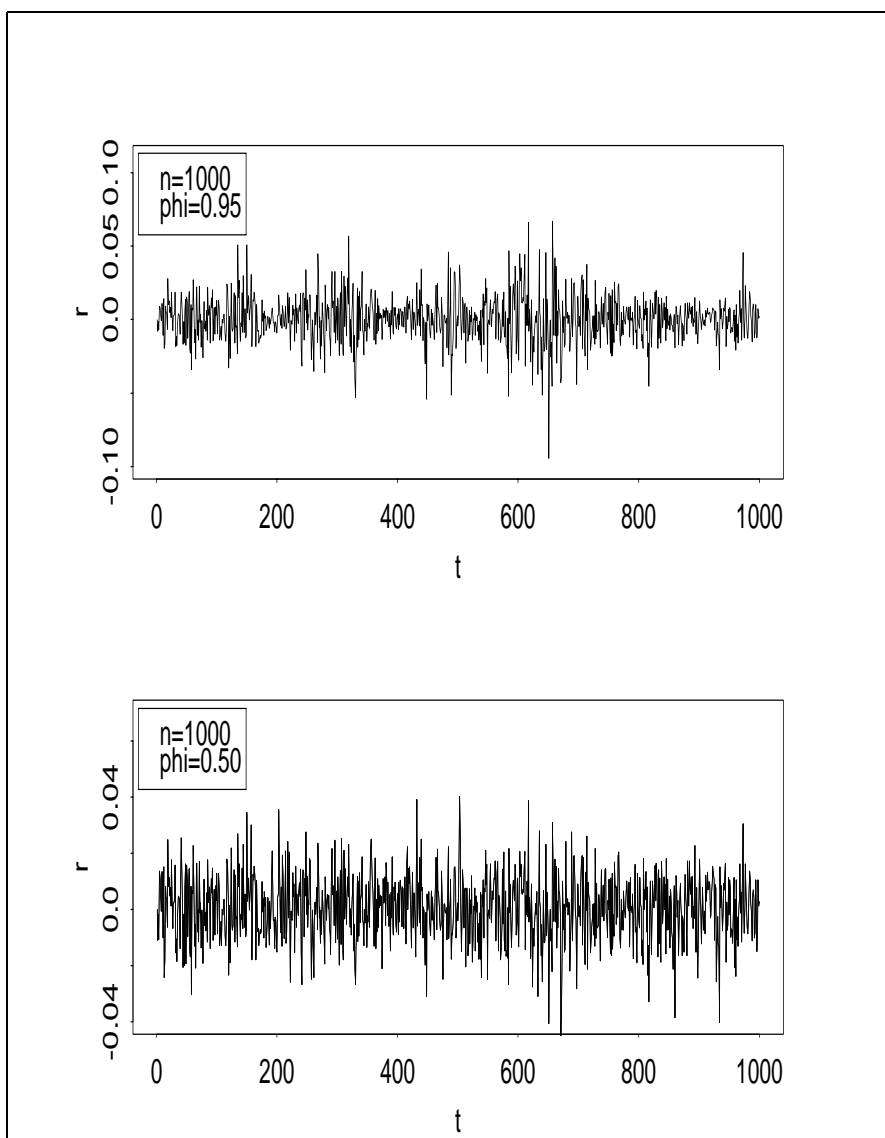


Figure 2.8: Mean Corrected Returns versus Time. *Seed Set #2.*

for $t=1, \dots, n$, where e_t is a white noise sequence with mean 0 and variance $\sigma_e^2 = \pi^2/2$ and $(h_t - \mu)$ is a zero-mean AR(1) process. Notice that $y_{t-1} = \mu^* + (h_{t-1} - \mu) + e_{t-1}$ and multiplying y_{t-1} by ϕ we get $\phi y_{t-1} = \phi\mu^* + \phi(h_{t-1} - \mu) + \phi e_{t-1}$. Subtracting ϕy_{t-1} from y_{t-1} , we get $y_t - \phi y_{t-1} = \mu^*(1 - \phi) + (h_t - \mu) - \phi(h_{t-1} - \mu) + (e_t - \phi e_{t-1})$. The right hand side is in fact same as $\mu^*(1 - \phi) + \eta_t + (e_t - \phi e_{t-1})$ since $(h_t - \mu) - \phi(h_{t-1} - \mu) = \eta_t$. If we let $Z_t = \eta_t + (e_t - \phi e_{t-1})$, then we find the autocovariances of the Z_t process as $\gamma_z(0) = \sigma_\eta^2 + (1 + \phi^2)\sigma_e^2$, $\gamma_z(1) = -\phi\sigma_e^2$, $\gamma_z(k) = 0$ for $k \geq 2$. This autocovariance structure is same as the autocovariance structure of an MA(1) process with error variance

$$\sigma^2 = \frac{(\sigma_\eta^2 + (1 + \phi^2)\sigma_e^2) + \sqrt{(\sigma_\eta^2 + (1 + \phi^2)\sigma_e^2)^2 - 4\phi^2(\sigma_e^2)^2}}{2} \quad (2.3)$$

and moving average parameter

$$\theta = \phi \frac{\sigma_e^2}{\sigma^2} \quad (2.4)$$

That is, $(\eta_t + e_t - \phi e_{t-1})$ has the correlation structure of an MA(1) process with moving average parameter θ and error variance σ^2 where σ^2 and θ are as given in (2.3) and (2.4). Hence it follows that $\log r_t^2$ has the same covariance structure as an ARMA(1,1) process with mean μ^* , AR parameter ϕ , MA parameter θ , and error variance σ^2 . Let us denote the error term in this corresponding ARMA(1,1) process by v_t . That is, this second order equivalent model for the time series y_t has the following structure;

$$(y_t - \mu^*) = \phi(y_{t-1} - \mu^*) + v_t - \theta v_{t-1} \quad (2.5)$$

$t=1, \dots, n$, where $y_0 = \mu^*$, $|\theta| < 1$ and $\{v_t\}$ is a sequence of uncorrelated random errors with mean 0 and variance σ^2 . All we know really is that y_t has the same covariance structure as the ARMA(1,1) process given in (2.5). If e_t and η_t are both Gaussian then there is a Gaussian v_t that satisfies (2.5) and our $y_t = \log r_t^2$ can exactly be represented in the form of (2.5). Otherwise, we don't know whether y_t can be expressed in this form. What we do know is that y_t has same second order properties as the time series in (2.5). Now, recall that our main goal is to figure out if the log-volatility process h_t

has a unit-root. In other words, the goal is to test the hypothesis $H_0 : \phi = 1$ where ϕ is the autoregressive parameter of the h_t process. Since h_t is an unobserved time series, we can not work with this autoregressive process to test the persistence parameter ϕ being 1. However, we showed that time series of the log-squared observations display an ARMA(1,1) structure with the same autoregressive parameter as the log-volatility time series. Therefore, testing for a unit-root in the time series of unobserved log-volatility boils down to testing for a unit-root in time series of observed log-squared return. As seen in Figure (1.1), θ of the corresponding ARMA(1,1) model is close to the AR parameter ϕ . This should be taken into account in order to get reliable results from the application of commonly used unit root procedures in SVM. Therefore, we look at the finite sampling distribution of several frequentist unit root test statistics with respect to different values of θ and use percentiles of those distributions as critical values while performing the tests in SVM.

- We now consider several methods for testing unit roots;
- Augmented Dickey Fuller Tests,
 - One Step Gauss-Newton Method (Said-Dickey Test),
 - Tests Based on Instrumental Variable Estimators.

2.2 Augmented Dickey-Fuller Unit-Root Test in SVM

Said and Dickey (1984) describe the Augmented Dickey-Fuller (ADF) test statistic for an unknown order ARMA model and derive the limiting distribution of the test statistic based on OLS estimators when the true value of the AR parameter is one. They show that the test statistic obtained by ADF approach has the same limiting distribution as the test statistic of Dickey and Fuller (1979). ADF testing procedure has widely been used to test for a unit root. In this section, we apply this testing procedure to $\log r_t^2$ where r_t is generated from model (2.1).

First of all, note that σ^2 , the error variance of the ARMA model that has similar second order properties to $\log r_t^2$, is obtained as a solution to a second order polynomial. This solution is given in equation (2.3). We choose this σ^2 so that $|\theta| < 1$ where $\theta = \phi \frac{\sigma^2}{\sigma^2}$, and the corresponding second order equivalent model to our $y_t = \log r_t^2$ is an invertible process.

In order to apply the ADF procedure to the mean corrected data y_t^* , where $y_t^* = y_t - \mu^*$ with $y_t = \log r_t^2$ where r_t is generated from (2.1), y_t^* is rewritten as an infinite order AR process such that $y_t^* = \sum_{j=1}^{\infty} \theta^{j-1} (\phi - \theta) y_{t-j}^* + v_t$ where $\{v_t\}$ are uncorrelated. This model is equivalent to $D_t = \phi^* y_{t-1}^* + a_1 D_{t-1} + a_2 D_{t-2} + \dots + v_t$ where $D_t = y_t^* - y_{t-1}^*$ and $\phi^* = (\phi - \theta) \sum_{j=1}^{\infty} \theta^{j-1} - 1$. This infinite sum is approximated by a long AR process. $\dot{y}_t = y_t - \bar{y}$ is usually used in practice. In practice, AR models with order 1 through order $[10n^{1/4}]$ are fit to \dot{y}_t and Akaike information criteria (AIC) is calculated for each model. Let p_{AIC} be the order picked by AIC. The final order of the model is chosen to be the the minimum of $p_{AIC} + 2$ and $[10n^{1/4}]$ where $[.]$ is the largest integer less than or equal to the argument. Pantula et al. (1994) approximate the ARMA model by an AR process whose order is equal to the order selected by the AIC plus 2. The reason for this is that, according to their simulations, this choice of order seems to maintain the level of the test better. In our study, we also use what we

call the maximum order ADF regression. That is, we take the order of the AR model to be $[10n^{1/4}]$. We use the terms maximum and full interchangeably for such fits. The ADF approach fits the model $D_t = \phi^* \dot{y}_{t-1} + a_1 D_{t-1} + \dots + a_{p-1} D_{t-p+1} + v_{tp}$ and test $\phi^* = 0$. We use two separate procedures to estimate the model parameters. The first one is obtained by minimizing $\sum_{t=p+1}^n (D_t - \phi^* \dot{y}_{t-1} - a_1 D_{t-1} - \dots - a_{p-1} D_{t-p+1})^2$. In other words we regress D_t on $\dot{y}_{t-1}, D_{t-1}, \dots, D_{t-p+1}$ and find the ordinary least square (OLS) estimates. Let $X' = [X'_1, X'_2 \dots X'_n]$, $X_t = [\dot{y}_{t-1}, D_{t-1}, D_{t-2}, \dots, D_{t-p+1}]$, $t = p + 1, \dots, n$. Then the corresponding ADF test statistic based on the ordinary least squares estimators in this regression is

$$t_{\mu,(ADF/OLS)} = \frac{\hat{\phi}_{ols}^*}{\sqrt{\hat{\sigma}_{ols}^2 c_{1,ols}}} \quad (2.6)$$

where $\hat{\phi}_{ols}^*$ is ordinary least squares estimator of ϕ^* , $c_{1,ols}$ is the element of $(X'X)^{-1}$ associated with ϕ^* , $\hat{\sigma}_{ols}^2$ is the error mean square, MSE, from our regression.

Another technique to estimate the model parameters of ADF regression is weighted-symmetric (WS) estimation. In this technique, estimates of the parameters are obtained by minimizing $\sum_{t=p+1}^n w_t [D_t - \phi^* \dot{y}_{t-1} - \sum_{i=1}^{p-1} a_i D_{t-i}]^2 + \sum_{t=1}^{n-p} (1 - w_{t+1}) [D_t + \phi^* \dot{y}_t - \sum_{i=1}^{p-1} a_i D_{t+i}]^2$

In our study, we take the order p to be $[10n^{1/4}]$ when we use ADF regression along with weighted-symmetric estimation procedure. We used the same weights as in Fuller (1996). Namely, w_t is 0 for $t=1,2,\dots,p$, $(n - 2p + 2)^{-1}(t - p)$ for $t=p+1,p+2,\dots,n-p+1$, and 1 for $t=n-p+2,n-p+3,\dots,n$. The table of corresponding dependent and independent vectors for WS regression is given in Table 2.1

Test statistic based on WS estimators is

$$t_{\mu,(ADF/WS)} = \frac{\hat{\phi}_{ws}^*}{\sqrt{\hat{\sigma}_{ws}^2 c_{1,ws}}} \quad (2.7)$$

where $\hat{\phi}_{ws}^*$ is WS estimator of ϕ^* , X being the matrix consisting of independent vectors as defined in Table 2.1, $c_{1,ws}$ is the element of $(X'X)^{-1}$ associated with ϕ^* , and $\hat{\sigma}_{ws}^2$ is the error mean square, MSE, from our WS regression. Asymptotically, $t_{\mu,(ADF/OLS)}$

Table 2.1: Data Arrangement of WS Regression

Weight	Dependent Variable	Parameters				
		ϕ^*	a_1	a_2	...	a_{p-1}
w_{p+1}	D_{p+1}	\dot{y}_p	D_p	D_{p-1}	...	D_2
w_{p+2}	D_{p+2}	\dot{y}_{p+1}	D_{p+1}	D_p	...	D_3
.
.
.
w_n	D_n	\dot{y}_{n-1}	D_{n-1}	D_{n-2}	...	D_{n-p+1}
$1 - w_{n-p+1}$	D_{n-p}	$-\dot{y}_{n-p}$	D_{n-p+1}	D_{n-p+2}	...	D_{n-1}
$1 - w_{n-p}$	D_{n-p-1}	$-\dot{y}_{n-p-1}$	D_{n-p}	D_{n-p+1}	...	D_{n-2}
.
.
.
$1 - w_2$	D_1	$-\dot{y}_1$	D_2	D_3	...	D_p

and $t_{\mu,(ADF/WS)}$ have the same distributions as the Dickey-Fuller test statistic based on OLS and WS respectively regardless of the value of the MA parameter θ . However, finite sampling distributions of these test statistics are affected by the true value of the MA parameter θ especially when it is close to one. Note that in SVM setup, the MA parameter of the second order equivalent model of $y_t = \log r_t^2$, as formulated in equation (2.4), is close to 1 when ϕ is close to 1. That is the motivation for the finite sampling study that we present next.

Now, let the level of the test be denoted by α and let $\tau_{n,\theta,\alpha}$ be the corresponding empirical percentile computed for a given value of θ and n by using Monte Carlo simulations. Then a unit root test criterion in the SVM is to reject $H_0 : \phi = 1$, if

$$t_{\mu,(ADF/.)} < \tau_{\mu,ADF/.,n,\theta} \quad (2.8)$$

where $\tau_{\mu,ADF/.,n,\theta}$ is the cutoff value obtained from the empirical percentiles of $\tau_{\mu,ADF/.$ that changes with respect to the sample size and the value of θ . How to compute $\tau_{\mu,ADF/.$ is explained below. In (2.8), $\{.\}$ stands for OLS if one uses ADF test with OLS estimators, for WS if one uses ADF test with WS estimators.

We use an ARMA(1,1) model to compute the empirical percentiles of $\tau_{\mu,ADF/}$ with respect to the sample size, n , and the value of θ . This ARMA(1,1) model is

$$(y_t - \mu^*) = \phi(y_{t-1} - \mu^*) + v_t - \theta v_{t-1}, \quad t \geq 1 \quad (2.9)$$

where $v_t \sim NID(0, \sigma^2)$ for $t \geq 0$, $h_1 \sim N(\mu^*, \sigma^2)$, $\mu^* = 0$, $\sigma^2 = 1$ and $\phi = 1$. For a given sample size n and MA parameter θ , we generated 10,000 samples and computed the test statistics $t_{\mu,(ADF/OLS)}$ and $t_{\mu,(ADF/WS)}$ by fitting an ADF model to the data generated from (2.9).

We compute the empirical percentiles, $\tau_{\mu,ADF/OLS}$, under two different cases one of which is the minimum order ADF model and the other one is the maximum (full) order ADF model. Here, maximum (full) order ADF regression consists of $p = \lceil 10n^{1/4} \rceil$ regressors and minimum order ADF regression consists of $\min(p_{AIC} + 2, \lceil 10n^{1/4} \rceil)$ regressors. Note that, when $n = 100$, $\lceil 10n^{1/4} \rceil = 31$. Similarly when $n = 500$ and $n = 1000$, it is 47 and 56 respectively. We compute the empirical percentiles of the test statistic $\tau_{\mu,ADF/WS}$ for only full ADF regression because choosing the order of the model according to AIC was very time consuming in the WS setup. In this part of this section, tables that exhibit the empirical finite sampling distribution of $\tau_{\mu,ADF/OLS,n,\theta}$ and $\tau_{\mu,ADF/WS,n,\theta}$ are presented.

Table 2.2: Percentiles of Dickey-Fuller Pivotal Test Statistic. *Reported from second parts of Table 10.A.2. and Table 10.A.4 of Fuller(1996)*

	n	100	500	1000
$t_{\mu,(DF/OLS)}, \alpha =$	0.05	-2.90	-2.87	-2.86
	0.10	-2.59	-2.57	-2.57
$t_{\mu,(DF/WS)}, \alpha =$	0.05	-2.55	-2.53	-2.52
	0.10	-2.24	-2.23	-2.22

In all of the following tables we see that, the empirical percentiles of $\tau_{\mu,ADF/OLS,n,\theta}$ and $\tau_{\mu,ADF/WS,n,\theta}$ test statistics are larger in magnitude than empirical percentiles of corresponding Dickey-Fuller test statistic if the true value of θ is close to 1 (the value

Table 2.3: Empirical 5th Percentiles for $\tau_{\mu,ADF/OLS,n,\theta}$. Order of the Augmented Dickey-Fuller model fit is $p_{min} = \min(p_{AIC} + 2, [10n^{1/4}])$

	n	100	500	1000
$\theta =$	0.00	-2.85	-2.88	-2.87
	0.10	-2.85	-2.88	-2.88
	0.20	-2.86	-2.88	-2.88
	0.30	-2.89	-2.88	-2.90
	0.40	-2.97	-2.91	-2.90
	0.50	-3.10	-2.95	-2.93
	0.60	-3.43	-3.05	-2.99
	0.70	-4.07	-3.23	-3.09
	0.80	-5.06	-3.73	-3.38
	0.85	-5.63	-4.39	-3.73
	0.90	-6.20	-6.46	-4.65
	0.95	-6.73	-12.29	-10.36
	0.99	-7.00	-13.91	-19.00

Table 2.4: Empirical 10th Percentiles for $\tau_{\mu,ADF/OLS,n,\theta}$. Order of the Augmented Dickey-Fuller model fit is $p_{min} = \min(p_{AIC} + 2, [10n^{1/4}])$

	n	100	500	1000
$\theta =$	0.00	-2.53	-2.59	-2.58
	0.10	-2.53	-2.59	-2.58
	0.20	-2.55	-2.60	-2.58
	0.30	-2.58	-2.60	-2.59
	0.40	-2.63	-2.62	-2.60
	0.50	-2.75	-2.65	-2.62
	0.60	-2.99	-2.73	-2.67
	0.70	-3.59	-2.88	-2.77
	0.80	-4.66	-3.28	-3.02
	0.85	-5.25	-3.84	-3.32
	0.90	-5.85	-5.32	-4.07
	0.95	-6.43	-11.83	-7.94
	0.99	-6.71	-13.65	-18.73

of ϕ under the null hypothesis). This implies that, if θ is close to ϕ under the null, it takes a larger sample for $\tau_{\mu,ADF/OLS,n,\theta}$ and $\tau_{\mu,ADF/WS,n,\theta}$ to converge to the limiting distribution of corresponding Dickey-Fuller test statistic. Obtaining this convergence

Table 2.5: Empirical 5th Percentiles for $\tau_{\mu,ADF/OLS,n,\theta}$. Order of the Augmented Dickey-Fuller model fit is $p_{full} = \lceil 10n^{1/4} \rceil$

	n	100	500	1000
$\theta =$	0.00	-2.02	-2.67	-2.76
	0.10	-2.01	-2.67	-2.76
	0.20	-2.01	-2.67	-2.76
	0.30	-2.01	-2.67	-2.77
	0.40	-2.02	-2.67	-2.77
	0.50	-2.04	-2.69	-2.77
	0.60	-2.09	-2.71	-2.79
	0.70	-2.19	-2.75	-2.81
	0.80	-2.42	-2.89	-2.86
	0.85	-2.64	-3.05	-2.96
	0.90	-2.94	-3.46	-3.23
	0.95	-3.26	-4.56	-4.24
0.99	-3.46	-6.28	-7.23	

Table 2.6: Empirical 10th Percentiles for $\tau_{\mu,ADF/OLS,n,\theta}$. Order of the Augmented Dickey-Fuller model fit is $p_{full} = \lceil 10n^{1/4} \rceil$

	n	100	500	1000
$\theta =$	0.00	-1.80	-2.40	-2.48
	0.10	-1.80	-2.40	-2.48
	0.20	-1.80	-2.40	-2.48
	0.30	-1.81	-2.40	-2.49
	0.40	-1.82	-2.41	-2.49
	0.50	-1.84	-2.42	-2.49
	0.60	-1.87	-2.43	-2.49
	0.70	-1.94	-2.45	-2.50
	0.80	-2.13	-2.54	-2.56
	0.85	-2.32	-2.66	-2.62
	0.90	-2.60	-2.97	-2.79
	0.95	-2.93	-3.95	-3.62
0.99	-3.13	-5.71	-6.52	

needs even a larger sample, if ARMA(1,1) model is approximated by a minimum order AR model. Comparing the empirical percentiles of $\tau_{\mu,ADF/OLS,n,\theta}$ obtained from the maximum order model for the small values of θ , we recognize that the empirical

Table 2.7: Empirical 5th Percentiles for $\tau_{\mu,ADF/WS,n,\theta}$. Order of the Augmented Dickey-Fuller model fit is $p_{full} = \lceil 10n^{1/4} \rceil$

	n	100	500	1000
$\theta =$	0.00	-2.73	-2.54	-2.56
	0.10	-2.72	-2.54	-2.55
	0.20	-2.71	-2.54	-2.56
	0.30	-2.70	-2.55	-2.55
	0.40	-2.70	-2.55	-2.55
	0.50	-2.70	-2.54	-2.55
	0.60	-2.72	-2.55	-2.55
	0.70	-2.70	-2.54	-2.55
	0.80	-2.67	-2.54	-2.55
	0.85	-2.67	-2.54	-2.55
	0.90	-2.69	-2.57	-2.57
	0.95	-2.89	-2.89	-2.78
	0.99	-3.09	-4.10	-4.62

Table 2.8: Empirical 10th Percentiles for $\tau_{\mu,ADF/WS,n,\theta}$. Order of the Augmented Dickey-Fuller model fit is $p_{full} = \lceil 10n^{1/4} \rceil$

	n	100	500	1000
$\theta =$	0.00	-2.37	-2.26	-2.24
	0.10	-2.37	-2.27	-2.24
	0.20	-2.37	-2.27	-2.24
	0.30	-2.37	-2.27	-2.24
	0.40	-2.36	-2.27	-2.24
	0.50	-2.36	-2.27	-2.24
	0.60	-2.35	-2.26	-2.24
	0.70	-2.35	-2.26	-2.25
	0.80	-2.35	-2.25	-2.24
	0.85	-2.34	-2.25	-2.25
	0.90	-2.38	-2.28	-2.26
	0.95	-2.57	-2.58	-2.48
	0.99	-2.79	-3.84	-4.34

percentiles from the full model are not close to the percentiles of Dickey-Fuller test statistic. This is probably because of the extensive number of nuisance parameters estimated in a full order fit.

The percentiles of the test statistic based on the weighted symmetric estimators in a full order ADF model is influenced only by the true values of θ which are very close to 1. The empirical percentiles of this test statistic are close to the asymptotic percentiles of Dickey-Fuller weighted symmetric test statistic even when the true value of θ is 0.6, 0.7, 0.8, or 0.85.

2.3 Said-Dickey (One Step Gauss-Newton) Unit-Root Test in SVM

Said and Dickey (1985) introduce a unit-root test in general ARMA models with known orders p and q . They derive the limiting distribution of the test statistic and show that it converges in distribution to the limiting distribution of Dickey-Fuller test statistic. In this section, we apply this testing procedure to $\log r_t^2$ by making use of the fact that our process has second order properties of an ARMA(1,1) model. First of all, we reexpress the simple ARMA(1,1) model as $e_t = y_t - \phi y_{t-1} + \theta e_{t-1}$. Then treating the initial condition on the random error term at lag 0 as a parameter, namely $e_0 = \delta$, we have the parameter vector $\beta' = (\phi, \theta, \delta)$. The reexpressed model can be written as $e_t(\beta) = y_t - \sum_{j=0}^{t-2} \theta^j (\phi - \theta) y_{t-1-j} + \theta^t \delta$ for $t \geq 1$. The first order Taylor's series expansion of $e_t(\beta)$ about an initial estimate $\hat{\beta}' = (\hat{\phi}, \hat{\theta}, \hat{\delta})$ is $e_t(\beta) = e_t(\hat{\beta}) - V_t(\hat{\beta})(\phi - \hat{\phi}) - W_t(\hat{\beta})(\theta - \hat{\theta}) - \Delta_t(\hat{\beta})(\delta - \hat{\delta}) + R_t$ where $-V_t(\hat{\beta})$, $-W_t(\hat{\beta})$, and $-\Delta_t(\hat{\beta})$ are the first derivatives of $e_t(\beta)$ with respect to ϕ , θ , and δ respectively evaluated at $\hat{\beta}$, and R_t is the remainder term. Ignoring the remainder term and rearranging this equation, one can get $e_t(\hat{\beta}) = V_t(\hat{\beta})(\phi - \hat{\phi}) + W_t(\hat{\beta})(\theta - \hat{\theta}) + \Delta_t(\hat{\beta})(\delta - \hat{\delta}) + e_t(\beta)$. Nonlinear least squares regression estimates of the parameters and the variance of the estimators are obtained by regressing $e_t(\hat{\beta})$ on $V_t(\hat{\beta})$, $W_t(\hat{\beta})$, and $\Delta_t(\hat{\beta})$. We perform one step iteration. We take 1 and 0 to be the initial values for $\hat{\phi}$ and $\hat{\delta}$ respectively. We use Durbin's estimate as an initial value for θ . In order to compute Durbin's estimate, we rewrite model (2.5) as $D_t = -(c_0 + 1)y_{t-1} + \sum_{j=1}^{t-1} c_j D_{t-j} + e_t - \theta^t e_0$ where

$-c_0 = (1 - \theta)^{-1}(\phi - \theta) = (1 - \theta)^{-1}(\phi - 1) + 1$ and $c_j = -(\theta)^j(1 - \theta)^{-1}(\phi - \theta) = \theta c_{j-1}$ for $j=1, \dots, (t-1)$. This is the same parameterization that Said and Dickey (1985) used. We truncate the sum in that representation at k . Estimates, $\hat{c}_0, \hat{c}_1, \dots, \hat{c}_k$ are obtained by regressing D_t on $y_{t-1}, D_{t-1}, \dots, D_{t-k}$. Then, θ is estimated by regressing \hat{c}_j on \hat{c}_{j-1} . Said-Dickey unit-root test statistic is

$$t_{\mu, SDD} = \frac{\hat{\phi}_{sdd} - 1}{\sqrt{S_{sdd}^2 c_{1, sdd}}} \quad (2.10)$$

where $\hat{\phi}_{sdd} = \gamma_1 + 1$ where γ_1 is the OLS estimate of the coefficient of $V_t(\hat{\beta})$ in the last iteration, $c_{1, sdd}$ is the entry of $[F^T(\hat{\beta})F(\hat{\beta})]^{-1}$ associated with ϕ where $F^T(\hat{\beta}) = [f_1^T(\hat{\beta}) f_2^T(\hat{\beta}) \dots f_n^T(\hat{\beta})]$ with $f_t(\hat{\beta}) = [-V_t(\hat{\beta}), -W_t(\hat{\beta}), -\Delta(\hat{\beta})]$ for $t = 1, 2, \dots, n$, and $S_{sdd}^2 = \frac{1}{n} \sum_{t=1}^n e_t^2(\hat{\beta})$. The test criterion for the unit-root in the log-squared mean-corrected return is to reject the null hypothesis if

$$t_{\mu, SDD} < \tau_{\mu, SDD, n, \theta} \quad (2.11)$$

where $\tau_{\mu, SDD, n, \theta}$ is the critical value that is determined with respect to n and a consistent estimate of θ .

In Table 2.9 and Table 2.10, we display the empirical 5th and 10th percentiles of $\tau_{\mu, SDD}$ with respect to different sample sizes as well as different values of θ . In order to generate the data to compute these empirical percentiles, we use the model (2.9). We generated 10,000 samples to study the finite sampling distribution of $\tau_{\mu, SDD}$ for various values of n and θ . By looking at the empirical results of Table 2.9 and Table 2.10, we can say that, if the true value of θ is close to ϕ under the null, large samples are needed to have the distribution of $\tau_{\mu, SDD, n, \theta}$ approximated by the limiting distribution of the DF test statistic. For $\theta=0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, \text{ or } 0.8$, this test statistic shows similar limiting properties as Dickey-Fuller test statistic when the sample size is $n=1000$.

Table 2.9: Empirical 5th Percentiles for $\tau_{\mu,SDD,n,\theta}$.

	n	100	500	1000
$\theta =$	0.00	-2.75	-2.73	-2.74
	0.10	-2.75	-2.83	-2.86
	0.20	-2.81	-2.87	-2.88
	0.30	-2.86	-2.87	-2.88
	0.40	-2.89	-2.87	-2.88
	0.50	-2.93	-2.88	-2.88
	0.60	-2.99	-2.87	-2.88
	0.70	-3.12	-2.88	-2.86
	0.80	-3.57	-2.92	-2.88
	0.85	-4.15	-3.07	-3.01
	0.90	-5.15	-3.64	-3.50
	0.95	-6.07	-6.09	-5.39
	0.99	-6.37	-11.95	-15.40

Table 2.10: Empirical 10th Percentiles for $\tau_{\mu,SDD,n,\theta}$.

	n	100	500	1000
$\theta =$	0.00	-2.43	-2.45	-2.44
	0.10	-2.45	-2.56	-2.57
	0.20	-2.51	-2.60	-2.58
	0.30	-2.56	-2.60	-2.58
	0.40	-2.59	-2.60	-2.58
	0.50	-2.61	-2.59	-2.58
	0.60	-2.64	-2.59	-2.58
	0.70	-2.73	-2.59	-2.57
	0.80	-3.03	-2.62	-2.58
	0.85	-3.43	-2.75	-2.66
	0.90	-4.20	-3.21	-3.11
	0.95	-5.05	-5.04	-4.77
	0.99	-5.40	-10.11	-12.72

2.4 Instrumental Variable Approach for a Unit-Root Test in SVM

Let $\{y_t\}$ be a time series generated by the model

$$\begin{aligned}(y_t - \mu^*) &= \phi(y_{t-1} - \mu^*) + \epsilon_t \\ \epsilon_t &= v_t - \theta v_{t-1} \quad t=1, \dots, n\end{aligned}\tag{2.12}$$

where y_0 has a fixed distribution independent of $\{v_t\}$, $\phi = 1$, and $\{v_t\}$ is independently and identically distributed with 0 mean and a finite variance. Also, $E[|\epsilon_t|^{\beta+\delta}] < \infty$ for some $\beta > 2$ and $\delta > 0$.

Here, the variable y_{t-1} is endogenous meaning that it is correlated with ϵ_t . In this case, OLS estimator of ϕ that is obtained by regressing y_t on y_{t-1} is not a consistent estimator of ϕ . Getting an OLS which is consistent requires variables which are known as instruments. A valid instrument, say y_{t-k} , must be correlated with y_{t-1} and uncorrelated with ϵ_t . Therefore, k is to be chosen in such a way that y_{t-k} is a valid instrument. In the model above, y_{t-2} appears to be a valid instrument. Let $\hat{\phi}_{IV}$ be an instrumental variable estimator of ϕ and also let $Y_t = y_t - \bar{y}$. $\hat{\phi}_{IV}$ is obtained as follows; Y_{t-1} can be written in terms of Y_{t-2} plus an error term, i.e., $Y_{t-1} = \gamma Y_{t-2} + w_t$. $\hat{\gamma}$ is obtained by regressing Y_{t-1} on Y_{t-2} and it is $\hat{\gamma} = \frac{\sum Y_{t-2} Y_{t-1}}{\sum Y_{t-2}^2}$. Hence fitted values of Y_{t-1} , i.e. \hat{Y}_{t-1} , are obtained as $\hat{Y}_{t-1} = \hat{\gamma} Y_{t-2} = \frac{\sum Y_{t-2} Y_{t-1}}{\sum Y_{t-2}^2} Y_{t-2}$. $\hat{\phi}_{IV}$ is obtained by regressing Y_t on \hat{Y}_{t-1} , that is, $\hat{\phi}_{IV} = \frac{\sum \hat{Y}_{t-1} Y_t}{\sum \hat{Y}_{t-1}^2} = \frac{\sum Y_{t-2} Y_t}{\sum Y_{t-2} Y_{t-1}}$.

Hall (1989) derives the limiting distributions of $n(\hat{\phi}_{IV} - 1)$ and instrumental variable t statistic multiplied by a correction factor (corrected t statistic) when the data are generated from an ARIMA(0,1,q). He looks at the limiting distribution of the model with no mean, with nonzero mean, as well as nonzero mean and a linear time trend. Limiting distribution of these test statistics for each of these models are same as the limiting distribution of Dickey-Fuller test statistic for the corresponding models.

In this section, we derive the limiting distribution of the test statistics $n(\hat{\phi}_{IV} - 1)$ when the true model for y_t is

$$\begin{aligned} y_t &= h_t + e_t \\ (h_t - \mu) &= \phi(h_{t-1} - \mu) + \eta_t \end{aligned} \quad (2.13)$$

for $t=1, \dots, n$ with the assumptions that $\phi = 1$, $h_0 = \mu$, and $e_t = \log u_t^2$ where u_t is a sequence of i.i.d. $N(0,1)$ random variables and η_t is a sequence of i.i.d. $N(0, \sigma_\eta^2)$ random variables. u_t and η_t are assumed to be independent.

We also derive the correction factor by which the instrumental variable t statistic, denoted by t_{IV} hereafter, is to be multiplied in order to converge to the limiting distribution of Dickey-Fuller test statistic under (2.13).

In addition, we look at the finite sampling distribution of the instrumental variable test statistics discussed here, computed from an ordinary ARMA(1,1) model with $\phi = 1$ so that we can view the behavior of these test statistics in relation to the value of the MA parameter θ . Then, we establish an instrumental variable unit root test criteria in SVM based on these percentiles.

Lemma 2.1. Let $\{\nu_t\}$ be a sequence of independent random variables. Assume that

$$\begin{aligned} \mathbb{E}[\nu_t] &= 0 \\ \mathbb{E}[\nu_t^2] &= \sigma_\nu^2 < \infty \text{ and} \\ \mathbb{E}[|\nu_t|^{2+\delta}] &< M < \infty \text{ for some } \delta > 0 \end{aligned}$$

Then,

$$\begin{aligned} &\left(n^{-1/2} \sum_{i=1}^n \nu_i, n^{-3/2} \sum_{t=2}^n \sum_{i=1}^{t-1} \nu_i, n^{-2} \sum_{t=2}^n \left(\sum_{i=1}^{t-1} \nu_i \right)^2 \right) \\ &\implies \left(\sigma_\nu W(1), \sigma_\nu \int_0^1 W(t) dt, \sigma_\nu^2 \int_0^1 W^2(t) dt \right) \end{aligned}$$

where $W(t)$ is a standard Brownian motion and \implies denotes convergence in distribution.

Proof: The proof is given in Fuller (1996).

Lemma 2.2. Let $\{e_t\}$ be a sequence of i.i.d. random variables with $E[e_t] = a$ and $E[e_t^2] = \sigma_e^2 + a^2 < \infty$ for all t . Also, let $\{\eta_t\}$ be a sequence of i.i.d. random variables with $E[\eta_t] = 0$ and $E[\eta_t^2] = \sigma_\eta^2 < \infty$ for all t . In addition, let $\text{Cov}(\eta_t, e_s) = 0 \forall t$ and s . Below, $e_t^* = e_t - a$. Then,

$$\begin{aligned}
a) \quad & \sum_{t=2}^n \eta_t \eta_{t-1} = O_p(n^{1/2}) \\
b) \quad & \sum_{t=3}^n e_{t-2}^* \eta_{t-1} = O_p(n^{1/2}) \\
c) \quad & \sum_{t=3}^n e_{t-2}^* e_{t-j}^* = O_p(n^{1/2}) \text{ for } \forall j \neq 2 \\
d) \quad & \sum_{t=3}^n e_t^* = O_p(n^{1/2}) \\
e) \quad & e_n^* - e_j^* = O_p(1) \forall j = 1, \dots, (n-1) \\
f) \quad & \sum_{t=1}^n e_{t-2}^{*2} = O_p(n) \\
g) \quad & \sum_{t=3}^n \left(\sum_{j=1}^{t-2} \eta_j \right) e_{t-2}^* = O_p(n^{3/2}) \\
h) \quad & \sum_{t=3}^n \sum_{j=1}^{t-2} \eta_j (e_t^* - e_{t-1}^*) = O_p(n^{1/2})
\end{aligned}$$

Proof:

a) Note that $E[\sum_{t=2}^n \eta_t \eta_{t-1}] = 0$. Also, note that,

$$\begin{aligned}
\text{Var} \left[\sum_{t=2}^n \eta_t \eta_{t-1} \right] &= E \left[\left(\sum_{t=2}^n \eta_t \eta_{t-1} \right)^2 \right] \\
&= \sum_{t=2}^n E[\eta_t^2 \eta_{t-1}^2] + 2 \sum_{t=2}^{n-1} \sum_{i=1}^{n-t} E[\eta_t \eta_{t-1} \eta_{t+i} \eta_{t-1+i}] \\
&= (n-1) (\sigma_\eta^2)^2
\end{aligned}$$

since random variables η_t are i.i.d. $(0, \sigma_\eta^2)$ random variables. Then, $\frac{1}{n} \text{Var}[\sum_{t=2}^n \eta_t \eta_{t-1}] = \frac{n-1}{n} (\sigma_\eta^2)^2 \rightarrow (\sigma_\eta^2)^2$. So, $\text{Var}[\sum_{t=2}^n \eta_t \eta_{t-1}] = O(n)$. As a result, $\sum_{t=2}^n \eta_t \eta_{t-1} = O_p(n^{1/2})$.

b) Note that, $E \left[\frac{1}{n} \left(\sum_{t=3}^n e_{t-2}^* \eta_{t-1} \right)^2 \right] = \frac{n-2}{n} \sigma_e^2 \sigma_\eta^2 = O(1)$.

Then, $\frac{1}{n^{1/2}} \sum_{t=3}^n e_{t-2}^* \eta_{t-1} = O_p(1)$.

c) Note that for $j \neq 2$, $E[\sum_{t=3}^n e_{t-2}^* e_{t-j}^*] = 0$. Then,

$$\begin{aligned} \text{Var} \left[\sum_{t=3}^n e_{t-2}^* e_{t-j}^* \right] &= E \left[\left(\sum_{t=3}^n e_{t-2}^* e_{t-j}^* \right)^2 \right] \\ &= \sum_{t=3}^n E[e_{t-2}^{*2} e_{t-j}^{*2}] \\ &\quad + 2 \sum_{t=3}^{n-1} \sum_{i=1}^{n-t} E[e_{t-2}^* e_{t-j}^* e_{t-2+i}^* e_{t-j+i}^*] \\ &= (n-2) (\sigma_e^2)^2 \end{aligned}$$

as the sequence of e_t are i.i.d. Then by a similar argument to that in (a), $\text{Var}[\sum_{t=3}^n e_{t-2}^* e_{t-j}^*] = O(n)$. Hence, $\sum_{t=3}^n e_{t-2}^* e_{t-j}^* = O_p(n^{1/2})$.

d) Note that, $E[\sum_{t=3}^n e_t^*] = 0$ and $\text{Var}[\sum_{t=3}^n e_t^*] = (n-2)\sigma_e^2 = O(n)$. Then, $\sum_{t=3}^n e_t^* = O_p(n^{1/2})$.

e) Note that, $E[(e_n^* - e_j^*)^2] = 2\sigma_e^2 = O(1)$. Hence, $e_n^* - e_j^* = O_p(1)$ for $j=1, \dots, (n-1)$.

f) Notice that, $E \left[\left[\frac{1}{n} \sum_{t=1}^n e_{t-2}^{*2} \right] \right] = \sigma_e^2 = O(1)$. Hence, $\frac{1}{n} \sum_{t=1}^n e_{t-2}^{*2} = O_p(1)$.

g) Let $S_{t-2} = \sum_{j=1}^{t-2} \eta_j$. Note that, $\sqrt{\sum_{t=3}^n e_{t-2}^{*2}} = O_p(n^{1/2})$ and $\sqrt{\sum_{t=3}^n S_{t-2}^2} = O_p(n)$ from (f) of Lemma 2.2. and from Lemma 2.1. respectively. Notice, by Cauchy-Schwarz inequality,

$$\left| \sum_{t=3}^n S_{t-2} e_{t-2}^* \right| \leq \sqrt{\sum_{t=3}^n S_{t-2}^2} \sqrt{\sum_{t=3}^n e_{t-2}^{*2}}$$

Therefore, $\sum_{t=3}^n \left(\sum_{j=1}^{t-2} \eta_j \right) e_{t-2}^* = O_p(n^{3/2})$.

h) Note that,

$$\begin{aligned} \sum_{t=3}^n \sum_{j=1}^{t-2} \eta_j (e_t^* - e_{t-1}^*) &= \sum_{t=3}^n \sum_{j=1}^{t-2} \eta_j e_t^* - \sum_{t=3}^n \sum_{j=1}^{t-2} \eta_j e_{t-1}^* \\ &= \sum_{j=1}^{n-2} \sum_{t=j+2}^n \eta_j e_t^* - \sum_{j=1}^{n-2} \sum_{t=j+2}^n \eta_j e_{t-1}^* \\ &= \sum_{j=1}^{n-2} \eta_j [e_{j+2}^* + \dots + e_n^* - e_{j+1}^* - e_{j+2}^* - \dots - e_{n-1}^*] \end{aligned}$$

$$\begin{aligned}
&= \left(\sum_{j=1}^{n-2} \eta_j \right) e_n^* - \sum_{j=1}^{n-2} \eta_j e_{j+1}^* \\
&= O_p(n^{1/2}).
\end{aligned}$$

Theorem 2.1. Consider the model

$$\begin{aligned}
y_t &= h_t + e_t \\
(h_t - \mu) &= \phi(h_{t-1} - \mu) + \eta_t
\end{aligned} \tag{2.14}$$

for $t=1, \dots, n$ with $\phi = 1$, $h_0 = \mu$, and $e_t = \log u_t^2$ where u_t is a sequence of i.i.d. $N(0,1)$ random variables and η_t is a sequence of i.i.d. $N(0, \sigma_\eta^2)$ random variables. u_t and η_t are assumed to be independent. Let a and σ_e^2 be the mean and variance of e_t respectively. In fact, $a = E[\log u_t^2] \approx -1.27$ and $\sigma_e^2 = \text{Var}[\log u_t^2] = \frac{\pi^2}{2}$. Let $\hat{\phi}_{IV}$ be an instrumental variable estimator of ϕ based on the instrument Y_{t-2} . That is,

$$\hat{\phi}_{IV} = \frac{\sum_{t=3}^n Y_t Y_{t-2}}{\sum_{t=3}^n Y_{t-1} Y_{t-2}} \tag{2.15}$$

where $Y_t = y_t - \bar{y} = (h_t - \bar{h}) + (e_t^* - \bar{e}^*)$. Here, $e_t^* = e_t - a$. Also, let $W(t)$ be the standard Brownian motion. Then,

$$n(\hat{\phi}_{IV} - 1) \implies \frac{0.5(W^2(1) - 1) - W(1) \int_0^1 W(t) dt}{\int_0^1 W^2(t) dt - \left(\int_0^1 W(t) dt \right)^2} \tag{2.16}$$

Proof: From the definition of $\hat{\phi}_{IV}$, we have,

$$n(\hat{\phi}_{IV} - 1) = \left[n^{-2} \sum_{t=3}^n Y_{t-1} Y_{t-2} \right]^{-1} \left[n^{-1} \sum_{t=3}^n Y_{t-2} (Y_t - Y_{t-1}) \right] \tag{2.17}$$

where $Y_t = y_t - \bar{y}$. Note that,

$$\begin{aligned}
\sum_{t=3}^n Y_{t-2} (Y_t - Y_{t-1}) &= \sum_{t=3}^n [(h_{t-2} - \bar{h})(e_{t-2}^* - \bar{e}^*)] [\eta_t + (e_t^* - e_{t-1}^*)] \\
&= \sum_{t=3}^n (h_{t-2} - \bar{h}) \eta_t + \sum_{t=3}^n (h_{t-2} - \bar{h})(e_t^* - e_{t-1}^*) \\
&\quad + \sum_{t=3}^n (e_{t-2}^* - \bar{e}^*) \eta_t + \sum_{t=3}^n (e_{t-2}^* - \bar{e}^*)(e_t^* - e_{t-1}^*)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t=3}^n h_{t-2} \eta_t - \bar{h} \sum_{t=3}^n \eta_t + O_p(n^{1/2}) + O_p(n^{1/2}) + O_p(n^{1/2}) \\
&= (0.5) \left(\sum_{t=1}^n \eta_t \right)^2 - (0.5) \sum_{t=1}^n \eta_t^2 - \sum_{t=2}^n \eta_t \eta_{t-1} \\
&\quad - n^{-1} \left(\sum_{t=1}^n \sum_{j=1}^t \eta_j \right) \left(\sum_{t=3}^n \eta_t \right) + O_p(n^{1/2})
\end{aligned}$$

Then,

$$\begin{aligned}
n^{-1} \sum_{t=3}^n Y_{t-2} (Y_t - Y_{t-1}) &= (0.5) \left(n^{-1/2} \sum_{t=1}^n \eta_t \right)^2 - (0.5) n^{-1} \sum_{t=1}^n \eta_t^2 \\
&\quad + O_p(n^{-1/2}) \\
&\quad - \left(n^{-3/2} \sum_{t=1}^n \sum_{j=1}^t \eta_j \right) \left(n^{-1/2} \sum_{t=3}^n \eta_t \right) \\
&\quad + O_p(n^{-1/2})
\end{aligned}$$

Also,

$$\begin{aligned}
\sum_{t=3}^n Y_{t-1} Y_{t-2} &= \sum_{t=3}^n \left[(h_{t-1} - \bar{h}) + (e_{t-1}^* - \bar{e}^*) \right] \left[(h_{t-2} - \bar{h}) + (e_{t-2}^* - \bar{e}^*) \right] \\
&= \sum_{t=3}^n (h_{t-1} - \bar{h})(h_{t-2} - \bar{h}) + \sum_{t=3}^n (h_{t-1} - \bar{h})(e_{t-2}^* - \bar{e}^*) \\
&\quad + \sum_{t=3}^n (e_{t-1}^* - \bar{e}^*)(h_{t-2} - \bar{h}) + \sum_{t=3}^n (e_{t-1}^* - \bar{e}^*)(e_{t-2}^* - \bar{e}^*) \\
&= \sum_{t=3}^n (h_{t-1} - \bar{h})(h_{t-2} - \bar{h}) + O_p(n^{3/2}) + O_p(n^{3/2}) + O_p(n^{1/2}) \\
&= \sum_{t=3}^n h_{t-1} h_{t-2} - \bar{h} \sum_{t=3}^n h_{t-1} - \bar{h} \sum_{t=3}^n h_{t-2} + (n-2)\bar{h}^2 + O_p(n^{3/2}) \\
&= \sum_{t=3}^n (h_{t-2} + \eta_{t-1}) h_{t-2} - \bar{h} \sum_{t=3}^n h_{t-1} - \bar{h} \sum_{t=3}^n h_{t-2} + (n-2)\bar{h}^2 + O_p(n^{3/2}) \\
&= \sum_{t=3}^n \left(\sum_{j=1}^{t-2} \eta_j \right)^2 + (0.5) \left(\sum_{t=1}^n \eta_t \right)^2 - (0.5) \sum_{t=1}^n \eta_t^2 \\
&\quad - n^{-1} \left(\sum_{t=1}^n \sum_{j=1}^t \eta_j \right) \left(\sum_{t=3}^n \sum_{j=1}^{t-1} \eta_j \right) - n^{-1} \left(\sum_{t=1}^n \sum_{j=1}^t \eta_j \right) \left(\sum_{t=3}^n \sum_{j=1}^{t-2} \eta_j \right) \\
&\quad + (n-2)n^{-2} \left(\sum_{t=3}^n \sum_{j=1}^t \eta_j \right)^2 + O_p(n^{3/2})
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t=3}^n \left(\sum_{j=1}^{t-2} \eta_j \right)^2 + O_p(n) + O_p(n) - n^{-1} \left(\sum_{t=1}^n \sum_{j=1}^t \eta_j \right) \left(\sum_{t=3}^n \sum_{j=1}^{t-1} \eta_j \right) \\
&\quad - n^{-1} \left(\sum_{t=1}^n \sum_{j=1}^t \eta_j \right) \left(\sum_{t=3}^n \sum_{j=1}^{t-2} \eta_j \right) \\
&\quad + (n-2)n^{-2} \left(\sum_{t=3}^n \sum_{j=1}^t \eta_j \right)^2 + O_p(n^{3/2})
\end{aligned}$$

Hence, by Lemma 2.1., The Continuous Mapping Theorem, Slutsky's Theorem, and Chan & Wei (1988),

$$\left[\begin{array}{c} n^{-1} \sum_{t=3}^n Y_{t-2} (Y_t - Y_{t-1}) \\ n^{-2} \sum_{t=3}^n Y_{t-1} Y_{t-2} \end{array} \right] \Longrightarrow \left[\begin{array}{c} \sigma_\eta^2 (0.5(W(1)^2 - 1) - W(1) \int_0^1 W(t) dt) \\ \sigma_\eta^2 \left(\int_0^1 W^2(t) dt - \left(\int_0^1 W(t) dt \right)^2 \right) \end{array} \right] \quad (2.18)$$

Note that, the random variable to which $n^{-2} \sum_{t=3}^n Y_{t-1} Y_{t-2}$ converges is positive.

Hence, by Continuous Mapping Theorem,

$$\left[n^{-2} \sum_{t=3}^n Y_{t-1} Y_{t-2} \right]^{-1} \Longrightarrow \left[\sigma_\eta^2 \left(\int_0^1 W^2(t) dt - \left(\int_0^1 W(t) dt \right)^2 \right) \right]^{-1}. \quad \text{Then, (2.16) follows.}$$

Theorem 2.2. Consider model (2.14). Let $\hat{\sigma}_{IV}^2$ be the variance of the residuals in instrumental variable regression. That is,

$$\hat{\sigma}_{IV}^2 = \frac{1}{n-4} \sum_{t=3}^n (Y_t - \hat{\phi}_{IV} Y_{t-1})^2 \quad (2.19)$$

Then, $\hat{\sigma}_{IV}^2$ is a consistent estimator of $2\sigma_e^2 + \sigma_\eta^2$.

Proof: Notice,

$$\begin{aligned}
\hat{\sigma}_{IV}^2 &= \frac{1}{n-4} \sum_{t=3}^n (Y_t - \hat{\phi}_{IV} Y_{t-1})^2 \\
&= \frac{1}{n-4} \sum_{t=3}^n \left[(Y_t - Y_{t-1}) - (\hat{\phi}_{IV} - 1) Y_{t-1} \right]^2 \\
&= \frac{1}{n-4} (\hat{\phi}_{IV} - 1)^2 \sum_{t=3}^n Y_{t-1}^2 - \frac{2}{n-4} (\hat{\phi}_{IV} - 1) \sum_{t=3}^n Y_{t-1} (Y_t - Y_{t-1}) \\
&\quad + \frac{1}{n-4} \sum_{t=3}^n (Y_t - Y_{t-1})^2
\end{aligned}$$

Note that,

$$\begin{aligned}
\sum_{t=3}^n Y_{t-1}^2 &= \sum_{t=3}^n \left[(h_{t-1} - \bar{h}) + (e_{t-1}^* - \bar{e}^*) \right]^2 \\
&= \sum_{t=3}^n \left(\sum_{j=1}^{t-1} \eta_j \right)^2 - 2n^{-1} \left[\sum_{t=1}^n \left(\sum_{j=1}^t \eta_j \right) \right] \left[\sum_{t=3}^n \left(\sum_{j=1}^t \eta_j \right) \right] \\
&\quad + (n-2)n^{-2} \left[\sum_{t=1}^n \left(\sum_{j=1}^t \eta_j \right) \right]^2 + \sum_{t=3}^n (e_{t-1}^* - \bar{e}^*)^2 \\
&\quad + 2 \sum_{t=3}^n (h_t - \bar{h})(e_{t-1}^* - \bar{e}^*) \\
&= O_p(n^2) + O_p(n^2) + O_p(n^2) + O_p(n) + O_p(n^{3/2}) \\
&= O_p(n^2)
\end{aligned}$$

and

$$\begin{aligned}
\sum_{t=3}^n Y_{t-1} (Y_t - Y_{t-1}) &= \sum_{t=3}^n \left[(h_{t-1} - \bar{h}) + (e_{t-1}^* - \bar{e}^*) \right] \left[\eta_t + (e_t^* - e_{t-1}^*) \right] \\
&= (0.5) \left(\sum_{t=1}^n \eta_t \right)^2 - (0.5) \sigma_\eta^2 \sum_{t=1}^n (\eta_t^2 / \sigma_\eta^2) - \bar{h} \sum_{t=3}^n \eta_t \\
&\quad + \sum_{t=3}^n \sum_{j=1}^t \eta_t (e_t^* - e_{t-1}^*) - \bar{h} \sum_{t=3}^n (e_t^* - e_{t-1}^*) \\
&\quad + \sum_{t=3}^n \eta_t e_{t-1}^* - \bar{e}^* \sum_{t=3}^n \eta_t + \sum_{t=3}^n e_t^* e_{t-1}^* - \bar{e}^* \sum_{t=3}^n e_t^* - \sum_{t=3}^n (e_{t-1}^*)^2 \\
&\quad + \bar{e}^* \sum_{t=3}^n e_{t-1}^* \\
&= O_p(n) + O_p(n) + O_p(n) + O_p(n^{1/2}) + O_p(n) \\
&\quad + O_p(n^{1/2}) + O_p(1) + O_p(n^{1/2}) + O_p(1) + O_p(n) + O_p(1) \\
&= O_p(n)
\end{aligned}$$

Also,

$$\sum_{t=3}^n (Y_t - Y_{t-1})^2 = \sum_{t=3}^n \eta_t^2 + \sum_{t=3}^n (e_t^*)^2 + \sum_{t=3}^n (e_{t-1}^*)^2 + 2 \sum_{t=3}^n \eta_t (e_t^* - e_{t-1}^*)$$

By the results in Lemma 2.2.,

$$\begin{aligned}\sum_{t=3}^n Y_{t-1}^2 &= O_p(n^2), \\ \sum_{t=3}^n Y_{t-1} (Y_t - Y_{t-1}) &= O_p(n) \\ \frac{1}{n-4} \sum_{t=3}^n (Y_t - Y_{t-1})^2 &\xrightarrow{p} 2\sigma_e^2 + \sigma_\eta^2\end{aligned}$$

Also, we have shown that $\hat{\phi}_{IV} - 1 = O_p(n^{-1})$ in Model (2.14). Therefore,

$$\begin{aligned}\hat{\sigma}_{IV}^2 &= O_p(n^{-1}) + O_p(n^{-1}) + \frac{1}{n-4} \sum_{t=3}^n (Y_t - Y_{t-1})^2 \\ &\xrightarrow{p} 2\sigma_e^2 + \sigma_\eta^2\end{aligned}$$

Theorem 2.3. Consider Model (2.14). Let t_{IV} be a pivotal test statistic based on the instrumental variable estimators. That is,

$$t_{IV} = \left(\hat{\sigma}_{IV}^2\right)^{-1/2} \left(\sum_{t=3}^n Y_{t-1} Y_{t-2}\right)^{1/2} \left(\hat{\phi}_{IV} - 1\right) \quad (2.20)$$

where

$$\hat{\sigma}_{IV}^2 = \frac{1}{n-4} \sum_{t=3}^n \left(Y_t - \hat{\phi}_{IV} Y_{t-1}\right)^2$$

Then,

$$\left(\hat{\sigma}_{IV}^2\right)^{1/2} \left|\left(\hat{\sigma}_{IV}^2 - 2\sigma_e^2\right)\right|^{-1/2} t_{IV} \implies \frac{0.5(W^2(1) - 1) - W(1) \int_0^1 W(t) dt}{\left[\int_0^1 W^2(t) dt - \left(\int_0^1 W(t) dt\right)^2\right]^{1/2}} \quad (2.21)$$

Proof: Notice,

$$\begin{aligned}t_{IV} &= \left(\hat{\sigma}_{IV}^2\right)^{-1/2} \left(\sum_{t=3}^n Y_{t-1} Y_{t-2}\right)^{1/2} \left(\hat{\phi}_{IV} - 1\right) \\ &= \left(\hat{\sigma}_{IV}^2\right)^{-1/2} \left[n^{-2} \sum_{t=3}^n Y_{t-1} Y_{t-2}\right]^{-1/2} \left[n^{-1} \sum_{t=3}^n Y_{t-2} (Y_t - Y_{t-1})\right]\end{aligned}$$

Note that, by the Continuous Mapping Theorem, $\left(\hat{\sigma}_{IV}^2\right)^{-1/2} \xrightarrow{p} (2\sigma_e^2 + \sigma_\eta^2)^{-1/2}$ as $\sigma_e^2 > 0$ and $\sigma_\eta^2 > 0$. Also, again by the Continuous Mapping Theorem, $\left[n^{-2} \sum_{t=3}^n Y_{t-1} Y_{t-2}\right]^{-1/2}$

$\implies (\sigma_\eta^2)^{-1/2} \left[\int_0^1 W^2(t) dt - \left(\int_0^1 W(t) dt \right)^2 \right]^{-1/2}$. Hence, by result (2.18) and Slutsky's Theorem,

$$t_{IV} \implies (2\sigma_e^2 + \sigma_\eta^2)^{-1/2} (\sigma_\eta^2)^{1/2} \frac{0.5(W^2(1) - 1) - W(1) \int_0^1 W(t) dt}{\left[\int_0^1 W^2(t) dt - \left(\int_0^1 W(t) dt \right)^2 \right]^{1/2}}$$

Now, notice that, σ_e^2 is a known quantity. Therefore, from Theorem 2.2., $|\hat{\sigma}_{IV}^2 - 2\sigma_e^2|^{-1/2} \xrightarrow{p} (\sigma_\eta^2)^{-1/2}$ as $\sigma_\eta^2 > 0$. Hence, (2.21) follows from Slutsky's Theorem.

Note that the estimator that is used in the correction factor in Theorem 2.3 is a consistent estimator for σ_η^2 if the data are generated from Model (2.14). We will now show that a more general form of the correction factor in Theorem 2.3 which is also consistent for σ_η^2 when $|\phi| < 1$.

Lemma 2.3. Consider Model (2.14) with $\phi \in (-1, 0) \cup (0, 1)$ and $h_0 \sim N(\mu, \frac{\sigma_\eta^2}{1-\phi^2})$. Then,

$$\begin{aligned} a) & \frac{1}{n} \sum_{t=3}^n (h_{t-i} - \mu)(e_{t-j} - a) = O_p(n^{-1/2}) \\ b) & \frac{1}{n} \sum_{t=3}^n (h_{t-i} - \bar{h})(e_{t-j} - \bar{e}) = O_p(n^{-1/2}) \end{aligned}$$

Proof: Let $h_t^* = h_t - \mu$ and $e_t^* = e_t - a$. Also, let \bar{h}^* and \bar{e}^* denote the sample means of $\{h_t^*\}$ and $\{e_t^*\}$ respectively.

a) First of all, note that $\frac{1}{n} \sum_{t=3}^n (h_{t-i} - \mu)(e_{t-j} - a) = \frac{1}{n} \sum_{t=3}^n h_{t-i}^* e_{t-j}^*$. Notice, $E \left[\frac{1}{n} \sum_{t=3}^n h_{t-i}^* e_{t-j}^* \right] = 0$ as $E [h_{t-i}^* e_{t-j}^*] = E [h_{t-i}^*] E [e_{t-j}^*] = 0$ since $\text{Cov}(h_t^*, e_s^*) = 0 \forall t$ and s . Hence,

$$\begin{aligned} E \left[\frac{1}{n^2} \left(\sum_{t=3}^n h_{t-i}^* e_{t-j}^* \right)^2 \right] &= \text{Var} \left[\frac{1}{n} \sum_{t=3}^n h_{t-i}^* e_{t-j}^* \right] \\ &= \frac{1}{n^2} \sum_{t=3}^n \text{Var} [h_{t-i}^* e_{t-j}^*] + \frac{2}{n^2} \sum_{k=3}^{n-1} \sum_{l=k+1}^n \text{Cov} [h_{k-i}^* e_{k-j}^*, h_{l-i}^* e_{l-j}^*] \\ &= \frac{1}{n^2} \sum_{t=3}^n E [(h_{t-i}^*)^2 (e_{t-j}^*)^2] \\ &= \frac{1}{n^2} \sum_{t=3}^n E [(h_{t-i} - \mu)^2] E [(e_{t-j} - a)^2] \end{aligned}$$

$$\begin{aligned}
&= \frac{n-2}{n^2} \frac{\sigma_\eta^2}{1-\phi^2} \sigma_e^2 \\
&= O(n^{-1})
\end{aligned}$$

This implies that $\frac{1}{n} \sum_{t=3}^n h_{t-i}^* e_{t-j}^* = O_p(n^{-1/2})$. Result follows.

b) Note that,

$$\begin{aligned}
\frac{1}{n} \sum_{t=3}^n (h_{t-i} - \bar{h})(e_{t-j} - \bar{e}) &= \frac{1}{n} \sum_{t=3}^n (h_{t-i}^* - \bar{h}^*)(e_{t-j}^* - \bar{e}^*) \\
&= \frac{1}{n} \sum_{t=3}^n h_{t-i}^* e_{t-j}^* - \bar{e}^* \frac{1}{n} \sum_{t=3}^n h_{t-i}^* - \bar{h}^* \frac{1}{n} \sum_{t=3}^n e_{t-j}^* + \frac{n-2}{n} \bar{h}^* \bar{e}^* \\
&= O_p(n^{-1/2}) + O_p(n^{-1}) + O_p(n^{-1}) + O_p(n^{-1})
\end{aligned}$$

Here, the order of the first term is due to Lemma 2.3. (a). Also, notice that $\bar{h}^* = O_p(n^{-1/2})$ as h_t is stationary.

Theorem 2.4. Consider Model (2.14) with $\phi \in (-1, 0) \cup (0, 1)$ and $h_0 \sim N(\mu, \frac{\sigma_\eta^2}{1-\phi^2})$. Let $\hat{\phi}_{IV}$ be defined as in (2.15). Then, $\hat{\phi}_{IV}$ is a consistent estimator of ϕ under this model.

Proof: Note that,

$$\hat{\phi}_{IV} - \phi = \frac{\frac{1}{n} \sum_{t=3}^n Y_{t-2} (Y_t - \phi Y_{t-1})}{\frac{1}{n} \sum_{t=3}^n Y_{t-1} Y_{t-2}}$$

First of all, let $\gamma_h(k)$ be the lag k autocovariance of the log-volatility series $\{h_t\}$.

The numerator is,

$$\begin{aligned}
\frac{1}{n} \sum_{t=3}^n Y_{t-2} (Y_t - \phi Y_{t-1}) &= \frac{1}{n} \sum_{t=3}^n Y_{t-2} Y_t - \phi \frac{1}{n} \sum_{t=3}^n Y_{t-2} Y_{t-1} \\
&= \frac{1}{n} \sum_{t=3}^n [(h_{t-2} - \bar{h}) + (e_{t-2}^* - \bar{e}^*)] [(h_t - \bar{h}) + (e_t^* - \bar{e}^*)] \\
&\quad - \phi \frac{1}{n} \sum_{t=3}^n [(h_{t-2} - \bar{h}) + (e_{t-2}^* - \bar{e}^*)] [(h_{t-1} - \bar{h}) + (e_{t-1}^* - \bar{e}^*)] \\
&= \frac{1}{n} \sum_{t=3}^n (h_{t-2} - \bar{h})(h_t - \bar{h}) + \frac{1}{n} \sum_{t=3}^n (h_{t-2} - \bar{h})(e_t^* - \bar{e}^*) \\
&\quad + \frac{1}{n} \sum_{t=3}^n (e_{t-2}^* - \bar{e}^*)(h_t - \bar{h}) + \frac{1}{n} \sum_{t=3}^n (e_{t-2}^* - \bar{e}^*)(e_t^* - \bar{e}^*)
\end{aligned}$$

$$\begin{aligned}
& -\phi \frac{1}{n} \sum_{t=3}^n (h_{t-2} - \bar{h})(h_{t-1} - \bar{h}) - \phi \frac{1}{n} \sum_{t=3}^n (h_{t-2} - \bar{h})(e_{t-1}^* - \bar{e}^*) \\
& -\phi \frac{1}{n} \sum_{t=3}^n (e_{t-2}^* - \bar{e}^*)(h_{t-1} - \bar{h}) - \phi \frac{1}{n} \sum_{t=3}^n (e_{t-2}^* - \bar{e}^*)(e_{t-1}^* - \bar{e}^*)
\end{aligned}$$

By Lemma 2.3. (b),

$$\begin{aligned}
\frac{1}{n} \sum_{t=3}^n Y_{t-2}(Y_t - \phi Y_{t-1}) &= \frac{1}{n} \sum_{t=3}^n (h_{t-2} - \bar{h})(h_t - \bar{h}) + O_p(n^{-1/2}) + O_p(n^{-1/2}) \\
&+ \frac{1}{n} \sum_{t=3}^n (e_{t-2}^* - \bar{e}^*)(e_t^* - \bar{e}^*) - \phi \frac{1}{n} \sum_{t=3}^n (h_{t-2} - \bar{h})(h_{t-1} - \bar{h}) \\
&+ O_p(n^{-1/2}) + O_p(n^{-1/2}) - \phi \frac{1}{n} \sum_{t=3}^n (e_{t-2}^* - \bar{e}^*)(e_{t-1}^* - \bar{e}^*)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{1}{n} \sum_{t=3}^n Y_{t-2}(Y_t - \phi Y_{t-1}) &\xrightarrow{p} \gamma_h(2) - \phi \gamma_h(1) + \text{Cov}(e_{t-2}^*, e_t^*) - \phi \text{Cov}(e_{t-2}^*, e_{t-1}^*) \\
&= 0 \tag{2.22}
\end{aligned}$$

as $\gamma_h(2) = \phi \gamma_h(1)$.

The denominator is,

$$\begin{aligned}
\frac{1}{n} \sum_{t=3}^n Y_{t-1} Y_{t-2} &= \frac{1}{n} \sum_{t=3}^n [(h_{t-1} - \bar{h}) + (e_{t-1}^* - \bar{e}^*)] [(h_{t-2} - \bar{h}) + (e_{t-2}^* - \bar{e}^*)] \\
&= \frac{1}{n} \sum_{t=3}^n (h_{t-1} - \bar{h})(h_{t-2} - \bar{h}) + \frac{1}{n} \sum_{t=3}^n (e_{t-1}^* - \bar{e}^*)(e_{t-2}^* - \bar{e}^*) \\
&\quad + \frac{1}{n} \sum_{t=3}^n (e_{t-2}^* - \bar{e}^*)(h_{t-1} - \bar{h}) + \frac{1}{n} \sum_{t=3}^n (e_{t-1}^* - \bar{e}^*)(h_{t-2} - \bar{h}) \\
&= \frac{1}{n} \sum_{t=3}^n (h_{t-1} - \bar{h})(h_{t-2} - \bar{h}) + \frac{1}{n} \sum_{t=3}^n (e_{t-1}^* - \bar{e}^*)(e_{t-2}^* - \bar{e}^*) \\
&\quad + O_p(n^{-1/2}) + O_p(n^{-1/2})
\end{aligned}$$

Therefore,

$$\frac{1}{n} \sum_{t=3}^n Y_{t-1} Y_{t-2} \xrightarrow{p} \gamma_h(1) + \text{Cov}(e_{t-1}^*, e_{t-2}^*) = \gamma_h(1) \neq 0 \tag{2.23}$$

Results in (2.22) and (2.23) lead to the conclusion that $\hat{\phi}_{IV} - \phi = o_p(1)$ under the alternative hypothesis.

Theorem 2.5. Consider Model (2.14) with $\phi \in (-1, 0) \cup (0, 1)$ and $h_0 \sim N(\mu, \frac{\sigma_\eta^2}{1-\phi^2})$. In this case, $\hat{\sigma}_{IV}^2$ as defined in (2.19) is a consistent estimator of $(1 + \phi^2)\sigma_e^2 + \sigma_\eta^2$.

Proof: Note that,

$$\begin{aligned}\hat{\sigma}_{IV}^2 &= \frac{1}{n-4} \sum_{t=3}^n (Y_t - \hat{\phi}_{IV} Y_{t-1})^2 \\ &= \frac{1}{n-4} \sum_{t=3}^n [(Y_t - \phi Y_{t-1}) - (\hat{\phi}_{IV} - \phi) Y_{t-1}]^2 \\ &= \frac{1}{n-4} \sum_{t=3}^n (Y_t - \phi Y_{t-1})^2 - 2(\hat{\phi}_{IV} - \phi) \frac{1}{n-4} \sum_{t=3}^n Y_{t-1} (Y_t - \phi Y_{t-1}) \\ &\quad + (\hat{\phi}_{IV} - \phi)^2 \frac{1}{n-4} \sum_{t=3}^n Y_{t-1}^2\end{aligned}$$

Now, notice that,

$$\begin{aligned}\frac{1}{n-4} \sum_{t=3}^n Y_t^2 &= \frac{1}{n-4} \sum_{t=3}^n [(h_t - \bar{h}) + (e_t^* - \bar{e}^*)]^2 \\ &= \frac{1}{n-4} \sum_{t=3}^n (h_t - \bar{h})^2 + \frac{1}{n-4} \sum_{t=3}^n (e_t^* - \bar{e}^*)^2 \\ &\quad + \frac{2}{n-4} \sum_{t=3}^n (h_t - \bar{h})(e_t^* - \bar{e}^*) \\ &= \frac{1}{n-4} \sum_{t=3}^n (h_t - \bar{h})^2 + \frac{1}{n-4} \sum_{t=3}^n (e_t^* - \bar{e}^*)^2 + O_p(n^{-1/2})\end{aligned}$$

Therefore,

$$\frac{1}{n-4} \sum_{t=3}^n Y_t^2 \xrightarrow{p} \gamma_h(0) + \sigma_e^2 \quad (2.24)$$

This result along with Theorem 2.4 and the result 2.23 in the proof of Theorem 2.4 imply that,

$$\begin{aligned}\hat{\sigma}_{IV}^2 &\xrightarrow{p} \gamma_h(0) + \sigma_e^2 - 2\phi\gamma_h(1) + \phi^2\gamma_h(0) + \phi^2\sigma_e^2 \\ &= (1 - \phi^2)\gamma_h(0) + (1 + \phi^2)\sigma_e^2 \\ &= (1 - \phi^2) \frac{\sigma_\eta^2}{(1 - \phi^2)} + (1 + \phi^2)\sigma_e^2 \\ &= \sigma_\eta^2 + (1 + \phi^2)\sigma_e^2.\end{aligned}$$

Note that $\hat{\sigma}_{IV}^2$ is a consistent estimator of $\sigma_\eta^2 + (1 + \phi^2)\sigma_e^2$ under both the null and the alternative SVM.

By Theorem 2.1 and Theorem 2.4, $\hat{\phi}_{IV} \xrightarrow{p} \phi$ under both the null and the alternative hypothesis which implies that, by continuity, $(1 + \hat{\phi}_{IV}^2)\sigma_e^2 \xrightarrow{p} (1 + \phi^2)\sigma_e^2$ under both null and alternative SVM. Then, $\sigma_{IV}^2 - (1 + \hat{\phi}_{IV}^2)\sigma_e^2 \xrightarrow{p} \sigma_\eta^2$. Hence, by Continuous Mapping Theorem, $|\sigma_{IV}^2 - (1 + \hat{\phi}_{IV}^2)\sigma_e^2|^{-1/2} \xrightarrow{p} (\sigma_\eta^2)^{-1/2}$ as $\sigma_\eta^2 > 0$.

The fruits of these theorems in practice are explained after introducing some notations for the pivotal test statistics considered here.

We address $(\hat{\sigma}_{IV}^2)^{1/2} |\hat{\sigma}_{IV}^2 - 2\sigma_e^2|^{-1/2}$ by the term “SVM correction factor which is consistent under H_0 ”. We denote the instrumental variable t statistic that is multiplied by the “SVM correction factor which is consistent under H_0 ” by “ $t_{\mu,IV,SVM}(H_0)$ ” which is equal to the left hand side of the expression in (2.21). Similarly, we address $(\hat{\sigma}_{IV}^2)^{1/2} |\hat{\sigma}_{IV}^2 - (1 + \hat{\phi}_{IV}^2)\sigma_e^2|^{-1/2}$ by the expression “SVM correction factor which is consistent under both H_0 and H_1 ”. In addition, we denote the instrumental variable t statistic which is multiplied by the “SVM correction factor which is consistent under both H_0 and H_1 ” by “ $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$ ”.

Having presented the notations for these test statistics, we are ready to develop the practical implications of the theoretical results regarding them. By Theorems 2.1, 2.3, and 2.5, the limiting distributions of $n(\hat{\phi}_{IV} - 1)$, $t_{\mu,IV,SVM}(H_0)$, and $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$ when the true model is (2.14) are same as the limiting distributions of the Dickey-Fuller test statistic based on $n(\hat{\phi} - 1)$ and the Dickey-Fuller pivotal test statistic when the true model is an AR(1) with $\phi = 1$. Since they share the same distributional properties in the limit, one can use the Dickey-Fuller cutoff points in performing a unit root test based on $t_{\mu,IV,SVM}(H_0)$ and $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$ in a SVM when the sample size is sufficiently large. A similar statement holds for the test statistic $n(\hat{\phi}_{IV} - 1)$. There is another implication as explained next. Hall (1989) showed that when the true model is an ARIMA(0,1,1), the distribution of the instrumental variable test statistic $n(\hat{\phi}_{IV} - 1)$ converges to the limiting distribution of

the Dickey-Fuller test statistic based on $n(\hat{\phi} - 1)$. Similarly, he showed that, when the true model is an ARIMA(0,1,1), the pivotal statistic in instrumental variable regression, t_{IV} , multiplied by a factor converges to the limiting distribution of the Dickey-Fuller test statistic. The multiplying factor is explained in detail in the next paragraph. This result of Hall along with the fact that (2.14) has second order properties that are similar to an ARMA(1,1) model motivates us to use the percentiles of the finite sampling distributions of these test statistics as cutoff points when we apply $n(\hat{\phi}_{IV} - 1)$, $t_{\mu,IV,SVM}(H_0)$ or $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$ to do a unit root test on a data set that is believed to be coming from a SVM.

Besides the t test statistics multiplied by SVM correction factors, $\log r_t^2$ with $\{r_t\}$ generated from (2.1) sharing the same second order properties with an ARMA(1,1) model motivates us to use the correction factor that is derived by Hall (1989) for data generated from an ARIMA(0,1,1). “Hall’s correction factor” is $s_\epsilon s^{-1}$ with s_ϵ^2 being a consistent estimator of the variance of ϵ_t where $\epsilon_t = v_t - \theta v_{t-1}$. Here, v_t is the random shock (2.5). Also, s^2 is a consistent estimator of σ_ϵ^2 where $\sigma_\epsilon^2 = \lim_{n \rightarrow \infty} n \text{Var}(\bar{\epsilon})$. We obtain s_ϵ^2 from ordinary least square estimation of $\epsilon_t = Y_t - \hat{\phi}_{IV} Y_{t-1} = v_t - \theta v_{t-1}$. We use a one step Gauss Newton estimation procedure in order to estimate θ and $\text{Var}(v_t) = \sigma^2$ from this regression and denote the resulting estimators by $\hat{\theta}$ and $\hat{\sigma}^2$ respectively. Then, $s_\epsilon^2 = (1 + \hat{\theta}^2)\hat{\sigma}^2$ and $s^2 = (1 - \hat{\theta})^2\hat{\sigma}^2$. We denote the instrumental variable test statistic that gets multiplied by “Hall’s correction factor” by “ $t_{\mu,IV,Hall}$ ”.

We apply these instrumental variable unit root test statistics to the data in the form of $\log r_t^2$ where $\{r_t\}$ are generated from the SVM in (2.1). The unit root test statistics we apply are; (i) $n(\hat{\phi}_{IV} - 1)$, (ii) $t_{\mu,IV,SVM}(H_0)$, (iii) $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$, and (iv) $t_{\mu,IV,Hall}$.

We use Y_{t-2} as an instrument. Below, we explain how we calculate $t_{\mu,IV,SVM}(H_0)$, $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$, and $t_{\mu,IV,Hall}$. In these calculations, we use a sandwich estimator formula to estimate $\text{Var}(\hat{\phi}_{IV})$ rather than the OLS formula in order to insure

positivity.

$$\begin{aligned}
t_{\mu,IV,SVM}(H_0) &= \left| \hat{\sigma}_{IV}^2 - 2\sigma_e^2 \right|^{-\frac{1}{2}} \left| \sum_{t=3}^n Y_{t-1} Y_{t-2} \right| \left(\sum_{t=3}^n Y_{t-2}^2 \right)^{-\frac{1}{2}} (\hat{\phi}_{IV} - 1) \\
t_{\mu,IV,SVM}(H_0 \text{ and } H_1) &= \left| \hat{\sigma}_{IV}^2 - (1 + \hat{\phi}_{IV}^2) \sigma_e^2 \right|^{-\frac{1}{2}} \left| \sum_{t=3}^n Y_{t-1} Y_{t-2} \right| \left(\sum_{t=3}^n Y_{t-2}^2 \right)^{-\frac{1}{2}} (\hat{\phi}_{IV} - 1) \\
t_{\mu,IV,Hall} &= \left| \sum_{t=3}^n Y_{t-1} Y_{t-2} \right| \left(\sum_{t=3}^n Y_{t-2}^2 \right)^{-\frac{1}{2}} s^{-1} (\hat{\phi}_{IV} - 1) \tag{2.25}
\end{aligned}$$

In the last one, s_ϵ disappears as it gets cancelled out. Here, $Y_t = y_t - \bar{y}$ where $y_t = \log r_t^2$.

We perform several tests based on the instrumental variable test statistics by comparing the values of $n(\hat{\phi}_{IV} - 1)$, $t_{\mu,IV,SVM}(H_0)$, $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$, and $t_{\mu,IV,Hall}$ against cutoff values $(n(\hat{\phi}_{IV} - 1))_{n,\theta}$ and $\tau_{\mu,IV,Hall,n,\theta}$ that are chosen from the empirical percentile tables we construct with respect to the sample size and the value of θ . Empirical percentiles $(n(\hat{\phi}_{IV} - 1))_{n,\theta}$ and $\tau_{\mu,IV,Hall,n,\theta}$ are calculated for the data generated from an ARMA(1,1) with $\phi = 1$ for different values of θ . $\tau_{\mu,IV,Hall,n,\theta}$ uses ‘‘Hall’s correction factor’’ as explained before. Here are the different rejection criteria, based on the instrumental variable test statistics that we consider for SV models:

Reject H_0 if

$$\begin{aligned}
n(\hat{\phi}_{IV} - 1) &< (n(\hat{\phi}_{IV} - 1))_{n,\theta} \\
t_{\mu,IV,Hall} &< \tau_{\mu,IV,Hall,n,\theta} \\
t_{\mu,IV,SVM}(H_0) &< \tau_{\mu,IV,Hall,n,\theta} \tag{2.26} \\
t_{\mu,IV,SVM}(H_0 \text{ and } H_1) &< \tau_{\mu,IV,Hall,n,\theta}
\end{aligned}$$

We digress from the rejection criteria for instrumental variable tests in SVM for a while to lay out some interesting results led by the following theorems. These theorems are related with the limiting behavior of the ‘‘SVM correction factor which is consistent under both H_0 and H_1 ’’ when the true model is an ARIMA(0,1,1).

Similar results could be obtained for “ SVM correction factor which is consistent under H_0 ” for such general models.

Theorem 2.7. Consider the model

$$\begin{aligned}(y_t - \mu^*) &= \phi(y_{t-1} - \mu^*) + \epsilon_t \\ \epsilon_t &= v_t - \theta v_{t-1} \quad t=1, \dots, n\end{aligned}$$

where $\phi = 1$, and $\{v_t\}$ is independently and identically distributed with 0 mean and a finite variance $\sigma^2 > 0$. Also, $E[|\epsilon_t|^{\beta+\delta}] < \infty$ for some $\beta > 2$ and $\delta > 0$. Then,

$$\left(\hat{\sigma}_{IV}^2\right)^{1/2} \left| \left(\hat{\sigma}_{IV}^2 - (1 + \hat{\phi}_{IV}^2)\sigma_e^2\right) \right|^{-1/2} \xrightarrow{p} (1 + \theta^2)^{1/2} \sigma \left| (1 + \theta^2)\sigma^2 - 2\sigma_e^2 \right|^{-1/2} \quad (2.27)$$

where $\sigma_e^2 = \frac{\pi^2}{2}$.

Proof: Note that,

$$\begin{aligned}\hat{\sigma}_{IV}^2 &= \frac{1}{n-4} \sum_{t=3}^n (Y_t - \hat{\phi}_{IV} Y_{t-1})^2 \\ &= \frac{1}{n-4} \sum_{t=3}^n \left[(Y_t - Y_{t-1}) - (\hat{\phi}_{IV} - 1) Y_{t-1} \right]^2 \\ &= \frac{1}{n-4} (\hat{\phi}_{IV} - 1)^2 \sum_{t=3}^n Y_{t-1}^2 - \frac{2}{n-4} (\hat{\phi}_{IV} - 1) \sum_{t=3}^n Y_{t-1} (Y_t - Y_{t-1}) \\ &\quad + \frac{1}{n-4} \sum_{t=3}^n (Y_t - Y_{t-1})^2\end{aligned}$$

where $Y_t = y_t - \bar{y}$. For an ARIMA(0,1,1) model, $\sum_{t=3}^n Y_{t-1}^2 = O_p(n^2)$ and $\sum_{t=3}^n Y_{t-1} (Y_t - Y_{t-1}) = O_p(n)$. In addition, by Hall (1989), $\hat{\phi}_{IV} - 1 = O_p(n^{-1})$. Also,

$$\begin{aligned}\frac{1}{n-4} \sum_{t=3}^n (Y_t - Y_{t-1})^2 &= \frac{1}{n-4} \sum_{t=3}^n \epsilon_t^2 \\ &= \sum_{t=3}^n (v_t - \theta v_{t-1})^2 \\ &= \frac{1}{n-4} \sum_{t=3}^n v_t^2 + \theta^2 \frac{1}{n-4} \sum_{t=3}^n v_{t-1}^2 - 2\theta \frac{1}{n-4} \sum_{t=3}^n v_t v_{t-1} \\ &\xrightarrow{p} (1 + \theta^2)\sigma^2\end{aligned}$$

as $\{v_t\}$ are i.i.d. Therefore, $\hat{\sigma}_{IV}^2$ is a consistent estimator of $(1 + \theta^2)\sigma^2$. This implies that $\left| \left(\hat{\sigma}_{IV}^2 - (1 + \hat{\phi}_{IV}^2)\sigma_e^2\right) \right|^{-1/2} \xrightarrow{p} \left| (1 + \theta^2)\sigma^2 - 2\sigma_e^2 \right|^{-1/2}$ from which the result follows.

Theorem 2.8. Consider the results in 2.27 with $\phi = 1$ and $\theta = \frac{\sigma_e^2}{\sigma^2}$ where $\sigma_e^2 = \frac{\pi^2}{2}$. Then,

$$\left(\hat{\sigma}_{IV}^2\right)^{1/2} \left| \left(\hat{\sigma}_{IV}^2 - (1 + \hat{\phi}_{IV}^2)\sigma_e^2\right) \right|^{-1/2} \xrightarrow{p} (1 - \theta)^{-1}(1 + \theta^2)^{1/2} \quad (2.28)$$

Proof: From the relationship between θ and σ^2 , $\sigma_e^2 = \theta\sigma^2$. Therefore,

$$\begin{aligned} \hat{\sigma}_{IV}^2 - (1 + \hat{\phi}_{IV}^2)\sigma_e^2 &= \hat{\sigma}_{IV}^2 - (1 + \hat{\phi}_{IV}^2)\theta\sigma^2 \\ &\xrightarrow{p} (1 + \theta^2)\sigma^2 - 2\theta\sigma^2 \\ &= (1 - \theta)^2\sigma^2 \end{aligned}$$

Hence, the result follows.

Before developing the practical implications of these theorems in SVM, it is necessary to indicate at this point that “Hall’s correction factor, ” $s_e s^{-1}$, converges to $(1 - \theta)^{-1}(1 + \theta^2)^{1/2}$ in probability when the true model is an ARIMA(0,1,1).

By Theorem 2.7, we observe that the “SVM correction factor which is consistent under both H_0 and H_1 ” converges to a quantity which depends on σ^2 if the true model is a general ARIMA(0,1,1). However, the same correction factor converges to a quantity which is free of σ^2 if the true model is an ARIMA(0,1,1) which possesses the restriction on σ^2 . We should note that this is the restriction on an ARMA(1,1) model that carries the same second order properties as $\log r_t^2$ where $\{r_t\}$ denote the mean corrected returns in the form of a SVM. The limiting value of the “SVM correction factor which is consistent under both H_0 and H_1 ” in the restricted ARIMA(0,1,1) is in fact same as the limiting value of “Hall’s correction factor ” which is computed for the data coming from a general ARIMA(0,1,1). That is, it converges in probability to $(1 - \theta)^{-1}(1 + \theta^2)^{1/2}$. These results altogether serve as a motive for calculating the percentiles of the finite sampling distribution of the instrumental variable t test statistic multiplied by “SVM correction factor which is consistent under both H_0 and H_1 ” where the data are generated from an ARIMA(0,1,1) satisfying the restriction on σ^2 and using these percentiles as critical points in instrumental variable testing

procedures. As for the other test statistics we considered before, we calculate the percentiles of this one for varying values of θ and sample size. Let us denote this test statistic by $\tau_{\mu,IV,SVM,n,\theta}^{rst}$ (H_0 and H_1) where “*rst*” in the superscript indicates that Monte Carlo data sets to calculate the empirical percentiles are generated from an ARIMA(0,1,1) with the restriction $\sigma^2 = \frac{\sigma_\varepsilon^2}{\theta}$. “(H_0 and H_1)” indicates that “SVM correction factor which is consistent under both H_0 and H_1 ” is used. Similarly, one could consider using the percentiles of the instrumental variable t test statistic multiplied by “SVM correction factor which is consistent under both H_0 and H_1 ” that are calculated by generating the data from a general ARMA(1,1) with $\phi = 1$ and an unrestricted σ^2 . However, by Theorem 2.7, the limiting value of this correction factor depends on the value of σ^2 which causes the percentiles of the test statistic to vary according to the true value of σ^2 and this is not practically useful.

Then, two other rejection criteria equipped with $\tau_{\mu,IV,SVM,n,\theta}^{rst}$ (H_0 and H_1) are put together with the ones presented in (2.26). Namely, reject H_0 if

$$\begin{aligned} t_{\mu,IV,SVM}(H_0) &< \tau_{\mu,IV,SVM,n,\theta}^{rst}(H_0 \text{ and } H_1) \\ t_{\mu,IV,SVM}(H_0 \text{ and } H_1) &< \tau_{\mu,IV,SVM,n,\theta}^{rst}(H_0 \text{ and } H_1) \end{aligned} \quad (2.29)$$

We also investigate the finite sampling distributions of the test statistics when the data are generated from a SVM with a given value of σ_η^2 and $\phi = 1$. From the formulas in equations (2.3) and (2.4), $\sigma_\eta^2 = \frac{(\theta-1)^2\pi^2}{2\theta}$. By exploiting this information, we obtain the value of σ_η^2 for various values of θ and generate η_t from $N(0, \sigma_\eta^2)$ for all $t = 1, \dots, n$. The data from SVM are generated accordingly. Then, another test criteria that we consider in addition to the ones in (2.26) and (2.29) are the ones that reject the unit root null when

$$\begin{aligned} n(\hat{\phi}_{IV} - 1) &< (n(\hat{\phi}_{IV} - 1))_{n,\theta}^{SVM} \\ t_{\mu,IV,Hall} &< \tau_{\mu,IV,Hall,n,\theta}^{SVM} \\ t_{\mu,IV,SVM}(H_0 \text{ and } H_1) &< \tau_{\mu,IV,SVM,n,\theta}^{SVM} \end{aligned} \quad (2.30)$$

where SVM on the superscript indicates that data are generated from SVM. Also, note that “SVM correction factor which is consistent under both H_0 and H_1 ” is used as a correction factor in computing $\tau_{\mu,IV,SVM,n,\theta}^{SVM}$ whereas “Hall’s correction factor” is used in computing $\tau_{\mu,IV,Hall,n,\theta}^{SVM}$.

In the following tables, we display the 5th and 10th percentiles of the finite sampling distributions of the test statistics $(n(\hat{\phi}_{IV} - 1))_{n,\theta}$, $\tau_{\mu,IV,Hall,n,\theta}$, $\tau_{\mu,IV,SVM,n,\theta}^{rst}$ (H_0 and H_1), $(n(\hat{\phi}_{IV} - 1))_{n,\theta}^{SVM}$, $\tau_{\mu,IV,Hall,n,\theta}^{SVM}$ and $\tau_{\mu,IV,SVM,n,\theta}^{SVM}$.

As we have done for the other test statistics we considered in the previous sections, we computed the test statistics, $(n(\hat{\phi}_{IV} - 1))_{n,\theta}$ and $\tau_{\mu,IV,Hall,n,\theta}$, for the data generated from an ARMA(1,1) model with mean zero, AR parameter with the value 1, and varying values of θ . The random error, v_t , is generated from $N(0, 1)$ for all t . In the data generation step for $\tau_{\mu,IV,SVM,n,\theta}^{rst}$ (H_0 and H_1), random errors of the ARMA(1,1), $\{v_t\}$, are generated independently and identically from $N\left(0, \frac{\pi^2}{2\theta}\right)$. We used different values for the MA parameter to see how the finite sampling properties of the test statistics respond to varying values of underlying MA parameter. In order to obtain the corresponding percentiles for $\tau_{\mu,IV,Hall,n,\theta}$ we use the third formula of the equations (2.25). In addition, to obtain the percentiles of $\tau_{\mu,IV,SVM,n,\theta}^{rst}$ (H_0 and H_1) we use the second of the equations given in (2.25). The number of Monte Carlo iterations is 10,000.

Table 2.11: Percentiles of Dickey-Fuller $n(\hat{\phi}_{ols} - 1)$ Test Statistic. *Reported from second part of Table 10.A.1 of Fuller(1996)*

n	100	500	1000
$\alpha = 0.05$	-13.70	-14.00	-14.10
0.10	-11.00	-11.20	-11.30

In Table 2.11, we display the percentiles of Dickey-Fuller $n(\hat{\phi}_{ols} - 1)$ test statistic. Comparing Tables 2.12 and 2.13 to Table 2.11, we see that convergence of the test

statistic $n(\hat{\phi}_{IV} - 1)$ to the limiting distribution of Dickey-Fuller test statistic $n(\hat{\phi}_{ols} - 1)$ is very slow for values of θ that are close to the value of ϕ under the null hypothesis. In addition, the outstanding jump in the percentiles of this test statistic from $\theta = 0.95$ to $\theta = 0.99$ holds the attention. These two characteristics might be explained by the results of Pantula (1991). He studies the limiting distributions of various unit root test statistics under an ARMA(1,1) model with the AR parameter ϕ being 1 as the MA parameter θ approaches to 1. He defines the MA parameter θ to be $1 - \frac{1}{n^\delta}$ where δ is the rate at which θ is approaching 1. One should note that θ is always fixed in all the cases. The idea of θ changing with n is to see how large a sample we need for the convergence to hold.

He derives the asymptotic distribution of various unit root test statistics for different values of δ . His results give us an idea about how large the sample should be to realistically use the asymptotic distribution, i.e. Dickey-Fuller limiting distribution, for different values of θ . The results imply that when θ is close to 1, convergence of the test statistic $n(\hat{\phi}_{IV} - 1)$ to the corresponding Dickey-Fuller limiting distribution is slower than the convergence of the ADF test statistic.

Table 2.12: Empirical 5th Percentiles for $(n(\hat{\phi}_{IV} - 1))_{n,\theta}$.

	n	100	500	1000
$\theta =$	0.00	-14.40	-14.31	-14.38
	0.10	-14.45	-14.28	-14.37
	0.20	-14.63	-14.29	-14.37
	0.30	-15.17	-14.35	-14.39
	0.40	-15.85	-14.59	-14.47
	0.50	-17.57	-15.23	-14.67
	0.60	-21.88	-16.80	-15.32
	0.70	-32.82	-21.26	-17.88
	0.80	-75.29	-35.94	-27.50
	0.85	-201.72	-57.23	-42.31
	0.90	-441.27	-117.70	-84.99
	0.95	-689.31	-575.39	-302.75
	0.99	-771.84	-3806.53	-6441.94

Table 2.13: Empirical 10th Percentiles for $(n(\hat{\phi}_{IV} - 1))_{n,\theta}$.

	n	100	500	1000
$\theta =$	0.00	-11.29	-11.47	-11.32
	0.10	-11.36	-11.51	-11.32
	0.20	-11.49	-11.53	-11.27
	0.30	-11.73	-11.54	-11.33
	0.40	-12.20	-11.64	-11.36
	0.50	-13.37	-11.93	-11.45
	0.60	-15.86	-12.66	-11.74
	0.70	-22.51	-15.17	-13.20
	0.80	-45.71	-24.25	-19.24
	0.85	-97.66	-38.11	-28.86
	0.90	-237.28	-78.09	-57.25
	0.95	-375.52	-324.39	-205.24
	0.99	-419.62	-2022.63	-3585.79

Table 2.14: Empirical 5th Percentiles for $\tau_{\mu,IV,Hall,n,\theta}$.

	n	100	500	1000
$\theta =$	0.00	-2.64	-2.82	-2.85
	0.10	-2.63	-2.82	-2.84
	0.20	-2.60	-2.81	-2.84
	0.30	-2.59	-2.82	-2.84
	0.40	-2.62	-2.85	-2.87
	0.50	-2.68	-2.96	-2.92
	0.60	-2.80	-3.17	-3.09
	0.70	-2.99	-3.62	-3.57
	0.80	-3.48	-4.37	-4.57
	0.85	-3.87	-5.06	-5.29
	0.90	-4.23	-6.26	-6.43
	0.95	-5.10	-6.53	-7.90
	0.99	-6.05	-5.01	-5.28

The finite sampling distribution of $\tau_{\mu,IV,Hall,n,\theta}$ draws attention as it has a distinguished behavior with respect to higher values of the MA parameter compared to the other t statistics we analyzed in the previous sections. This behavior is seen in Tables 2.14 and 2.15. The test statistics, $\tau_{\mu,(ADF/OLS),n,\theta}$, $\tau_{\mu,(ADF/WS),n,\theta}$, $\tau_{\mu,SDD,n,\theta}$, exhibit

Table 2.15: Empirical 10th Percentiles for $\tau_{\mu,IV,Hall,n,\theta}$.

	n	100	500	1000
$\theta =$	0.00	-2.39	-2.56	-2.57
	0.10	-2.37	-2.55	-2.56
	0.20	-2.35	-2.54	-2.55
	0.30	-2.34	-2.55	-2.55
	0.40	-2.36	-2.55	-2.55
	0.50	-2.40	-2.60	-2.59
	0.60	-2.48	-2.75	-2.70
	0.70	-2.61	-3.14	-3.04
	0.80	-2.84	-3.82	-3.95
	0.85	-3.07	-4.31	-4.62
	0.90	-3.33	-5.16	-5.46
	0.95	-3.92	-5.47	-6.54
	0.99	-4.65	-3.97	-4.20

Table 2.16: Empirical 5th Percentiles for $\tau_{\mu,IV,SVM,n,\theta}^{rst}$.

	n	100	500	1000
$\theta =$	0.10	-2.82	-2.87	-2.87
	0.20	-2.91	-2.90	-2.89
	0.30	-3.16	-2.96	-2.91
	0.40	-3.86	-3.10	-2.99
	0.50	-5.79	-3.62	-3.24
	0.60	-6.48	-5.74	-4.24
	0.70	-6.44	-7.30	-7.08
	0.80	-6.82	-7.44	-7.79
	0.85	-6.98	-7.88	-7.89
	0.90	-7.79	-8.81	-8.62
	0.95	-8.73	-11.01	-10.96
	0.99	-9.31	-13.84	-15.70

noticeably slow convergence when the true value of the MA parameter is real close to unity, e.g. for $\theta = 0.99$. On the other hand, the test statistic, $\tau_{\mu,IV,Hall,n,\theta}$, has this characteristic for true values of the MA parameters which are slightly away from unity also. In other words, the distribution of $\tau_{\mu,IV,Hall,n,\theta}$ is more sensitive to the values of θ compared to the other pivotal test statistics mentioned just above. For

Table 2.17: Empirical 10th Percentiles for $\tau_{\mu,IV, SVM, n, \theta}^{rst}$.

	n	100	500	1000
$\theta =$	0.10	-2.51	-2.57	-2.58
	0.20	-2.58	-2.58	-2.58
	0.30	-2.72	-2.62	-2.59
	0.40	-3.15	-2.71	-2.66
	0.50	-4.14	-3.01	-2.76
	0.60	-4.46	-4.07	-3.37
	0.70	-4.52	-5.05	-4.95
	0.80	-4.65	-5.06	-5.24
	0.85	-4.89	-5.26	-5.25
	0.90	-5.35	-5.96	-5.88
	0.95	-6.05	-7.61	-7.45
	0.99	-6.49	-9.48	-10.91

Table 2.18: Empirical 5th Percentiles for $(n(\hat{\phi}_{IV} - 1))_{n, \theta}^{SVM}$.

	n	100	500	1000
$\theta =$	0.10	-15.04	-14.32	-14.24
	0.20	-15.34	-14.26	-14.29
	0.30	-15.68	-14.19	-14.37
	0.40	-16.44	-14.44	-14.32
	0.50	-18.37	-14.95	-14.44
	0.60	-22.54	-16.49	-15.30
	0.70	-33.66	-21.09	-18.01
	0.80	-81.33	-36.36	-27.72
	0.85	-198.64	-57.84	-42.48
	0.90	-429.20	-119.87	-86.18
	0.95	-665.60	-526.08	-308.54
	0.99	-707.44	-3975.94	-6949.78

instance, if the true value of θ is 0.5, $\tau_{\mu,IV, Hall, n, \theta}$ doesn't seem to have converged to the limiting distribution yet even for large samples like $n = 1000$. We ran one more set of simulations to see the 5th and 10th percentiles of the finite sampling distribution of $\tau_{\mu,IV, Hall, n, \theta}$ when the sample size is very large. The sample size for this case is 10,000 and they come from an ARMA(1,1) process with zero mean, AR parameter 1, and MA parameter 0.80. The number of Monte Carlo samples is 1000. 5th and

Table 2.19: Empirical 10th Percentiles for $(n(\hat{\phi}_{IV} - 1))_{n,\theta}^{SVM}$.

	n	100	500	1000
$\theta =$	0.10	-11.85	-11.35	-11.33
	0.20	-11.90	-11.40	-11.38
	0.30	-12.08	-11.48	-11.40
	0.40	-12.43	-11.58	-11.50
	0.50	-13.73	-11.69	-11.58
	0.60	-16.45	-12.54	-11.96
	0.70	-23.21	-15.04	-13.42
	0.80	-48.73	-24.53	-19.36
	0.85	-100.18	-38.81	-29.16
	0.90	-238.09	-78.96	-58.30
	0.95	-373.34	-317.82	-210.57
	0.99	-394.41	-2094.53	-3709.71

Table 2.20: Empirical 5th Percentiles for $\tau_{\mu,IV,Hall,n,\theta}^{SVM}$.

	n	100	500	1000
$\theta =$	0.10	-2.64	-2.83	-2.86
	0.20	-2.62	-2.83	-2.85
	0.30	-2.67	-2.84	-2.84
	0.40	-2.65	-2.87	-2.86
	0.50	-2.71	-2.97	-2.93
	0.60	-2.81	-3.22	-3.11
	0.70	-2.98	-3.69	-3.60
	0.80	-3.45	-4.36	-4.61
	0.85	-3.80	-5.04	-5.35
	0.90	-4.26	-6.24	-6.52
	0.95	-5.09	-6.60	-7.99
	0.99	-5.96	-4.94	-5.26

10th percentiles of $\tau_{\mu,IV,Hall,n,\theta}$ in this case are -3.73 and -3.18 respectively. On the contrary, 5th and 10th percentiles of the distribution of $\tau_{\mu,ADF/OLS,n,\theta}$ is same as that of DF test statistic for $n = 1000$. The same is true for $\tau_{\mu,ADF/WS,n,\theta}$ and $\tau_{\mu,SDD,n,\theta}$. We see that convergence of the test statistic $\tau_{\mu,IV,Hall,n,\theta}$ to DF limiting distribution is comparatively slow even if the true value of the MA parameter is not very close to the true value of the AR parameter under the null model.

Table 2.21: Empirical 10th Percentiles for $\tau_{\mu,IV,Hall,n,\theta}^{SVM}$.

	n	100	500	1000
$\theta =$	0.10	-2.40	-2.54	-2.57
	0.20	-2.37	-2.54	-2.57
	0.30	-2.39	-2.54	-2.57
	0.40	-2.38	-2.55	-2.57
	0.50	-2.43	-2.61	-2.60
	0.60	-2.49	-2.78	-2.72
	0.70	-2.60	-3.17	-3.07
	0.80	-2.84	-3.85	-3.98
	0.85	-3.10	-4.32	-4.64
	0.90	-3.39	-5.20	-5.50
	0.95	-3.87	-5.46	-6.63
	0.99	-4.46	-3.93	-4.21

Table 2.22: Empirical 5th Percentiles for $\tau_{\mu,IV,SVM,n,\theta}^{SVM}$.

	n	100	500	1000
$\theta =$	0.10	-2.86	-2.89	-2.88
	0.20	-2.98	-2.91	-2.89
	0.30	-3.32	-3.00	-2.92
	0.40	-4.61	-3.22	-3.03
	0.50	-6.30	-4.22	-3.42
	0.60	-6.11	-6.39	-5.40
	0.70	-5.84	-6.43	-6.55
	0.80	-5.69	-6.09	-6.02
	0.85	-6.15	-6.37	-6.51
	0.90	-6.92	-7.10	-7.34
	0.95	-7.21	-9.05	-9.21
	0.99	-7.65	-11.38	-12.96

From Tables 2.16 and 2.17, we get an idea about the finite sampling distribution of $\tau_{\mu,IV,SVM,n,\theta}$ when the true model is a restricted ARMA(1,1) with $\phi = 1$. For a fixed θ , $\tau_{\mu,IV,SVM,n,\theta}^{rst}$ and $\tau_{\mu,IV,Hall,n,\theta}$ converge to the same distribution as n increases.

Tables 2.18-2.23 present the finite sampling distribution of the test statistics for given values of θ , when the data are generated from a SVM with $\phi = 1$. By comparing the Tables 2.18 and 2.19 to the Tables 2.12 and 2.13, we see that $n(\hat{\phi}_{IV} - 1)$ have

Table 2.23: Empirical 10th Percentiles for $\tau_{\mu,IV,SVM,n,\theta}^{SVM}$.

	n	100	500	1000
$\theta =$	0.10	-2.54	-2.57	-2.57
	0.20	-2.61	-2.58	-2.59
	0.30	-2.83	-2.64	-2.61
	0.40	-3.54	-2.79	-2.67
	0.50	-4.34	-3.33	-2.90
	0.60	-4.33	-4.47	-3.84
	0.70	-3.93	-4.41	-4.54
	0.80	-3.90	-4.13	-4.17
	0.85	-4.14	-4.32	-4.26
	0.90	-4.54	-4.80	-4.79
	0.95	-5.08	-6.04	-6.14
	0.99	-5.18	-7.73	-8.99

similar finite sampling properties for an ARMA(1,1) model with $\phi = 1$ and for an SVM with $\phi = 1$ for given values of θ . We observe a similar behaviour for the pivotal test statistic that uses “Hall’s correction factor” by comparing Tables 2.20 and 2.21 to the Tables 2.14 and 2.15. That is, it has similar finite sampling properties for these two models mentioned right above.

2.5 Choosing the Critical Points

Critical points for a particular unit root test that we discuss in this chapter are chosen from the empirical percentiles of the particular test statistic with respect to the sample size n and θ . Since the true value of θ is unknown, it is to be estimated from the data set. That is, we need to estimate the MA parameter of the ARMA(1,1) model that corresponds to the SVM we fit to the data set.

We consider three procedures to determine the critical point after having estimated θ . We call them Method 1, Method 2, and Method 3. Method 1 uses the tables constructed earlier in this chapter. Let us denote a consistent estimator of θ by $\hat{\theta}$. Let θ^* be the value of θ in our tables that is above $\hat{\theta}$ and is the closest to $\hat{\theta}$. Then,

the empirical percentile that corresponds to n and θ^* is selected as the critical point.

Method 2 requires smoothing of the empirical percentiles over the values of the MA parameter. Notice that we constructed the empirical percentile tables for discrete values of θ . In order to be able to predict the corresponding percentile for any given value of MA parameter, we smooth the percentiles over the values of θ for a given sample size. That is, for a fixed sample size, we fit a model on the percentiles which is nonlinear in θ . Details regarding the smoothing are given in Appendix B.

For Method 3, one needs to conduct a Monte Carlo simulation to obtain the percentils of the test statistic that corresponds to $\hat{\theta}$. Data sets should be generated from the model where $\hat{\theta}$ is the MA parameter of the model.

Method 1 is the procedure we use to determine the critical points in Monte Carlo power study.

2.6 Monte Carlo Power Study

In this section, we carry out an extensive power study on the frequentist unit root tests discussed earlier in this chapter to find out about how they perform when they applied to a set of mean corrected return data assumed to have a SVM structure.

We generate the return data $\{r_t\}$ from (2.1) (SVM). In order to generate the data $\{r_t\}$ from SVM, we take $\mu = -9$ and $\sigma_\eta^2 = 0.1$. These are the values that So and Li (1999) use in their simulation study. They fix μ at -9 in their data generation step since this value implies that the modal instantaneous volatility, $e^{\mu/2}$, is around 1.1% which is a reasonable number in daily returns data. The value of σ_η^2 is taken to be 0.1 because estimates of σ_η^2 are usually around this value in many real data studies. We generate the log-volatility at time $t = 1$, h_1 , from $N(\mu, \sigma_\eta^2)$ for $\phi = 1$ and from $N\left(\mu, \frac{\sigma_\eta^2}{1-\phi^2}\right)$ for $|\phi| < 1$. We consider different sample sizes changing from small to moderately large, namely, $n = 100, 500, \text{ and } 1000$. For the power study, we consider $\phi = 0.98$ and $\phi = 0.95$. Our empirical power study is based on 1000 Monte Carlo samples.

We calculate the test statistics, $t_{\mu,(ADF/OLS),p_{min}}$, $t_{\mu,(ADF/OLS),p_{full}}$, $t_{\mu,(ADF/WS),p_{full}}$, $t_{\mu,SDD}$, $t_{\mu,IV,Hall}$, $t_{\mu,IV,SVM}(H_0)$, $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$, and $n(\hat{\phi}_{IV} - 1)$ for the data sets generated from the SVM as specified above.

We make use of the Tables 2.3-2.10 and 2.12-2.19 in order to determine the critical values for the tests we consider.

In the tables below, the rows corresponding to $t_{\mu,(ADF/OLS),p_{min}}$, $t_{\mu,(ADF/OLS),p_{full}}$, $t_{\mu,(ADF/WS),p_{full}}$, and $t_{\mu,SDD}$, are obtained by the unit root tests that compare the values of these test statistics which are computed for the data set, $\log r_t^2$, against $\tau_{\mu,ADF/OLS,p_{min},n,\theta}$, $\tau_{\mu,ADF/OLS,p_{full},n,\theta}$, $\tau_{\mu,ADF/WS,p_{full},n,\theta}$, and $\tau_{\mu,SDD,n,\theta}$ respectively.

The rows corresponding to $t_{\mu,IV,Hall}$, $t_{\mu,IV,SVM}(H_0)$, and $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$ are obtained by comparing these test statistics computed for the data sets against $\tau_{\mu,IV,Hall,n,\theta}$. Similarly, the rows related with $n(\hat{\phi}_{IV} - 1)$ are computed by comparing

the value of this test statistic against $(n(\hat{\phi}_{IV} - 1))_{n,\theta}$.

The rows related with $t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$ are obtained by comparing the value of the test statistic, $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$, against $\tau_{\mu,IV,SVM,n,\theta}^{rst}(H_0 \text{ and } H_1)$.

The results corresponding to $t_{\mu,IV,Hall}^{(dgp:SVM)}$, $t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$, and $n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$ are obtained by comparing the test statistics $t_{\mu,IV,Hall}$, $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$, and $n(\hat{\phi}_{IV} - 1)$ against $\tau_{\mu,IV,Hall,n,\theta}^{SVM}$, $\tau_{\mu,IV,SVM,n,\theta}^{SVM}$, and $(n(\hat{\phi}_{IV} - 1))_{n,\theta}^{SVM}$ respectively.

In the following tables, columns with the heading DF contain the empirical levels of the tests when Dickey-Fuller cutoff points are used. These cutoff points are obtained from Fuller (1996). The empirical levels under the last three columns of each table here correspond to the levels of the tests when we use the empirical percentiles from sections 2.2., 2.3., and 2.4. as cutoff points. In Section 2.5., we mentioned two different methods of picking a critical value with respect to the value of the MA parameter. In the power study here, we apply the one which gets the critical values directly from the empirical tables constructed. This is what we call Method 1 in Section 2.5.

In order to be able to pick a critical value from the empirical percentile tables that we exhibit in the previous sections, one has to estimate θ . We consider three different methods to estimate θ . The first one we consider is Durbin's estimator of θ which is denoted by $\hat{\theta}_{durbin}$. We have explained this estimation procedure in Section 2.3. Our second type of estimator is what we call Durbin's adjusted estimator that is denoted by $\hat{\theta}_{adj. \text{ durbin}}$. This estimator is obtained by adding $2(\text{standard deviation of Durbin's estimator})$ to $\hat{\theta}_{durbin}$. The last one we consider is one-step Gauss-Newton estimator of θ that uses Durbin's estimator as a starting point and is denoted by $\hat{\theta}_{one \text{ step } GN}$. This estimation procedure is also explained in Section 2.3.

First of the last three columns of the following tables is constructed by using $\hat{\theta}_{durbin}$ to get a critical value by applying Method 1 of Section 2.5. Similarly, content of the next two columns are computed by using $\hat{\theta}_{adj. \text{ durbin}}$ and $\hat{\theta}_{one \text{ step } GN}$ respectively in choosing the critical points according to Method 1.

In Table 2.24, we display the average of the orders of ADF regression over 1000 Monte Carlo replications when data are generated from Model (2.1). That is, for each data set generated from a SVM, their log-squared transformation is taken, and a minimum order ADF model is fit as we know that log-squared return data has an autocovariance structure of an ARMA(1,1) model. Recall that in that procedure, the order of the AR model that approximates the ARMA(1,1) model is taken to be $p = \min(p_{AIC} + 2, [10n^{1/4}])$. According to the results displayed in the table, selecting

Table 2.24: Average of p over 1000 Monte Carlo Iterations. *Data are generated from SVM as in (2.1).*

	n	100	500	1000
$\phi =$	1.00	3.86	9.76	12.45
	0.98	3.56	7.95	10.59
	0.95	3.30	5.52	7.80

the order by this way tends to pick a smaller order when the data are from a SVM.

First of all we should remind that all the results layed out in this section are based on Method 1 of Section 2.5 to choose the critical points.

Note that, $\pm 2(\text{standard deviation})$ Monte Carlo intervals of the nominal levels 0.05 and 0.10 in 1000 Monte Carlo samples are approximately $[0.036, 0.063]$ and $[0.081, 0.120]$ respectively. Examining whether the empirical significance levels computed from the Monte Carlo method fall into the corresponding such intervals help us judge the test criteria considered here.

From the results of the extensive Monte Carlo power study we lay out in this section, we see that, for sufficiently large samples, the test that compares $t_{\mu, (ADF/OLS), p_{min}}$ against $\tau_{\mu, ADF/OLS, n, \theta, (p_{min})}$ maintains the level when our cutoffs are used. One should notice in these tables that the empirical level of the test that compares the value of the test statistic, $t_{\mu, (ADF/OLS), p_{min}}$, against the Dickey-Fuller cutoff point is much higher than the nominal level. This results in unreliable conclusions made regarding a unit-root in volatility when this test is used. Therefore one is better off

using $\tau_{\mu,ADF/OLS,n,\theta}(p_{min})$ as suggested here if one decides to apply small order ADF unit root test with OLS regression to a moderately large data set that is assumed to be coming from a SVM.

The level of the unit root test that compares the value of the test statistic $t_{\mu,(ADF/OLS),p_{full}}$ against $\tau_{\mu,ADF/OLS,n,\theta}(p_{full})$ maintains the level for sufficiently large samples like size $n = 500$ and $n = 1000$ when our cutoff points are used. For large data sets e.g. $n = 1000$, the test that compares $t_{\mu,(ADF/OLS),p_{full}}$ against the Dickey-Fuller cutoff point is also loyal to the nominal level. Therefore the overall conclusion regarding the maximum order ADF unit root tests with full order OLS regression is that one should use the critical points that are obtained with respect to the values of θ (by Method 1 of Section 2.5) rather than usual Dickey-Fuller critical points if the sample size is moderately large, e.g. $n = 500$. For larger samples, either one of them can be used. We observe that the unit root test based on Dickey-Fuller critical points has slightly higher power than the other one for large data sets.

According to what we observe from the tables, the test that compares the value of $t_{\mu,(ADF/WS),p_{full}}$ to our critical points can be confidently used for all sample sizes. The test that compares its value against Dickey-Fuller cutoff point should be used if sample size is sufficiently large, e.g. greater than 500. From the tables, we see that, for $n = 500$, the latter test has slightly higher power.

Another observation we make is that the unit root test statistic based on comparing $t_{\mu,SDD}$ against $\tau_{\mu,SDD,n,\theta}$ seems to have improved the size substantially for a small sample size like $n = 100$ compared to the test that uses Dickey-Fuller cutoff points for the same test statistic. However, it still doesn't quite maintain the level for that size of samples. This test keeps up with the nominal level for $n = 500$ when $\tau_{\mu,SDD,n,\theta}$ is used as a cutoff point. It maintains the nominal level for large sample sizes like $n = 1000$ when the Dickey-Fuller critical value as well as $\tau_{\mu,SDD,n,\theta}$ are used. In large samples, the test based on the Dickey-Fuller cutoff points has slightly higher power.

We also performed basically eight different tests based on the instrumental vari-

able estimators which we discussed in detail in Section 2.4. Two of them are based on $n(\hat{\phi}_{IV} - 1)$ and the rest are based on pivotal test statistics multiplied by some correction factors.

Comparing $t_{\mu,IV,Hall}$ against $\tau_{\mu,IV,Hall,n,\theta}$ maintains the level well.

The results that correspond to the row indicated by $t_{\mu,IV,Hall}^{dgp:SVM}$ are the results of the test that compares the value of the test statistic $t_{\mu,IV,Hall}$ against $\tau_{\mu,IV,Hall,n,\theta}^{SVM}$. This test has similar empirical significance levels and powers as the test that compares $t_{\mu,IV,Hall}$ against $\tau_{\mu,IV,Hall,n,\theta}$. This test also maintains the level for all the sample sizes we look at here.

Although the test that compares the value of $t_{\mu,IV,SVM}(H_0)$ against $\tau_{\mu,IV,Hall,n,\theta}$ appears to be maintaining the nominal level in most cases, it exhibits an odd power structure. That is, the power of this test is decreasing as the true value of ϕ is getting away from the value in the null hypothesis. This might be because the correction factor is computed from an estimator of σ_η^2 which is not consistent under the alternative model.

In the tables, the rows corresponding to $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$, $t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$, and $t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$ are obtained by comparing the value of the test statistic $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$ against three different cutoff values.

The first one that we look at is the row corresponding to $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$. As we show in Section 2.4, this t_{IV} test statistic multiplied by “SVM correction factor which is consistent under both H_0 and H_1 ” under Model(2.13) has the same limiting properties as the t_{IV} test statistic multiplied by “Hall’s correction factor” under a general ARIMA(0,1,1). That is, they converge to the same distribution as the sample size increases. This property is revealed in this finite sample study here. We see that the level is satisfied for $n=500$ and 1000 . These results are displayed on the row corresponding to $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$ in the tables.

The second one is the row corresponding to $t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$. In Section 2.4, we showed that, for the data generated from ARIMA(0,1,1) with the restriction as

specified above, “SVM correction factor which is consistent under both H_0 and H_1 ” and “Hall’s correction factor” converge to the same quantity in probability. That in turn implies that when the data are generated from this restricted ARIMA(0,1,1), the distribution of the test statistic t_{IV} multiplied by either one of these correction factors converges to the same distribution. Therefore, one could expect to see similar finite sampling results here in this study on the rows corresponding to $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$ and $t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$. However, as we see here, this situation is not mirrored in finite samples.

The last of the trio is the row corresponding to $t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$. Nominal level is maintained well for all the sample sizes we consider here when the value of the test statistic $t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$ is compared against $\tau_{\mu,IV,SVM,n,\theta}^{SVM}$. Notice that, in this one the data is generated exactly from a SVM to compute the empirical percentiles.

Another instrumental variable unit root test statistics we consider is the one that is based on $n(\hat{\phi}_{IV} - 1)$. Previously, we theoretically proved that $n(\hat{\phi}_{IV} - 1)$ when the true model is (2.13) converges to the same distribution as $n(\hat{\phi}_{IV} - 1)$ when the true model is (2.12). This result is not really mirrored in the tables.

The last one of the tests based on the instrumental variable estimators is the one that compares the value of the test statistic $n(\hat{\phi}_{IV} - 1)$ against the percentiles obtained by generating the data exactly from a SVM. One expects to see that this test maintains the level and yet this doesn’t seem to be the case here as far as the results in the tables are concerned. This is the affect of using Method 1 to determine the critical value. If one uses the percentile that is computed exactly for the estimate of θ as critical value instead of the one that is determined by Method 1, then the test will maintain the level. We performed a little simulation study to compute the percentiles of $(n(\hat{\phi}_{IV} - 1))_{n,\theta}^{SVM}$ by generating the data from a SVM with $\phi = 1$ and $\sigma_\eta^2 = \frac{(\theta-1)^2\pi^2}{2\theta}$ where we take θ to be 0.87 which is the true value of θ in this section. We see that using this percentile as the cutoff point, the test maintains the level. In

the following, the first element of the pair is the empirical level of the 0.05 level test whereas the second one is the empirical level of the 0.10 level test; for $n = 100$ it is (0.046, 0.087), for $n = 500$ it is (0.045, 0.085), and for $n = 1000$ it is (0.043, 0.092).

Comparing the test statistics based on the instrumental variable estimators against Dickey-Fuller cutoff points does not maintain the level for any of the sample sizes we consider here.

The overall picture regarding the different estimates of θ is that using $\hat{\theta}_{one\ step\ GN}$ produces conservative unit root tests especially when the sample size is large, e.g. $n = 1000$. Another conclusion concerning estimating θ to determine the critical points is that, estimation procedure of θ seems to be crucial in decision making.

Overall, instrumental variable unit root tests that maintain the level are not as powerful as the other test statistics which maintain the level.

For sufficiently large samples, we observe that, power of the tests based on $t_{\mu,SDD}$ are higher than that of the tests based on $t_{\mu,(ADF/OLS),p_{full}}$ and $t_{\mu,(ADF/WS),p_{full}}$.

Another conclusion that we get from the extensive Monte Carlo study conducted in this section is that one should use the test statistics $t_{\mu,(ADF/OLS),p_{full}}$, $t_{\mu,(ADF/OLS),p_{min}}$, $t_{\mu,(ADF/WS),p_{full}}$, and $t_{\mu,SDD}$ along with our cutoff points if the data set is moderately large, e.g. $n = 500$. For larger samples, one can use $t_{\mu,(ADF/OLS),p_{full}}$, $t_{\mu,(ADF/WS),p_{full}}$, and $t_{\mu,SDD}$ along with Dickey-Fuller cutoff points as well. If one has a small data set, e.g. $n = 100$, the most reasonable test to use is $t_{\mu,(ADF/WS),p_{full}}$ along with our cutoff values.

If one uses $t_{\mu,(ADF/OLS),p_{full}}$, $t_{\mu,(ADF/WS),p_{full}}$, or $t_{\mu,SDD}$ to test the unit root hypothesis in SV models, the critical points we suggest and Dickey-Fuller critical points can be used if the sample size is sufficiently large. Nevertheless, it would be a wise choice to use $\tau_{\mu,ADF/OLS,p_{full},n,\theta}$, $\tau_{\mu,ADF/WS,p_{full},n,\theta}$, and $\tau_{\mu,SDD,n,\theta}$ as critical points when testing for a unit root in SVM.

The conclusions regarding the unit root test statistics considered here are based on the method used here (Method 1) to choose the critical points from the percentile

tables.

In order to improve the tests, one can calculate the exact empirical percentile that corresponds to the estimate of θ instead of using Method 1 of Section 2.5.

Notice that, we obtain the critical values by generating the data from an ARMA(1,1) with $\phi = 1$. We compare the values of the test statistics calculated for the data from SVM against those critical values. Note that these two models have the same autocovariance structure and that is the only common property they have. Their likelihoods are different. From the tables, we see that, for sufficiently large samples, comparing the test statistic obtained from SVM against the percentile obtained from an ARMA model that has the same second order properties work in terms of maintaining the level.

Table 2.25: Empirical Significance Levels and Powers for 5% Level Test Criteria.
($n=100$ and 1,000 Monte Carlo replications)

ϕ	Test Statistic	DF	$\hat{\theta}_{durbin}$	$\hat{\theta}_{adj. durbin}$	$\hat{\theta}_{one step GN}$
1	$t_{\mu,(ADF/OLS),p_{min}}$	0.710	0.244	0.027	0.360
	$t_{\mu,(ADF/OLS),p_{full}}$	0.036	0.090	0.021	0.085
	$t_{\mu,(ADF/WS),p_{full}}$	0.049	0.032	0.019	0.031
	$t_{\mu,SDD}$	0.168	0.115	0.038	0.131
	$t_{\mu,IV,Hall}$	0.102	0.054	0.011	0.050
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.102	0.054	0.011	0.049
	$t_{\mu,IV,SVM}(H_0)$	0.046	0.030	0.015	0.031
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.173	0.125	0.062	0.124
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.173	0.113	0.155	0.127
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.173	0.049	0.031	0.047
	$n(\hat{\phi}_{IV} - 1)$	0.469	0.186	0.038	0.152
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.469	0.179	0.039	0.148
	0.98	$t_{\mu,(ADF/OLS),p_{min}}$	0.835	0.336	0.039
$t_{\mu,(ADF/OLS),p_{full}}$		0.046	0.119	0.033	0.124
$t_{\mu,(ADF/WS),p_{full}}$		0.091	0.071	0.033	0.069
$t_{\mu,SDD}$		0.236	0.162	0.055	0.189
$t_{\mu,IV,Hall}$		0.101	0.063	0.011	0.058
$t_{\mu,IV,Hall}^{(dgp:SVM)}$		0.101	0.064	0.011	0.056
$t_{\mu,IV,SVM}(H_0)$		0.047	0.039	0.016	0.038
$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$		0.190	0.183	0.120	0.184
$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$		0.190	0.131	0.149	0.139
$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$		0.190	0.070	0.048	0.069
$n(\hat{\phi}_{IV} - 1)$		0.534	0.236	0.049	0.216
$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$		0.534	0.233	0.050	0.214
0.95		$t_{\mu,(ADF/OLS),p_{min}}$	0.938	0.445	0.056
	$t_{\mu,(ADF/OLS),p_{full}}$	0.056	0.155	0.044	0.164
	$t_{\mu,(ADF/WS),p_{full}}$	0.123	0.080	0.043	0.083
	$t_{\mu,SDD}$	0.318	0.217	0.068	0.257
	$t_{\mu,IV,Hall}$	0.088	0.065	0.010	0.062
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.088	0.063	0.010	0.064
	$t_{\mu,IV,SVM}(H_0)$	0.055	0.043	0.017	0.041
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.216	0.199	0.125	0.199
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.216	0.141	0.123	0.138
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.216	0.068	0.043	0.068
	$n(\hat{\phi}_{IV} - 1)$	0.607	0.334	0.068	0.329
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.607	0.327	0.069	0.320

Table 2.26: Empirical Significance Levels and Powers for 5% Level Test Criteria.
($n=500$ and $1,000$ Monte Carlo replications)

ϕ	Test Statistic	DF	$\hat{\theta}_{durbin}$	$\hat{\theta}_{adj. durbin}$	$\hat{\theta}_{one step GN}$
1	$t_{\mu,(ADF/OLS),p_{min}}$	0.315	0.100	0.026	0.052
	$t_{\mu,(ADF/OLS),p_{full}}$	0.081	0.069	0.026	0.045
	$t_{\mu,(ADF/WS),p_{full}}$	0.048	0.044	0.041	0.044
	$t_{\mu,SDD}$	0.070	0.056	0.023	0.046
	$t_{\mu,IV,Hall}$	0.302	0.046	0.013	0.027
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.302	0.046	0.012	0.027
	$t_{\mu,IV,SVM}(H_0)$	0.141	0.048	0.032	0.036
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.178	0.078	0.051	0.060
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.178	0.037	0.026	0.031
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.178	0.047	0.035	0.040
	$n(\hat{\phi}_{IV} - 1)$	0.317	0.084	0.018	0.036
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.317	0.083	0.017	0.035
	0.98	$t_{\mu,(ADF/OLS),p_{min}}$	0.910	0.546	0.175
$t_{\mu,(ADF/OLS),p_{full}}$		0.257	0.223	0.096	0.181
$t_{\mu,(ADF/WS),p_{full}}$		0.286	0.283	0.269	0.275
$t_{\mu,SDD}$		0.402	0.330	0.152	0.278
$t_{\mu,IV,Hall}$		0.291	0.070	0.023	0.036
$t_{\mu,IV,Hall}^{(dgp:SVM)}$		0.291	0.070	0.023	0.036
$t_{\mu,IV,SVM}(H_0)$		0.125	0.030	0.018	0.021
$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$		0.260	0.131	0.088	0.106
$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$		0.260	0.055	0.041	0.052
$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$		0.260	0.082	0.059	0.074
$n(\hat{\phi}_{IV} - 1)$		0.502	0.303	0.101	0.204
$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$		0.502	0.302	0.095	0.197
0.95		$t_{\mu,(ADF/OLS),p_{min}}$	1.000	0.936	0.440
	$t_{\mu,(ADF/OLS),p_{full}}$	0.573	0.383	0.117	0.430
	$t_{\mu,(ADF/WS),p_{full}}$	0.671	0.503	0.250	0.461
	$t_{\mu,SDD}$	0.889	0.810	0.432	0.701
	$t_{\mu,IV,Hall}$	0.259	0.070	0.020	0.049
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.259	0.068	0.017	0.048
	$t_{\mu,IV,SVM}(H_0)$	0.060	0.008	0.004	0.008
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.294	0.176	0.112	0.152
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.294	0.071	0.039	0.063
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.294	0.104	0.065	0.095
	$n(\hat{\phi}_{IV} - 1)$	0.544	0.459	0.201	0.397
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.544	0.455	0.198	0.390

Table 2.27: Empirical Significance Levels and Powers for 5% Level Test Criteria.
 ($n=1000$ and $1,000$ Monte Carlo replications)

ϕ	Test Statistic	DF	$\hat{\theta}_{durbin}$	$\hat{\theta}_{adj. durbin}$	$\hat{\theta}_{one step GN}$
1	$t_{\mu,(ADF/OLS),p_{min}}$	0.220	0.065	0.024	0.033
	$t_{\mu,(ADF/OLS),p_{full}}$	0.064	0.049	0.028	0.037
	$t_{\mu,(ADF/WS),p_{full}}$	0.050	0.041	0.039	0.039
	$t_{\mu,SDD}$	0.052	0.037	0.021	0.028
	$t_{\mu,IV,Hall}$	0.345	0.064	0.023	0.023
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.345	0.057	0.019	0.021
	$t_{\mu,IV,SVM}(H_0)$	0.184	0.060	0.045	0.045
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.196	0.079	0.063	0.062
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.196	0.034	0.029	0.029
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.196	0.057	0.040	0.041
	$n(\hat{\phi}_{IV} - 1)$	0.266	0.073	0.020	0.021
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.266	0.073	0.019	0.020
0.98	$t_{\mu,(ADF/OLS),p_{min}}$	0.991	0.881	0.594	0.682
	$t_{\mu,(ADF/OLS),p_{full}}$	0.578	0.522	0.381	0.442
	$t_{\mu,(ADF/WS),p_{full}}$	0.729	0.709	0.686	0.699
	$t_{\mu,SDD}$	0.840	0.765	0.528	0.582
	$t_{\mu,IV,Hall}$	0.418	0.115	0.048	0.064
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.418	0.112	0.048	0.063
	$t_{\mu,IV,SVM}(H_0)$	0.220	0.055	0.038	0.039
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.318	0.162	0.121	0.130
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.318	0.091	0.081	0.085
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.318	0.118	0.106	0.107
	$n(\hat{\phi}_{IV} - 1)$	0.567	0.456	0.280	0.338
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.567	0.455	0.277	0.335
0.95	$t_{\mu,(ADF/OLS),p_{min}}$	1.000	1.000	0.849	0.985
	$t_{\mu,(ADF/OLS),p_{full}}$	0.936	0.913	0.753	0.867
	$t_{\mu,(ADF/WS),p_{full}}$	0.972	0.969	0.956	0.965
	$t_{\mu,SDD}$	0.999	0.995	0.845	0.982
	$t_{\mu,IV,Hall}$	0.362	0.100	0.047	0.072
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.362	0.098	0.045	0.070
	$t_{\mu,IV,SVM}(H_0)$	0.124	0.012	0.005	0.012
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.348	0.213	0.155	0.192
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.348	0.111	0.105	0.109
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.348	0.153	0.124	0.141
	$n(\hat{\phi}_{IV} - 1)$	0.599	0.552	0.430	0.523
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.599	0.552	0.430	0.523

Table 2.28: Empirical Significance Levels and Powers for 10% Level Test Criteria.
 ($n=100$ and $1,000$ Monte Carlo replications)

ϕ	Test Statistic	DF	$\hat{\theta}_{durbin}$	$\hat{\theta}_{adj. durbin}$	$\hat{\theta}_{one step GN}$
1	$t_{\mu,(ADF/OLS),p_{min}}$	0.770	0.314	0.036	0.425
	$t_{\mu,(ADF/OLS),p_{full}}$	0.060	0.143	0.040	0.138
	$t_{\mu,(ADF/WS),p_{full}}$	0.131	0.093	0.043	0.096
	$t_{\mu,SDD}$	0.214	0.160	0.057	0.177
	$t_{\mu,IV,Hall}$	0.133	0.096	0.030	0.093
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.133	0.097	0.030	0.093
	$t_{\mu,IV,SVM}(H_0)$	0.062	0.047	0.023	0.042
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.199	0.169	0.098	0.162
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.199	0.147	0.181	0.166
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.199	0.095	0.065	0.096
	$n(\hat{\phi}_{IV} - 1)$	0.498	0.250	0.062	0.214
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.498	0.246	0.063	0.209
	0.98	$t_{\mu,(ADF/OLS),p_{min}}$	0.881	0.416	0.055
$t_{\mu,(ADF/OLS),p_{full}}$		0.084	0.192	0.051	0.190
$t_{\mu,(ADF/WS),p_{full}}$		0.180	0.138	0.069	0.142
$t_{\mu,SDD}$		0.306	0.219	0.070	0.252
$t_{\mu,IV,Hall}$		0.130	0.108	0.038	0.108
$t_{\mu,IV,Hall}^{(dgp:SVM)}$		0.130	0.106	0.038	0.106
$t_{\mu,IV,SVM}(H_0)$		0.060	0.049	0.030	0.055
$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$		0.222	0.199	0.108	0.194
$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$		0.222	0.171	0.181	0.183
$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$		0.222	0.108	0.082	0.109
$n(\hat{\phi}_{IV} - 1)$		0.548	0.328	0.073	0.303
$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$		0.548	0.321	0.077	0.297
0.95		$t_{\mu,(ADF/OLS),p_{min}}$	0.959	0.529	0.086
	$t_{\mu,(ADF/OLS),p_{full}}$	0.103	0.229	0.065	0.239
	$t_{\mu,(ADF/WS),p_{full}}$	0.223	0.174	0.086	0.179
	$t_{\mu,SDD}$	0.408	0.290	0.093	0.323
	$t_{\mu,IV,Hall}$	0.117	0.102	0.030	0.096
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.117	0.100	0.033	0.094
	$t_{\mu,IV,SVM}(H_0)$	0.073	0.056	0.031	0.058
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.238	0.213	0.111	0.208
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.238	0.180	0.157	0.175
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.238	0.111	0.080	0.109
	$n(\hat{\phi}_{IV} - 1)$	0.615	0.419	0.109	0.409
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.615	0.416	0.109	0.404

Table 2.29: Empirical Significance Levels and Powers for 10% Level Test Criteria.
($n=500$ and 1,000 Monte Carlo replications)

ϕ	Test Statistic	DF	$\hat{\theta}_{durbin}$	$\hat{\theta}_{adj. durbin}$	$\hat{\theta}_{one step GN}$
1	$t_{\mu,(ADF/OLS),p_{min}}$	0.415	0.150	0.052	0.090
	$t_{\mu,(ADF/OLS),p_{full}}$	0.122	0.110	0.064	0.084
	$t_{\mu,(ADF/WS),p_{full}}$	0.095	0.092	0.080	0.090
	$t_{\mu,SDD}$	0.118	0.101	0.048	0.080
	$t_{\mu,IV,Hall}$	0.345	0.097	0.036	0.054
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.345	0.095	0.034	0.052
	$t_{\mu,IV,SVM}(H_0)$	0.171	0.069	0.043	0.054
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.208	0.103	0.074	0.079
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.208	0.069	0.047	0.057
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.208	0.101	0.078	0.083
	$n(\hat{\phi}_{IV} - 1)$	0.358	0.133	0.044	0.072
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.358	0.131	0.043	0.071
	0.98	$t_{\mu,(ADF/OLS),p_{min}}$	0.959	0.674	0.286
$t_{\mu,(ADF/OLS),p_{full}}$		0.388	0.360	0.197	0.309
$t_{\mu,(ADF/WS),p_{full}}$		0.490	0.477	0.451	0.463
$t_{\mu,SDD}$		0.573	0.481	0.237	0.391
$t_{\mu,IV,Hall}$		0.355	0.124	0.060	0.092
$t_{\mu,IV,Hall}^{(dgp:SVM)}$		0.355	0.120	0.060	0.090
$t_{\mu,IV,SVM}(H_0)$		0.157	0.047	0.024	0.031
$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$		0.289	0.164	0.124	0.140
$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$		0.289	0.121	0.086	0.106
$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$		0.289	0.156	0.127	0.141
$n(\hat{\phi}_{IV} - 1)$		0.510	0.391	0.172	0.291
$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$		0.510	0.389	0.169	0.288
0.95		$t_{\mu,(ADF/OLS),p_{min}}$	1.000	0.965	0.547
	$t_{\mu,(ADF/OLS),p_{full}}$	0.722	0.692	0.383	0.617
	$t_{\mu,(ADF/WS),p_{full}}$	0.846	0.836	0.706	0.821
	$t_{\mu,SDD}$	0.954	0.895	0.537	0.799
	$t_{\mu,IV,Hall}$	0.292	0.112	0.057	0.096
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.292	0.107	0.058	0.091
	$t_{\mu,IV,SVM}(H_0)$	0.089	0.015	0.006	0.010
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.320	0.212	0.160	0.196
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.320	0.148	0.099	0.138
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.320	0.202	0.151	0.197
	$n(\hat{\phi}_{IV} - 1)$	0.550	0.498	0.251	0.448
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.550	0.498	0.249	0.448

Table 2.30: Empirical Significance Levels and Powers for 10% Level Test Criteria.
($n=1000$ and 1,000 Monte Carlo replications)

ϕ	Test Statistic	DF	$\hat{\theta}_{durbin}$	$\hat{\theta}_{adj. durbin}$	$\hat{\theta}_{one step GN}$
1	$t_{\mu,(ADF/OLS),p_{min}}$	0.270	0.119	0.050	0.067
	$t_{\mu,(ADF/OLS),p_{full}}$	0.106	0.103	0.061	0.078
	$t_{\mu,(ADF/WS),p_{full}}$	0.097	0.093	0.090	0.092
	$t_{\mu,SDD}$	0.100	0.083	0.035	0.047
	$t_{\mu,IV,Hall}$	0.383	0.129	0.061	0.060
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.383	0.128	0.059	0.057
	$t_{\mu,IV,SVM}(H_0)$	0.207	0.080	0.057	0.061
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.218	0.093	0.080	0.076
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.218	0.078	0.070	0.067
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.218	0.104	0.089	0.089
	$n(\hat{\phi}_{IV} - 1)$	0.316	0.128	0.053	0.057
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.316	0.127	0.053	0.057
	0.98	$t_{\mu,(ADF/OLS),p_{min}}$	0.998	0.952	0.768
$t_{\mu,(ADF/OLS),p_{full}}$		0.745	0.721	0.557	0.650
$t_{\mu,(ADF/WS),p_{full}}$		0.879	0.864	0.848	0.854
$t_{\mu,SDD}$		0.947	0.912	0.710	0.745
$t_{\mu,IV,Hall}$		0.458	0.177	0.110	0.124
$t_{\mu,IV,Hall}^{(dgp:SVM)}$		0.458	0.174	0.102	0.116
$t_{\mu,IV,SVM}(H_0)$		0.264	0.081	0.056	0.062
$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$		0.342	0.198	0.156	0.171
$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$		0.342	0.154	0.133	0.140
$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$		0.342	0.218	0.184	0.200
$n(\hat{\phi}_{IV} - 1)$		0.577	0.506	0.390	0.434
$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$		0.577	0.506	0.385	0.432
0.95		$t_{\mu,(ADF/OLS),p_{min}}$	1.000	1.000	0.875
	$t_{\mu,(ADF/OLS),p_{full}}$	0.977	0.974	0.893	0.957
	$t_{\mu,(ADF/WS),p_{full}}$	0.993	0.993	0.988	0.992
	$t_{\mu,SDD}$	1.000	0.999	0.872	0.987
	$t_{\mu,IV,Hall}$	0.412	0.160	0.088	0.131
	$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.412	0.160	0.087	0.130
	$t_{\mu,IV,SVM}(H_0)$	0.171	0.022	0.010	0.022
	$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.367	0.250	0.193	0.229
	$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.367	0.198	0.168	0.190
	$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.367	0.259	0.229	0.245
	$n(\hat{\phi}_{IV} - 1)$	0.605	0.583	0.476	0.547
	$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.605	0.583	0.475	0.547

Chapter 3

Bayesian Unit-Root Tests in Stochastic Volatility Models

3.1 Introduction

Bayesian data analysis is becoming more and more appealing because of its flexibility in handling complex models with many parameters and/or missing observations. This is one of the reasons that we consider a Bayesian perspective for unit-root testing in analysis of the data generated from a SVM as the number of unobserved volatilities (h_t) in a SVM is equal to the sample size. To perform Bayesian data analysis, one needs to set up the full probability model, i.e. a joint model for the observations as well as parameters. This is typically done in two stages. First a conditional distribution of the observations given the unknown parameters is specified; then the marginal distributions of the parameters are specified. Bayesian statistical inference on these unknown quantities is obtained based on their conditional distributions after having observed the data. Similarly, a unit-root test is performed based on the posterior information that is obtained through Bayesian statistical analyses. In this chapter, we develop two different Bayesian unit-root tests in SVM. We test the null hypothesis $H_0 : \phi = 1$, against the alternative hypothesis $H_1 : \phi < 1$ within the SVM framework.

Let us denote the null model by M_0 and the model under the alternative by M_1 . Unlike the frequentist unit-root tests that we considered in the previous section, we perform Bayesian unit-root tests on the mean corrected return data itself instead of the log-squared data. We express the full SVM as a hierarchical Bayesian model. This makes it more convenient to comprehend the structure of the model;

$$\begin{aligned}
r_t|h_t &\sim N(0, e^{h_t}) \text{ for } t \geq 1, \\
h_t|h_{t-1}, \mu, \phi, \sigma_\eta^2 &\sim N(\mu(1-\phi) + \phi h_{t-1}, \sigma_\eta^2) \text{ for } t \geq 2, \\
h_1|\mu, \phi, \sigma_\eta^2 &\sim N(\mu, \sigma_{h_1}^2) \text{ where } \sigma_{h_1}^2 = \begin{cases} \frac{\sigma_\eta^2}{(1-\phi^2)} & \text{if } \phi < 1 \\ \sigma_\eta^2 & \text{if } \phi = 1 \end{cases} \\
(\mu, \phi, \sigma_\eta^2) &\sim f
\end{aligned} \tag{3.1}$$

where f is the prior density of $(\mu, \sigma_\eta^2, \phi)$ which reflects our prior beliefs about the unknown parameters. As the first step of the Bayesian data analysis, we will choose the prior densities for the model parameters in the first section of this chapter. The most noticeable part of this section is composed of the new prior for the persistence parameter ϕ that we introduce to the SVM literature. The joint posterior density of the parameters and the unobserved log-volatilities is given by,

$$f(\mu, \sigma_\eta^2, \phi, \underline{h}|\underline{r}) \propto f(r|\underline{h}, \mu, \sigma_\eta^2, \phi) \times f(\underline{h}, \mu, \sigma_\eta^2, \phi)$$

Inference about a parameter, say ϕ , is obtained from the marginal posterior distribution of ϕ which is obtained by integrating the above density with respect to μ , σ_η^2 , and \underline{h} . Interval estimates and testing for ϕ are based on

$$f(\phi|\underline{r}) \propto \int f(\underline{h}, \mu, \sigma_\eta^2, \phi|\underline{r}) d\mu d\sigma_\eta^2 d\underline{h}$$

Note that, Bayesian tests are based on the exact SV process rather than just matching another process up to second order. In this respect, these tests are different than the frequentist tests in the previous chapter.

The first one of the Bayesian test statistics we are using for unit-root testing in the SVM is based on the Bayes Factor (BF). Our method to compute the BF differs

from that of So and Li (1999) described in Section 1.3. For their method, see Section 1.3 in this thesis. For the second Bayesian unit-root test we consider here, we make use of a 95% Posterior Interval for ϕ . Calculation of the BF and construction of the posterior interval are done using Gibbs Sampling in a software called BUGS (Version 0.603 for unix systems). To understand how BUGS works, see Gilks and Spiegelhalter (1994).

3.2 Choice of Prior Densities and Introducing Mixed Prior Density for ϕ

We consider several prior densities for each of the SVM parameters. The main reason to consider more than one prior density for each parameter is to do a sensitivity analysis so that we can see how sensitive our Bayesian testing procedures are to the choice of the prior densities. The prior we use for μ , the population mean of the log-volatilities, is $N(\mu_0, \sigma_0^2)$ where μ_0 and σ_0^2 are assumed to be known quantities. We consider both informative and noninformative priors for μ . Noninformative prior is used to make the prior belief have a minimal role on the posterior distribution. For noninformative prior, we set the variance of this Normal distribution to be large. For our analysis we set it at 10^6 to approximate a diffuse prior. On the other hand, we use $N(0, 10)$ as an informative prior distribution for μ .

We employ the inverse-gamma prior distribution, $IG(a, b)$, for σ_η^2 , sampling variance of the random shocks in the log-volatility that is generated from an $AR(1)$ process. Like we do for μ , we consider both informative and noninformative priors for σ_η^2 . The variance of a random variable having the inverse-gamma distribution is $\frac{b^2}{(a-1)^2(a-2)}$ with $a > 2$. We set $a = 2 + 10^{-10}$ as in So and Li (1999) to obtain a noninformative prior for σ_η^2 . Therefore, the flat prior for σ_η^2 is approximated by $IG(2 + 10^{-10}, 0.1)$ where the hyperparameter b is fixed at 0.1. In addition, we use $IG(5, 0.1)$ as an informative prior density for σ_η^2 .

In the literature, continuous densities such as uniform, truncated normal, and beta distributions that are defined over the interval $(0,1)$ are used as prior densities for ϕ . We think that using a prior density that is defined on an interval that excludes the point 1 is not suitable for the statistical inference where one's main concern is testing $H_0 : \phi = 1$. As a prior density for ϕ , we suggest using a density that is defined on an interval that includes the point 1 or a density that assigns a positive mass to 1. In this thesis, we introduce a density that has a positive mass on the point 1 and use it in the Bayesian unit root tests. Assigning a density that is defined on an interval that includes 1 will be considered in a future study.

We propose a new prior what we call the *MIXED PRIOR* for ϕ . Let b be a Bernoulli random variable with success probability $p = P(b = 1)$. In addition, let U be a continuous random variable having support on the open interval $(0,1)$. U is assumed to be independent of b . We define the persistence parameter ϕ as

$$\phi = b + (1 - b)U \quad (3.2)$$

Here, p can be referred as the mixing probability. The mixing probability can be assumed to be a constant or a parameter. When it is taken to be a constant, the marginal prior density of ϕ is

$$f_\phi(\phi) = pI(\phi = 1) + (1 - p)f_U(\phi)I(0 < \phi < 1) \quad (3.3)$$

This is a normalized proper density. When p is assumed to be a parameter, we choose a $\text{Beta}(\alpha, \beta)$ distribution to be a prior distribution for it. Then,

$$\begin{aligned} f(\phi|p) &= pI(\phi = 1) + (1 - p)f_U(\phi)I(0 < \phi < 1) \quad \text{and} \\ f_\phi(\phi) &= \int f(\phi|p)f(p) dp \\ &= E[p]I(\phi = 1) + (1 - E[p])f_U(\phi)I(0 < \phi < 1) \\ &= \frac{\alpha}{\alpha + \beta}I(\phi = 1) + \frac{\beta}{\alpha + \beta}f_U(\phi)I(0 < \phi < 1) \end{aligned}$$

That is, the marginal distribution of ϕ is also a mixture distribution given in (3.3) with $p = \frac{\alpha}{\alpha + \beta}$.

Here, U can be assumed to have a Beta distribution defined over $(0,1)$ with parameters α_U and β_U . In this thesis, we consider the case where both α_U and β_U are equal to 1. That is, we take $U \sim U(0, 1)$.

Marginal distribution of ϕ can also be represented in a hierarchical form as follows: When p is assumed to be known;

$$\begin{aligned}\phi|b &\sim U(b, 1) \\ b &\sim \text{Bernoulli}(p)\end{aligned}$$

and, when p is assumed to be a parameter, the level that defines the distribution of p is added as the last level of the hierarchy. For example, we may assume that $p \sim \text{Beta}(\alpha, \beta)$.

In addition to the mixed density, we suggest flat priors that are defined on a region that includes the point one. Phillips (1994) argued that any prior density with a support that doesn't include the point one downweights the larger values of the AR parameter and this causes the posterior inference of the parameter to be biased towards the stationarity. Note that, he made this argument in a general sense, not specifically for SVM. He calls such priors informative priors as they do not include the nonstationarity values of the parameter. He suggests using Jeffrey's prior instead as an objective alternative.

In our study, we suggest the use of priors in the form of $U(a,b)$ where $b > 1$. For example, $U(-2,2)$ and $U(0,1+\epsilon)$. In the second one, $\epsilon > 0$ can be chosen arbitrarily. We choose ϵ such that the expected value of ϕ having this density is equal to the expected value of ϕ having a mixed density with mixing probability p . Note that, here, p is the prior probability that we assign to point one in the Mixed Density. In this case, $\epsilon = p$. Another alternative for ϵ is to choose it such that $P(1 < \phi < 1 + \epsilon) = p$. Here, this probability is taken over the $U(0,1+\epsilon)$. In this case, $\epsilon = \frac{p}{1-p}$.

3.3 Posterior Inference in SVM

3.3.1 Gibbs Sampling in SVM

As pointed out by many authors, the main difficulty in obtaining the joint posterior density of the parameters of SVM is that it requires an n-dimensional integration. The joint density of the parameters given the data is

$$f(\mu, \sigma_\eta^2, \phi | \underline{r}) = \int f(\mu, \sigma_\eta^2, \phi, \underline{h} | \underline{r}) d\underline{h}$$

As seen above, deriving the joint posterior density of the SVM parameters amounts to integrating out the unobserved log-volatilities \underline{h} which is difficult to perform both analytically and numerically as the number of unobserved log-volatilities is equal to the sample size. This makes obtaining posterior inference about the parameters very difficult. At this point, simulation based techniques provide an alternative approach to obtain posterior information on the parameters. Such techniques are called Markov Chain Monte Carlo (MCMC) methods. MCMC methods are being increasingly used for the cases where marginal distribution of the random variables can not be obtained either analytically or numerically. MCMC methods consist of algorithms to construct a Markov Chain of the parameters such that its stationary distribution is our distribution of interest, i.e the posterior distribution of the parameter of interest. That means, under some regularity conditions, the realization of this Markov Chain can be thought of as points sampled from the posterior distribution. MCMC methods have wide applications in different problems particularly in Bayesian analyses. A Bayesian statistician needs to obtain the marginal posterior density of the parameters to extract the properties of the posterior distribution such as mean and variance. Obtaining the marginal posterior density might be difficult since it requires multiple integration over the other parameters. One can use MCMC methods to draw samples from the posterior distribution and hence make a Bayesian inference based on mean, variance, and percentiles of the posterior distribution of the parameters. One way to

construct such a Markov chain is introduced by Hastings (1970). This algorithm is a generalization of a method proposed by Metropolis et al. (1953). Therefore this MCMC algorithm is called Metropolis-Hastings algorithm. For more information on MCMC techniques see, for example, Gilks et al. (1996). We use Gibbs Sampler (Gelfand and Smith, 1990), a widely used MCMC method, to simulate the parameters indirectly from the posterior density $f(\mu, \sigma_\eta^2, \phi | \underline{r})$. Gibbs Sampling algorithm is a special case of Metropolis-Hastings algorithm. The Gibbs Sampling algorithm in our SVM proceeds as follows: at each cycle k , points $\mu^{(k)}$, $\phi^{(k)}$, $\sigma_\eta^{2(k)}$, and $h_t^{(k)}$ ($t=1, \dots, n$) are sampled from their corresponding full conditional densities. By full conditional density, we mean the density of the parameter given the observed data and the rest of the parameters in the model. Let $f(\mu | \phi, \sigma_\eta^2, \underline{h}, \underline{r})$, $f(\sigma_\eta^2 | \mu, \phi, \underline{h}, \underline{r})$, $f(\phi | \mu, \sigma_\eta^2, \underline{h}, \underline{r})$, and $f(h_t | \mu, \phi, \sigma_\eta^2, \underline{h}_{-t}, \underline{r})$ ($t=1, \dots, n$) denote the full conditional densities of μ , σ_η^2 , ϕ , and h_t ($t=1, \dots, n$) in SVM respectively. Here \underline{h}_{-t} is the vector of unobserved log-volatilities excluding the t th one.

The full conditional densities of the parameters as well as h_t ($t=1, \dots, n$) are as follows;

$$\begin{aligned} \mu | \phi, \sigma_\eta^2, \underline{h}, \underline{r} &\sim N\left(\frac{c_2}{c_1}, \frac{\sigma_\eta^2}{c_1}\right) \\ \sigma_\eta^2 | \mu, \phi, \underline{h}, \underline{r} &\sim IG\left(\frac{n}{2} + a, \frac{c_3}{2} + b\right) \end{aligned} \tag{3.4}$$

$$f(h_t | \mu, \phi, \sigma_\eta^2, \underline{h}_{-t}, \underline{r}) \propto \begin{cases} f(r_t | h_t) f(h_t | \mu, \sigma_\eta^2, \phi) f(h_{t+1} | h_t, \mu, \sigma_\eta^2, \phi) & \text{if } t = 1 \\ f(r_t | h_t) f(h_t | h_{t-1}, \mu, \sigma_\eta^2, \phi) f(h_{t+1} | h_t, \mu, \sigma_\eta^2, \phi) & \text{if } 1 < t < n \\ f(r_t | h_t) f(h_t | h_{t-1}, \mu, \sigma_\eta^2, \phi) & \text{if } t = n \end{cases}$$

where

$$\begin{aligned} c_1 &= (1 - \phi^2) + (n - 1)(1 - \phi)^2 + \frac{\sigma_\eta^2}{\sigma_0^2} \\ c_2 &= h_1(1 - \phi^2) + (1 - \phi) \sum_{t=2}^n (h_t - \phi h_{t-1}) + \mu_0 \frac{\sigma_\eta^2}{\sigma_0^2} \end{aligned}$$

$$c_3 = (1 - \phi^2)(h_1 - \mu)^2 + \sum_{t=2}^n (h_t - \mu - \phi(h_{t-1} - \mu))^2 \quad (3.5)$$

Notice that, $f(r_t|h_t)$ is the probability density function of $N(0, e^{h_t})$.

Similarly, $f(h_1|\mu, \sigma_\eta^2, \phi)$, $f(h_t|h_{t-1}, \mu, \sigma_\eta^2, \phi)$, and $f(h_{t+1}|h_t, \mu, \sigma_\eta^2, \phi)$ are the probability density functions of $N(\mu, \sigma_\eta^2)$, $N(\mu(1-\phi) + \phi h_t, \sigma_\eta^2)$, and $N(\mu(1-\phi) + \phi h_{t-1}, \sigma_\eta^2)$ respectively.

If we use U(0,1) distribution for a prior on ϕ , the full conditional density of ϕ that should be considered along with the ones in (3.4) is

$$f(\phi|\mu, \sigma_\eta^2, \underline{h}, \underline{r}) \propto \sqrt{1 - \phi^2} e^{\left[\phi^2 \frac{((h_1 - \mu)^2 - \sum_{t=2}^n (h_{t-1} - \mu)^2)}{2\sigma_\eta^2} + \phi \frac{\sum_{t=2}^n (h_t - \mu)(h_{t-1} - \mu)}{\sigma_\eta^2} \right]} \quad (3.6)$$

When we use a mixed prior for ϕ with a constant mixing probability p , the additional full conditional densities corresponding to the ones shown in (3.4) are

$$f(\phi|\mu, \sigma_\eta^2, \underline{h}, \underline{r}, b) \propto \begin{cases} \left[\prod_{t=2}^n f(h_t|h_{t-1}, \mu, \sigma_\eta^2) \right] f(h_1|\mu, \sigma_{h_1}^2) p & \text{if } \phi = 1 \\ \left[\prod_{t=2}^n f(h_t|h_{t-1}, \mu, \phi, \sigma_\eta^2) \right] f(h_1|\mu, \phi, \sigma_{h_1}^2) (1 - p) & \text{if } \phi < 1 \end{cases} \quad (3.7)$$

Given an arbitrary choice of starting values for the parameters at the very beginning of the algorithm, namely $\mu^{(0)}$, $\phi^{(0)}$, $\sigma_\eta^{2(0)}$, and $h_t^{(0)}$ ($t=1, \dots, n$), Gibbs Sampling algorithm for cycle k is;

- Draw $h_t^{(k)}$ from $f(h_t|\mu^{(k-1)}, \phi^{(k-1)}, \sigma_\eta^{2(k-1)}, \underline{h}_{-t}^{(k-1)}, \underline{r})$ for $t=1, \dots, n$
- Draw $\mu^{(k)}$ from $f(\mu|\phi^{(k-1)}, \sigma_\eta^{2(k-1)}, \underline{h}^{(k)}, \underline{r})$
- Draw $\sigma_\eta^{2(k)}$ from $f(\sigma_\eta^2|\mu^{(k)}, \phi^{(k-1)}, \underline{h}^{(k)}, \underline{r})$
- Draw $\phi^{(k)}$ from $f(\phi|\mu^{(k)}, \sigma_\eta^{2(k-1)}, \underline{h}^{(k)}, \underline{r})$

Repeating the same process $(m+M)$ times, we get a Markov Chain whose stationary distribution is the $f(\mu, \sigma_\eta^2, \phi, \underline{h}|\underline{r})$ under the model M_1 .

We also need to lay out the full conditional densities under the null model, M_0 , as we will need them to compute the Bayes Factor in the next section. These are as follows;

$$\begin{aligned}\mu|\sigma_\eta^2, \underline{h}, \underline{r} &\sim N\left(\frac{d_2}{d_1}, \frac{\sigma_\eta^2}{d_1}\right) \\ \sigma_\eta^2|\mu, \underline{h}, \underline{r} &\sim IG\left(\frac{n}{2} + a, \frac{d_3}{2} + b\right)\end{aligned}\quad (3.8)$$

$$f(h_t|\mu, \sigma_\eta^2, \underline{h}_{-t}, \underline{r}) \propto \begin{cases} f(r_t|h_t)f(h_t|\mu, \sigma_\eta^2)f(h_{t+1}|h_t, \sigma_\eta^2) & \text{if } t = 1 \\ f(r_t|h_t)f(h_t|h_{t-1}, \sigma_\eta^2)f(h_{t+1}|h_t, \sigma_\eta^2) & \text{if } 1 < t < n \\ f(r_t|h_t)f(h_t|h_{t-1}, \sigma_\eta^2) & \text{if } t = n \end{cases}$$

in (3.8), d_1 , d_2 , and d_3 are as follows;

$$\begin{aligned}d_1 &= 1 + \frac{\sigma_\eta^2}{\sigma_0^2} \\ d_2 &= h_1 + \mu_0 \frac{\sigma_\eta^2}{\sigma_0^2} \\ d_3 &= (h_1 - \mu)^2 + \sum_{t=2}^n (h_t - h_{t-1})^2\end{aligned}$$

The following theorem ensures that we can use the Markov chain constructed to serve to our purpose as long as it has some specific properties that are stated below.

Ergodic Theorem: Let $\mathbf{X} = \{\mathbf{X}_0, \mathbf{X}_1, \dots\}$ be a Markov Chain. If \mathbf{X} is irreducible, positive recurrent, and aperiodic then, it is said that \mathbf{X} is *ergodic* and the following consequences hold;

(i) There exists a stationary distribution, $\pi(\cdot)$, such that the transition density, $f_{X_t|X_0}(y|x)$ converges to this stationary density $\pi(y)$ as $t \rightarrow \infty$ for all x and y .

(ii) Let $f(\cdot)$ be a function defined on \mathbf{X} . Then, $\frac{1}{N} \sum_{t=1}^N f(X_t) \xrightarrow{a.s.} E_\pi(f(X))$ where E_π is the expectation of $f(X)$ with respect to $\pi(\cdot)$.

The Markov Chain for ϕ that is constructed via BUGS is obviously irreducible as there is a positive probability going from one state (e.g. (0,1)) to another state (e.g. 1) in finite number of iterations. This feature of the chain for other parameters and unobserved log-volatilities, as well as ϕ , is easily seen in the BUGS output.

We know that the Metropolis-Hasting algorithm constructs a positive recurrent chain as long as the joint posterior density is a proper density. In Appendix A, we show that the joint posterior density of the parameters in the SVM is a proper density even when we use improper prior densities for the parameters μ and σ_η^2 . The Gibbs Sampling algorithm is a special case of the Metropolis-Hasting Algorithm. Thus the Markov Chain of the SVM parameters and the unobserved log-volatilities we construct via a Gibbs Sampling simulator is positive recurrent. In other words, once the point drawn is from some distribution, all the subsequent points are from the same distribution. Positive recurrence implies that a stationary distribution exists. If the chain is aperiodic, i.e. if the points in the chain are not oscillating between the two states, the transition probability that is specified by the full conditional densities converges to this stationary density. However, checking aperiodicity of the chain theoretically is not possible for most of the applications. Tackling this problem is done by using softwares like CODA (Best et al. 1995) and WinBUGS (Spiegelhalter et al. 1999). One can use those softwares to plot the chain and see if it is converging. If the chain is converging, it must be converging to the posterior density. We used CODA to plot the points in the chain for each of the SVM parameters and determined that after some burn-in point, the points in the chain are converging in distribution to the points from the target density. In Bayesian analysis, the target density is the posterior density. That ensures that, after a burn-in point, the points sampled via Gibbs Sampling algorithm are from the joint posterior density of the parameters.

3.3.2 Gibbs Sampling in SVM via BUGS

We carry out the Gibbs Sampling algorithm as proposed above by means of a software package known as BUGS (Spiegelhalter et al. (1996)). We suggest reading Spiegelhalter et al. (1996) to get a satisfactory insight to BUGS. Also, see Gilks et al. (1994).

Full conditional densities in BUGS are constructed by utilizing the idea of a graph-

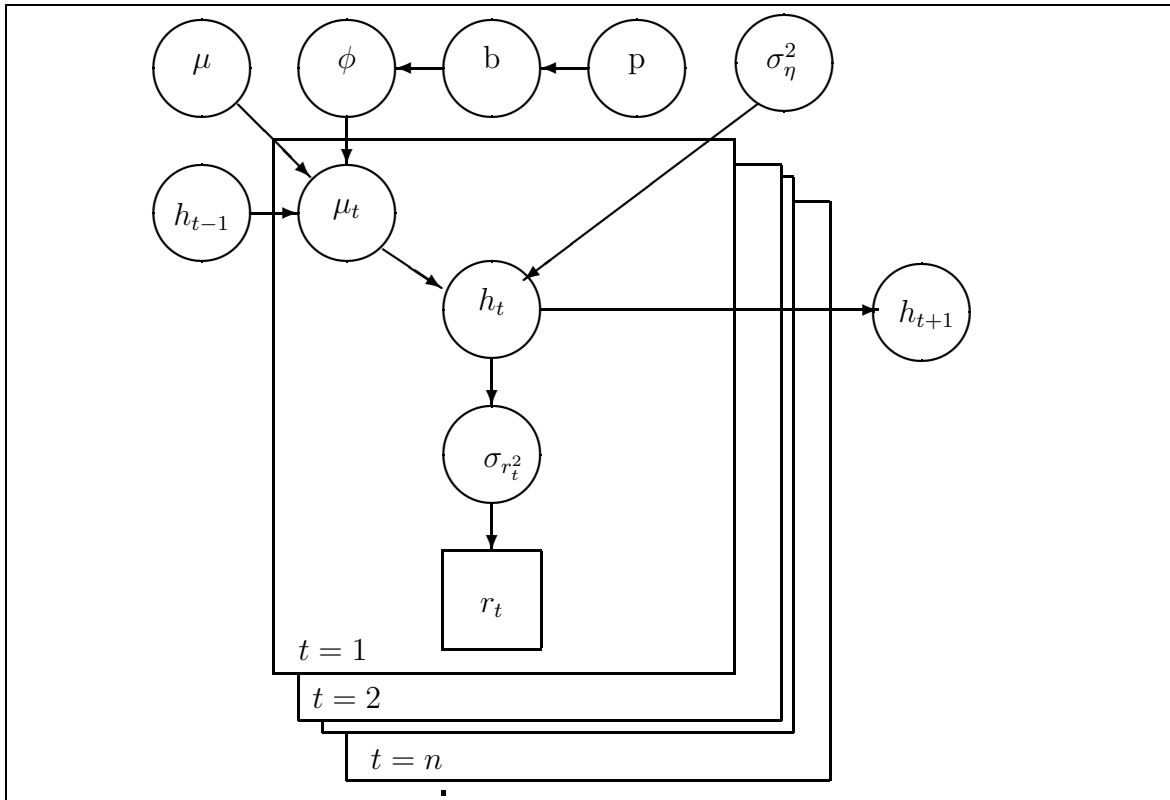


Figure 3.1: DAG representation of SVM with mixed density on ϕ .

ical representation of the model which is called a Directed Acyclic Graph (DAG). By using DAG, one can depict the full SVM in a graphical format. First, we explain how to draw the DAG of SVM and how to write down the full conditional densities of the parameters from the DAG of SVM. Then, we explain how the software BUGS draws on the model entered to obtain full conditional densities by using DAG logic.

In DAG, each quantity and variable in the model is represented as a node.

Each one of the squares and circles represents a node. A square node represents an observed quantity like a data point while a circular node represents an unobserved quantity such as a parameter or unobserved log-volatility. Similar figures are stacked up for as many observations as needed. See Figure 3.1. In DAG terminology, V usually denotes the set of all nodes. Each node in this set is denoted by v . In

addition, the predecessors of each node are called the parents of that node. Given its parents, each node is independent of all the other nodes except its descendants. This is the conditional independence.

Let V_{-v} denotes the subset of V which has all the nodes excluding the specific node v . Then, full conditional density of node v is

$$f(v|V_{-v}) \propto f(v|\text{parents of } v) \times \prod_{w \in [\text{parents of } v]} f(w|\text{parents of } w)$$

This formula when applied to the SVM, whose DAG representation for a mixed prior on ϕ is given in Figure 3.1, produces the full conditional densities in (3.4), (3.7). Obviously, it is not convenient to input the full model as a diagram into BUGS. The user enters a program that defines all the nodes as in DAG. BUGS compiles and executes the statements of the program, extracts the full conditional densities by a logic which is conceptually the same as writing them out from DAG, and performs the Gibbs Sampling algorithm. As one can see from (3.4), full conditional densities of μ and σ_η^2 are known distributions. Therefore, drawing samples from these distributions at each cycle of Gibbs Sampling is straightforward. Notice that the full conditional densities of the others are not standard known distributions. However, those full conditionals are log-concave as those densities are from an exponential family. BUGS employs Adaptive Rejection Sampling (ARS) method to sample from such distributions. For more information on ARS and other sampling methods implemented in BUGS, see Gilks et al. (1994) and (1996).

3.4 Bayesian Unit-Root Test Based on the Bayes Factor

Testing the null hypothesis $H_0 : \phi = 1$ versus the alternative hypothesis $H_1 : |\phi| < 1$ can be thought of as comparing the two proposed models with each other after having observed the data \underline{r} . In this sense the ratio of the posterior probabilities of the models, $\frac{Pr(H_0|\underline{r})}{Pr(H_1|\underline{r})}$, is considered as a suitable test statistic. This ratio is called Posterior Odds. We can conduct the Bayesian unit-root test in SVM based on the Posterior Odds. From Bayes's theorem, we obtain

$$Pr(H_i|\underline{r}) = \frac{Pr(\underline{r}|H_i)Pr(H_i)}{Pr(\underline{r}|H_0)Pr(H_0) + Pr(\underline{r}|H_1)Pr(H_1)} \quad \text{for } i=0,1$$

so that the Posterior Odds (PO) is

$$\frac{Pr(H_0|\underline{r})}{Pr(H_1|\underline{r})} = \frac{Pr(\underline{r}|H_0)Pr(H_0)}{Pr(\underline{r}|H_1)Pr(H_1)}$$

Here, $Pr(H_0)$ and $Pr(H_1)$ are the prior probabilities assigned to the models under the hypotheses H_0 and H_1 respectively and the related ratio is called Prior Odds. Also, the ratio of the marginal likelihoods of the observed returns under the model of H_0 and the model of H_1 is called Bayes Factor. Hence

$$\text{PO} = \text{Bayes Factor} \times \text{Prior Odds}$$

In the analyses, Prior Odds is usually taken to be 1 to indicate prior ignorance. As said earlier in this chapter, we consider two different prior distributions for ϕ one of which is a continuous density and the other one is a mixture of continuous and discrete densities. For the first one, we take the Prior Odds to be 1. Note that the second one, i.e. the mixture density, has a positive mass on the point 1. In this situation, from Section 2.2., $Pr(H_0) = Pr(\phi = 1) = P(b = 1) = p$ and $Pr(H_1) = Pr(0 < \phi < 1) = P(b = 0) = 1 - p$. Therefore, the Prior Odds is $\frac{p}{1-p}$ and so PO is $\left(\text{Bayes Factor} \times \frac{p}{1-p}\right)$.

Notice that, the Bayes Factor is the ratio of the Posterior Odds to the Prior Odds. It quantifies the amount of information gained from the data with respect to the prior information. It can also be viewed as the ratio of the marginal probability of the data under model H_0 to the one under H_1 . In this case, it has an objective interpretation as compared to Posterior Odds. Therefore, in Section 3.6, we will present results concerning Bayes Factor besides the Posterior Odds.

Let β_i denote the vector of the parameters under the model in H_i for $i=0,1$. Assuming that $h_0 = \mu$ both under the null and the alternative hypotheses, $\beta_0 = (\mu, \sigma_\eta^2)$ and $\beta_1 = (\mu, \sigma_\eta^2, \phi)$. For notational convenience, let us denote the marginal likelihood of the data \underline{r} under the model in H_i , $Pr(\underline{r}|H_i)$, by $m_i(\underline{r})$. Also, let us denote the Bayes Factor by BF. Then

$$BF = \frac{m_0(\underline{r})}{m_1(\underline{r})} \quad (3.9)$$

Our goal is to compute the Bayes Factor. We suggest two algorithms to estimate the marginal likelihoods $m_0(\underline{r})$ and $m_1(\underline{r})$.

Algorithm 1: The marginal likelihood of the observed mean corrected returns can be defined as

$$m_i(\underline{r}) = \frac{f(\underline{r}|\beta_i)f(\beta_i)}{f(\beta_i|\underline{r})} \quad \text{for } i=0,1 \quad (3.10)$$

where $f(\underline{r}|\beta_i)$ is the likelihood of observed returns conditional on the parameters under the model in H_i , $f(\beta_i)$ is the joint probability density function of the parameters of the model under H_i and $f(\beta_i|\underline{r})$ is the joint posterior density of the parameters under H_i . Finding the joint posterior density of the parameters amounts to integrating out the unobserved log-volatilities represented by \underline{h} . That is

$$f(\beta_i|\underline{r}) = \int f(\beta_i, \underline{h}|\underline{r})d\underline{h}$$

Note that the dimension of this integral is equal to the sample size. Therefore solving this integral is intractable. We can write the equation in (3.10) in such a way that

we can obtain an alternative expression for the marginal likelihood of the observed returns. Writing

$$\frac{f(\beta_{\mathbf{i}}|\underline{r})}{f(\underline{r}|\beta_{\mathbf{i}})} = \frac{f(\beta_{\mathbf{i}})}{m_i(\underline{r})}$$

and integrating both sides of the equation with respect to $\beta_{\mathbf{i}}$

$$\int \frac{1}{f(\underline{r}|\beta_{\mathbf{i}})} f(\beta_{\mathbf{i}}|\underline{r}) d\beta_{\mathbf{i}} = \int \frac{1}{m_i(\underline{r})} f(\beta_{\mathbf{i}}) d\beta_{\mathbf{i}}$$

we get

$$E_{\beta_{\mathbf{i}}|\underline{r}} \left(\frac{1}{f(\underline{r}|\beta_{\mathbf{i}})} \right) = \frac{1}{m_i(\underline{r})}$$

so that

$$m_i(\underline{r}) = \frac{1}{E_{\beta_{\mathbf{i}}|\underline{r}} \left(\frac{1}{f(\underline{r}|\beta_{\mathbf{i}})} \right)} \quad (3.11)$$

The expectation in the denominator is to be obtained by integrating $\frac{1}{f(\underline{r}|\beta_{\mathbf{i}})}$ over the posterior distribution of the model parameters given the data. We can estimate $m_i(\underline{r})$ by substituting this expectation in the denominator by its Monte Carlo integration estimate. In other words, we use the harmonic mean of $f(\underline{r}|\beta_{\mathbf{i}})$ as an estimate of this expectation. Then,

$$\hat{m}_i(\underline{r}) = \frac{1}{\frac{1}{M} \sum_{j=1}^M \frac{1}{f(\underline{r}|\beta_{\mathbf{i}}^{(j)})}} \quad (3.12)$$

where $\beta_{\mathbf{i}}^{(j)}$ are samples drawn from the joint posterior distribution of the parameters and M is a suitably chosen integer.

From the hierarchical expression in (3.1)

$$\begin{aligned} f(\underline{r}|\beta_{\mathbf{i}}^{(j)}) &= \int f(\underline{r}, \underline{h}|\beta_{\mathbf{i}}^{(j)}) d\underline{h} \\ &= \int f(\underline{r}|\underline{h}, \beta_{\mathbf{i}}^{(j)}) f(\underline{h}|\beta_{\mathbf{i}}^{(j)}) d\underline{h} \\ &= \int f(\underline{r}|\underline{h}) f(\underline{h}|\beta_{\mathbf{i}}^{(j)}) d\underline{h} \end{aligned}$$

The last equality is due to the fact that mean corrected returns are not dependent on the parameters in the presence of the log-volatilities. As seen from the chain of the equations right above, computing $f(\underline{r}|\beta_i^{(j)})$ in expression (3.12) boils down to integrating $f(\underline{r}|\underline{h})$ over the likelihood of the unobserved log-volatility \underline{h} . Again, this integration is hindered by the high-dimensionality problem. One can use the Monte Carlo integration estimate. The Monte Carlo integration estimate of this quantity is

$$\hat{f}(\underline{r}|\beta_i^{(j)}) = \frac{1}{N} \sum_{k=1}^N f(\underline{r}|\underline{h}^{(k)}) \quad (3.13)$$

where each of $\{\underline{h}^{(k)} : k = 1, \dots, N\}$ is a set of sample of log-volatilities simulated from $f(\underline{h}|\beta_i^{(j)})$ for a given value of $\beta_i^{(j)}$ and N is a suitably chosen integer. By plugging the estimator in (3.13) back into the expression in (3.12) we obtain

$$\hat{m}_i(\underline{r}) = \frac{1}{\frac{N}{M} \sum_{j=1}^M \frac{1}{\sum_{k=1}^N f(\underline{r}|\underline{h}^{(k)})}} \quad \text{for } i=0,1 \quad (3.14)$$

Steps to compute the estimate of the marginal likelihood of observed mean corrected returns as described above are laid out as follows;

- Draw $\beta_i^{(j)}$ ($j=1, \dots, M$) from $f(\beta_i|\underline{r})$
- For each $\beta_i^{(j)}$ ($j=1, \dots, M$), generate $\underline{h}^{(k)}$ ($k=1, \dots, N$) from $f(\underline{h}|\beta_i^{(j)})$ and compute $f(\underline{r}|\underline{h}^{(k)})$.

Algorithm 2: Note that,

$$\begin{aligned} f(\underline{h}, \beta_i|\underline{r})m_i(\underline{r}) &= f(\underline{h}, \beta_i, \underline{r}) \\ &= f(\underline{r}, \underline{h}|\beta_i)f(\beta_i) \\ &= f(\underline{r}|\underline{h}, \beta_i)f(\underline{h}|\beta_i)f(\beta_i) \\ &= f(\underline{r}|\underline{h})f(\underline{h}|\beta_i)f(\beta_i) \end{aligned}$$

From this expression, we have,

$$\frac{f(\underline{h}, \beta_i|\underline{r})}{f(\underline{r}|\underline{h})} = \frac{f(\underline{h}|\beta_i)f(\beta_i)}{m_i(\underline{r})}$$

Integrating both sides of the equation with respect to \underline{h} and β_i

$$\int \frac{1}{f(\underline{r}|\underline{h})} f(\underline{h}|\underline{r}) d\underline{h} = \frac{1}{m_i(\underline{r})}$$

gives us

$$E_{\underline{h}|\underline{r}} \left(\frac{1}{f(\underline{r}|\underline{h})} \right) = \frac{1}{m_i(\underline{r})}$$

so that

$$m_i(\underline{r}) = \frac{1}{E_{\underline{h}|\underline{r}} \left(\frac{1}{f(\underline{r}|\underline{h})} \right)} \quad (3.15)$$

Using the Monte Carlo integration estimate for the denominator in the expression (3.15), brings us an estimator of the marginal likelihood of the data. That is,

$$\hat{m}_i(\underline{r}) = \frac{1}{\frac{1}{M} \sum_{l=1}^M \frac{1}{f(\underline{r}|\underline{h}^{(l)})}} \quad (3.16)$$

can be used to estimate $m_i(\underline{r})$. Then, the steps to estimate $m_i(\underline{r})$ are presented as follows;

- Draw $\underline{h}^{(l)}$ ($l=1, \dots, M$), from $f(\underline{h}|\underline{r})$
- For each $\underline{h}^{(l)}$, compute $f(\underline{r}|\underline{h}^{(l)})$.

These algorithms suggested to estimate the marginal likelihood of the mean corrected returns as seen above require the sampling of the parameters, β_i , and the unobserved log-volatilities, \underline{h} , from their posterior densities. As pointed out earlier, obtaining $f(\beta_i|\underline{r})$ requires an n dimensional integration and therefore it is difficult to perform. Even if the integral were tractable to calculate, one would still have the problem of drawing samples from the resulting density. At this point, we make use of the Markov Chain which is constructed by the Gibbs Sampling algorithm as explained before.

The algorithm that So and Li (1999) use to compute BF in SVM and the algorithms we propose are different. Their algorithm is explained in the first chapter. The main difference between the procedure that So and Li (1999) follow to compute BF

in SVM and our methods is that they write the SVM in a linear state space form and approximate the exact distribution of $\log u_t^2$ by a mixture of Normal distributions to draw $\{h_t\}$. We don't need to write the SVM in that form and so don't need to make any approximation to the exact distribution.

In terms of the Monte Carlo simulations in Section 3.6, we investigate the performance of the test that uses Algorithm 2 as this algorithm is computationally more efficient than Algorithm 1.

3.5 Bayesian Unit-Root Test Based on the Posterior Interval

In frequentist analysis, one way to perform hypothesis testing is to use confidence intervals as an alternative to using test statistics. Similarly, one can employ the posterior interval of the parameter of interest to conduct hypothesis testing in Bayesian analysis. The posterior interval is constructed from the quantiles of the posterior distribution of the parameter. For instance, a 95% posterior interval is formed from 2.5th and 97.5th percentiles of the posterior distribution. In the earlier sections, we have explained how to obtain posterior inference about a parameter by simulations without actually deriving the posterior probability density functions analytically. Accordingly, a central 95% posterior interval is obtained from the corresponding quantiles of the Markov Chain that we construct via the Gibbs Sampling algorithm. A Bayesian unit-root test for SVM that we introduce in this section is carried out by checking whether the null hypothesis value of ϕ is included in the posterior interval or not.

When one uses a continuous density which is defined on an interval that doesn't include 1 as a prior density for ϕ , 1 can never get to be picked by the Gibbs Sampling simulator. Therefore, any Markov Chain constructed that way doesn't contain 1. This situation implies that 1 is never included in the posterior interval. However, this problem is avoided by using the *MIXED PRIOR* that we introduced with a

reasonably large mass on the point 1. The reason is that, there is a positive mass of the posterior density of ϕ at the point 1, if prior used has a positive probability on 1. This is proved in Appendix A.

3.6 Monte Carlo Power Study

In this section, we display the results of the simulation experiment we conducted in order to determine the powers of the BF unit root test and posterior interval unit root test we introduce in SVM. In order to perform the simulation experiments, we simulate the data from the model (2.1) by using the same parameter values as in the power study of Chapter 2. In other words, each data set throughout the simulation experiments here is generated with volatility mean $\mu = -9$ and variance of the shocks to volatility $\sigma_\eta^2 = 0.1$. These are the values that are used by So and Li (1999) in their Monte Carlo experiment. We use diffuse densities as prior densities for μ and σ_η^2 for our experiment related with BF whereas we use both diffuse and informative prior densities for these parameters for the experiment related with the posterior interval. For ϕ , we consider mixed density with a fixed mixing probability, $U(-2, 2)$, and a $U(0, 1)$ density as prior densities in BF experiments. We consider two types of mixed densities for the simulation study for the unit root test based on the posterior interval; mixed density with a fixed mixing probability and a mixed density with a random mixing probability. The mixed distribution we consider is the one that is mixed with $U(0, 1)$. For this unit root testing procedure, we also consider $U(0, 1 + \epsilon)$ for ϕ besides the uniform densities stated above. For one case, we take $\epsilon = p = 0.95$ and for another case we take $\epsilon = \frac{p}{1-p} = 19$ given that $p = 0.95$. Here, p is the probability that the mixed prior density has on the point one. We compute the number of correct decisions by using SAS and BUGS together; Gibbs Sampling is performed via BUGS whereas BF calculations are performed by SAS. Recall that, m is used to denote the burn-in point whereas M is used to denote the number of Markov Chain points that

are used to make Bayesian inference about the parameters after the burn-in point. We use the plot of the points in the chain for each parameter in order to determine the burn-in point. We started from three different set of starting points, obtained these three parallel chains and plotted them on the same figure to see the point where all the three chains seem to have forgotten their starting point. This point is taken to be the burn in point. Another method to determine the burn-in point is Gelman-Rubin test. This test is available in CODA and that's what we used to apply Gelman-Rubin test on our BUGS output. See Gilks et al. (1996) for details on Gelman-Rubin test. For this study, by using those two methods explained above, we decided to take the burn-in point to be 1000. That means, first 1000 iterations are discarded. In addition, we needed to determine the number of points sampled that should be used in estimating the posterior quantities. There are several ways in the MCMC literature to come up with this number. We used Raftery-Lewis (1992) test that is available in CODA. Please, see Gilks et al. (1996) for more information regarding this test. According to the test results, we decided to take 5000 iterations after the burn-in point to use for the Bayesian inference.

In Table 3.1, we display the average value of $\log_{10}(PO)$ over 500 Monte Carlo replications as well as the proportion of correct decisions in the same 500 Monte Carlo replications. When we use a continuous density as the prior density for ϕ , $\log_{10}(PO) = \log_{10}(BF)$. When we use Mixed Density with a fixed mixing probability p as the prior density for ϕ , $\log_{10}(PO) = \log_{10}(BF) + \log_{10}(p) - \log_{10}(1-p)$ where $p = P(\phi = 1)$ is taken to be 0.95 and 0.5. The test based on PO rejects if $\log_{10}(PO) < 0$. Similarly, the test based on BF rejects the unit root null if $\log_{10}(BF) < 0$.

Note that, the algorithms we presented in Section 3.4 are useful for the cases where the data set is small, e.g. $n = 100$. Computing the density $f(\underline{r}|\underline{h})$ in the second steps of Algorithm 1 and Algorithm 2 are likely to have some computational problems for large data sets. Therefore, we didn't consider sample sizes like $n = 500$ and $n = 1000$ in the construction of Table 3.1.

Table 3.1: Average Values of $\log_{10}(PO)$ and $\log_{10}(BF)$ and the Proportion of Correct Decisions (PCD). (*Sample Size $n=100$, 500 replications*).

ϕ	Prior for ϕ	$\log_{10}(PO)$	PCD	$\log_{10}(BF)$	PCD
1	Mixed with U(0,1) ($p = 0.95$)	1.54 (0.92)	0.97	0.27 (0.92)	0.62
	Mixed with U(0,1) ($p = 0.5$)	0.23 (0.93)	0.57	0.23 (0.93)	0.57
	U(-2,2)	-0.45 (0.91)	0.26	-0.45 (0.91)	0.26
	U(0,1)	0.68 (1.02)	0.75	0.68 (1.02)	0.75
0.98	Mixed with U(0,1) ($p = 0.95$)	1.62 (0.93)	0.03	0.35 (0.93)	0.37
	Mixed with U(0,1) ($p = 0.5$)	0.47 (0.96)	0.32	0.47 (0.96)	0.32
	U(-2,2)	-0.90 (0.82)	0.89	-0.90 (0.82)	0.89
	U(0,1)	0.76 (1.06)	0.24	0.76 (1.06)	0.24
0.95	Mixed with U(0,1) ($p = 0.95$)	1.70 (0.92)	0.03	0.42 (0.92)	0.32
	Mixed with U(0,1) ($p = 0.5$)	0.47 (0.95)	0.31	0.47 (0.95)	0.31
	U(-2,2)	-1.31 (0.82)	0.98	-1.31 (0.82)	0.98
	U(0,1)	0.95 (1.06)	0.17	0.95 (1.06)	0.17

(Monte Carlo standard errors of $\log_{10}(PO)$ and $\log_{10}(BF)$ are given in the parantheses.)

From the results in Table 3.1, we see that, in all the cases shown here, Monte Carlo standard errors of $\log_{10}(PO)$ for which the prior used for ϕ is a mixed prior or U(-2,2) are less than that of the case where U(0,1) is used as a prior density for ϕ . We observe a similar behavior for $\log_{10}(BF)$. Notice that, the proportion of correct decisions made by the test based on $\log_{10}(PO)$ is remarkably small if the true value of ϕ is 0.98 or 0.95. In this case, one might wonder if a mixed prior with a smaller p would work better in terms of the proportion of the correct decisions. To get an idea

about this, we did a simulation study where we take the mixing probability, p , to be 0.5. As we see in Table 3.1, using a prior with a smaller p increases the power at the cost of a high Type 1 error. Also, note that, the prior $U(-2,2)$ perform significantly well in terms of the Type 2 error at the cost of Type 1 error.

Another conclusion that Table 3.1 brings our attention is that Bayesian unit root test based on $\log_{10}(PO)$ in SVM is more sensitive to the unit root than the test based on $\log_{10}(BF)$. However, the total error rates committed by both of the tests are quite high. This issue is considered further in Chapter 5.

Table 3.2: Proportion of Correct Decisions (PCD) of the Bayesian Unit Root Test based on The Posterior Interval. (*Sample Size $n=100$, 500 replications*).

Prior for ϕ	ϕ	PCD
Mixed with $U(0,1)$ ($p = 0.95$)	1	0.70
	0.98	0.39
	0.95	0.60
Mixed with $U(0,1)$ ($p = 0.50$)	1	0.41
	0.98	0.78
	0.95	0.94
$U(-2,2)$	1	0.87
	0.98	0.19
	0.95	0.87
$U(0,1)$	1	0.00
	0.98	1.00
	0.95	1.00

(Prior for μ is $N(0, 10^6)$ and prior for σ_η^2 is $IG(2 + 10^{-10}, 0.1)$)

In Table 3.2, we present the proportion of correct decisions made by the posterior interval testing criteria for testing $\phi = 1$. In posterior interval test, as indicated in Section 3.5, the decision is based on whether the null hypothesis value of ϕ is included in the posterior interval or not. When the true value of ϕ is 1, the number of correct decisions is simply the number of samples for which the point 1 is covered by the posterior interval of ϕ . When the true value of ϕ is 0.98 or 0.95, the number of correct decisions is the number of posterior intervals that do not include 1.

By comparing the proportion of correct decisions in Table 3.1 that correspond to mixed prior versus the ones in Table 3.2, we can get an idea which Bayesian unit root testing procedure gives more reliable results in small samples in SVM when we use a mixed prior for ϕ . The proportion of correct decisions made by the Bayesian test based on $\log_{10}(PO)$ is much smaller than that of the test which is based on the posterior interval when the data is generated with $\phi = 0.95$. However, in terms of sensitivity of the test to the unit root, the test based on $\log_{10}(PO)$ performs better. Overall reliability of these tests for small samples in SVM are further discussed in Chapter 5.

In Table 3.3, we compare the power of the Bayesian unit root testing criteria of So and Li to our test that is based on the posterior interval of ϕ . So and Li use four different continuous densities as a prior density for ϕ in their Monte Carlo study. Also, they use a flat Normal density and a flat Inverse Gamma density as prior densities for μ and σ_{η}^2 respectively. The results that correspond to the continuous priors for ϕ are reported from their Monte Carlo study. In the simulation study that we carry out for the posterior interval unit root test in SVM, we consider the following prior densities for ϕ ; mixed prior density, and flat densities that are defined on both stationary and nonstationary values of ϕ . The decision of So and Li about the unit root in volatility is based on whether $\log_{10}(PO)$ is greater than 0 or not. In their case, $\log_{10}(PO) = \log_{10}(BF)$ where their method to calculate BF is described in Section 1.3. For their priors, the prior odds is 1.

In Table 3.3, we see that, when the true value of ϕ is 1, number of correct decisions made based on the test criteria of So and Li decreases as the sample size increases in all the cases except for the Truncated Normal (TN) case. In the unit root models, especially when the sample size is moderately large e.g. 1000, correct selection rates of our tests are higher than theirs.

The mixed priors which have a high probability on the point $\phi = 1$ seem to perform better as compared to the ones that have a lower probability on the point

Table 3.3: Number of Correct Decisions. (*100 replications*).

ϕ	Prior Density	$n = 500$	$n = 1000$
1	U(0,1)	96	90
	TN(0.9, 0.05 ²)	88	94
	Beta(10,1)	86	75
	Beta(20,2)	85	84
	U(0,1.95)	93	93
	U(0,20)	93	93
	U(-2,2)	93	94
	Mixed with U(0,1), $p = 0.95$	94	98
	Mixed* with U(0,1), $p = 0.95$	96	97
	Mixed with U(0,1), $p \sim \text{Beta}(0.5, 0.5)$	79	86
	Mixed with U(0,1), $p \sim \text{Beta}(2.66, 0.14)$	95	97
0.98	U(0,1)	36	64
	TN(0.9, 0.05 ²)	44	67
	Beta(10,1)	60	66
	Beta(20,2)	50	80
	U(0,1.95)	43	86
	U(0,20)	42	85
	U(-2,2)	47	87
	Mixed with U(0,1), $p = 0.95$	41	77
	Mixed* with U(0,1), $p = 0.95$	25	62
	Mixed with U(0,1), $p \sim \text{Beta}(0.5, 0.5)$	77	93
	Mixed with U(0,1), $p \sim \text{Beta}(2.66, 0.14)$	42	82
0.95	U(0,1)	82	98
	TN(0.9, 0.05 ²)	87	98
	Beta(10,1)	89	97
	Beta(20,2)	93	98
	U(0,1.95)	94	99
	U(0,20)	92	99
	U(-2,2)	94	99
	Mixed with U(0,1), $p = 0.95$	89	98
	Mixed* with U(0,1), $p = 0.95$	69	96
	Mixed with U(0,1), $p \sim \text{Beta}(0.5, 0.5)$	97	99
	Mixed with U(0,1), $p \sim \text{Beta}(2.66, 0.14)$	85	96

(Mixed* indicates that the priors on μ and σ_n^2 are informative.)

$\phi = 1$. Based on our simulation study, we suggest that practitioners assign a high probability to the value to be tested in order to get high power. We also checked the

correct decisions made by the posterior interval unit root test when the true value of ϕ is 0.7 and 0.5. We looked at these cases when we used a mixed prior with a mixing probability fixed at 0.95 and $n=1000$. For both studies, number of correct decisions out of 100 Monte Carlo replications came out to be 100.

Results obtained by using a uniform density that includes the nonstationary values of ϕ are very impressive. Note that, Type 2 errors committed by the tests that use such prior densities are a little bit less than that of mixed priors in general. We note that, the performance of different procedures with priors $U(-2,2)$, $U(0,1.95)$, and $U(0,20)$ is similar.

Table 3.3 demonstrates the advantages of our unit root testing procedure based on the posterior interval of ϕ along with the mixed prior we introduced and continuous priors that include the nonstationary values of ϕ . From the Table 3.3, we see that, in general, So and Li's test yields high power at the cost of high Type 1 Error. We observe similar behavior for the posterior interval unit root test if the mixing probability p is taken to be random with the distribution $Beta(0.5, 0.5)$. The posterior interval testing procedure with mixed prior having higher probability on the point 1 provides reliable results for a unit root test in SVM. We have the similar conclusion for the cases where a continuous prior that includes nonstationary region as well as stationary region for ϕ is used. Besides, computing the posterior intervals is very straightforward and is not computationally intensive. A software in the WinBUGS environment is available from the author.

3.7 Some Other Simulation Results

In this section, we display some simulation results that are not our primary interest in testing for a unit root in SV models.

3.7.1 Some distributional properties of the parameters after having observed the data

In the tables below, $\hat{\lambda}_{mean}$, $\hat{\lambda}_{std.err.}$, $\hat{\lambda}_{lower}$ and $\hat{\lambda}_{upper}$ represent respectively the average of the posterior mean estimates, the average of the posterior variance estimates, the average of the lower bound estimate of the posterior intervals, and the average of the upper bound estimates of the posterior intervals of the parameters over 100 Monte Carlo Samples. In other words,

$$\begin{aligned}\hat{\lambda}_{mean} &= \frac{1}{100} \sum_{k=1}^{100} \hat{E}[\beta_{1(j)}|\mathcal{L}_k] \\ \hat{\lambda}_{std.err.} &= \frac{1}{100} \sum_{k=1}^{100} \left(\hat{\text{Var}}[\beta_{1(j)}|\mathcal{L}_k] \right)^{1/2} \\ \hat{\lambda}_{lower} &= \frac{1}{100} \sum_{k=1}^{100} \hat{F}_{\beta_{1(j)}|\mathcal{L}_k}^{-1}(0.025) \\ \hat{\lambda}_{upper} &= \frac{1}{100} \sum_{k=1}^{100} \hat{F}_{\beta_{1(j)}|\mathcal{L}_k}^{-1}(0.975)\end{aligned}$$

for all $j=1,2,3$, where $\beta_1 = (\mu, \sigma_\eta, \phi)$ and $\beta_{1(j)}$ is the j th element of the parameter vector β_1 . Also, note that

$$\begin{aligned}\hat{E}[\beta_{1(j)}|\mathcal{L}] &= \frac{1}{M} \sum_{i=m+1}^{m+M} \beta_{1(j)}^{(i)} \\ \hat{\text{Var}}[\beta_{1(j)}|\mathcal{L}] &= \frac{1}{M-1} \sum_{i=m+1}^{m+M} \left(\beta_{1(j)}^{(i)} - \hat{E}[\beta_{1(j)}|\mathcal{L}] \right)^2\end{aligned}$$

where $\beta_{1(j)}^{(i)}$ is the i th element of the Markov Chain that is used for the inferential purposes after the burn-in point. In addition, $\hat{F}_{\beta_{1(j)}|\mathcal{L}_k}^{-1}(x)$ is the x th percentile of the Markov Chain of the parameter $\beta_{1(j)}$ after the burn-in point.

Irreducibility and positive recurrence imply that $\hat{E}[\beta_{1(j)}|\mathcal{L}]$ converges almost surely to $E[\beta_{1(j)}|\mathcal{L}]$. Similar result also holds for $\hat{\text{Var}}[\beta_{1(j)}|\mathcal{L}]$ as it is also a sample path average.

As seen in Tables 3.4-3.9, posterior intervals for the parameters μ and σ_η^2 include the true values regardless of the prior distributions. Bayesian estimates for these parameters are not sensitive to the prior distributions.

Table 3.4: Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. *Sample Size $n=100$, Flat Priors on μ and σ_η^2 .*

ϕ	Prior Density for ϕ	Parameter	$\hat{\lambda}_{mean}$	$\hat{\lambda}_{std.err.}$	$\hat{\lambda}_{lower}$	$\hat{\lambda}_{upper}$
1	U(0,1)	ϕ	0.877	0.094	0.657	0.989
		μ	-8.997	0.607	-10.188	-7.798
		σ_η	0.356	0.112	0.195	0.626
	Mixed Prior ($p = 0.95$)	ϕ	0.898	0.087	0.678	0.997
		μ	-9.012	0.618	-10.217	-7.789
		σ_η	0.354	0.109	0.197	0.618
0.98	U(0,1)	ϕ	0.856	0.114	0.578	0.989
		μ	-8.925	0.553	-10.027	-7.855
		σ_η	0.348	0.114	0.187	0.626
	Mixed Prior ($p = 0.95$)	ϕ	0.894	0.104	0.630	0.998
		μ	-7.970	0.388	-8.812	-7.238
		σ_η	0.316	0.107	0.175	0.579
0.95	U(0,1)	ϕ	0.787	0.159	0.409	0.982
		μ	-8.965	0.453	-9.873	-8.089
		σ_η	0.339	0.120	0.175	0.637
	Mixed Prior ($p = 0.95$)	ϕ	0.791	0.195	0.219	0.994
		μ	-8.255	0.311	-8.896	-7.667
		σ_η	0.280	0.098	0.146	0.530

Table 3.5: Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. *Sample Size $n=500$, Flat Priors on μ and σ_η^2 .*

ϕ	Prior Density for ϕ	Parameter	$\hat{\lambda}_{mean}$	$\hat{\lambda}_{std.err.}$	$\hat{\lambda}_{lower}$	$\hat{\lambda}_{upper}$
1	U(0,1)	ϕ	0.992	0.005	0.979	0.999
		μ	-9.045	0.719	-10.411	-7.606
		σ_η	0.329	0.045	0.249	0.423
	Mixed Prior ($p = 0.95$)	ϕ	0.993	0.006	0.979	0.999
		μ	-9.051	0.720	-10.409	-7.596
		σ_η	0.324	0.044	0.247	0.418
	Mixed Prior ($p \sim Beta(0.5, 0.5)$)	ϕ	0.992	0.006	0.979	0.999
		μ	-9.017	0.718	-10.383	-7.563
		σ_η	0.322	0.043	0.246	0.417
	Mixed Prior ($p \sim Beta(2.66, 0.14)$)	ϕ	0.993	0.006	0.979	0.999
		μ	-9.030	0.722	-10.395	-7.571
		σ_η	0.322	0.043	0.246	0.415
0.98	U(0,1)	ϕ	0.976	0.013	0.946	0.996
		μ	-8.954	0.589	-10.108	0.236
		σ_η	0.319	0.048	-7.796	0.425
	Mixed Prior ($p = 0.95$)	ϕ	0.977	0.014	0.945	0.998
		μ	-8.957	0.596	-10.112	-7.784
		σ_η	0.312	0.049	0.234	0.427
	Mixed Prior ($p \sim Beta(0.5, 0.5)$)	ϕ	0.977	0.013	0.947	0.996
		μ	-8.941	0.592	-10.101	-7.782
		σ_η	0.317	0.048	0.235	0.420
	Mixed Prior ($p \sim Beta(2.66, 0.14)$)	ϕ	0.977	0.014	0.946	0.998
		μ	-8.928	0.606	-10.109	-7.736
		σ_η	0.318	0.049	0.234	0.423
0.95	U(0,1)	ϕ	0.945	0.025	0.888	0.984
		μ	-8.971	0.352	-9.679	-8.279
		σ_η	0.326	0.058	0.227	0.454
	Mixed Prior ($p = 0.95$)	ϕ	0.947	0.026	0.889	0.991
		μ	-8.981	0.369	-9.721	-8.259
		σ_η	0.319	0.056	0.224	0.442
	Mixed Prior ($p \sim Beta(0.5, 0.5)$)	ϕ	0.945	0.025	0.888	0.985
		μ	-8.968	0.359	-9.692	-8.262
		σ_η	0.249	0.056	0.221	0.438
	Mixed Prior ($p \sim Beta(2.66, 0.14)$)	ϕ	0.947	0.026	0.889	0.992
		μ	-8.953	0.371	-9.682	-8.203
		σ_η	0.316	0.056	0.221	0.441

Table 3.6: Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. *Sample Size $n=1000$, Flat Priors on μ and σ_η^2 .*

ϕ	Prior Density for ϕ	Parameter	$\hat{\lambda}_{mean}$	$\hat{\lambda}_{std.err.}$	$\hat{\lambda}_{lower}$	$\hat{\lambda}_{upper}$
1	U(0,1)	ϕ	0.996	0.002	0.991	0.999
		μ	-9.050	0.726	-10.419	-7.589
		σ_η	0.320	0.030	0.265	0.383
	Mixed Prior ($p = 0.95$)	ϕ	0.997	0.002	0.992	0.999
		μ	-9.047	0.727	-10.408	-7.561
		σ_η	0.317	0.031	0.262	0.382
	Mixed Prior ($p \sim Beta(0.5, 0.5)$)	ϕ	0.997	0.002	0.992	0.999
		μ	-9.050	0.726	-10.427	-7.583
		σ_η	0.318	0.030	0.263	0.381
	Mixed Prior ($p \sim Beta(2.66, 0.14)$)	ϕ	0.997	0.002	0.992	0.999
		μ	-9.048	0.725	-10.409	-7.572
		σ_η	0.381	0.030	0.263	0.382
0.98	U(0,1)	ϕ	0.979	0.009	0.961	0.993
		μ	-8.942	0.520	-9.969	-7.938
		σ_η	0.315	0.035	0.252	0.387
	Mixed Prior ($p = 0.95$)	ϕ	0.980	0.009	0.960	0.997
		μ	-8.954	0.528	-9.986	-7.917
		σ_η	0.316	0.035	0.252	0.389
	Mixed Prior ($p \sim Beta(0.5, 0.5)$)	ϕ	0.979	0.009	0.960	0.994
		μ	-8.965	0.519	-9.979	-7.952
		σ_η	0.318	0.035	0.256	0.391
	Mixed Prior ($p \sim Beta(2.66, 0.14)$)	ϕ	0.979	0.009	0.959	0.995
		μ	-8.905	0.503	-9.895	-7.921
		σ_η	0.317	0.035	0.255	0.390
0.95	U(0,1)	ϕ	0.947	0.017	0.910	0.975
		μ	-8.969	0.232	-9.429	-8.505
		σ_η	0.316	0.042	0.242	0.406
	Mixed Prior ($p = 0.95$)	ϕ	0.950	0.018	0.912	0.982
		μ	-8.978	0.249	-9.472	-8.491
		σ_η	0.316	0.041	0.244	0.404
	Mixed Prior ($p \sim Beta(0.5, 0.5)$)	ϕ	0.949	0.016	0.914	0.977
		μ	-8.960	0.243	-9.429	-8.473
		σ_η	0.319	0.040	0.248	0.405
	Mixed Prior ($p \sim Beta(2.66, 0.14)$)	ϕ	0.949	0.018	0.911	0.984
		μ	-8.984	0.257	-9.486	-8.490
		σ_η	0.316	0.041	0.242	0.404

Table 3.7: Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. *Sample Size $n=100$, Informative Priors on μ and σ_η^2 .*

ϕ	Prior Density for ϕ	Parameter	$\hat{\lambda}_{mean}$	$\hat{\lambda}_{std.err.}$	$\hat{\lambda}_{lower}$	$\hat{\lambda}_{upper}$
1	U(0,1)	ϕ	0.909	0.077	0.718	0.995
		μ	-8.667	0.517	-9.662	-7.633
		σ_η	0.221	0.055	0.139	0.352
	Mixed Prior ($p = 0.95$)	ϕ	0.921	0.072	0.742	0.998
		μ	-8.679	0.523	-9.672	-7.618
		σ_η	0.222	0.054	0.142	0.349
0.98	U(0,1)	ϕ	0.871	0.107	0.618	0.992
		μ	-8.652	0.462	-9.542	-7.704
		σ_η	0.213	0.055	0.133	0.346
	Mixed Prior ($p = 0.95$)	ϕ	0.904	0.091	0.665	0.998
		μ	-8.646	0.476	-9.561	-7.682
		σ_η	0.208	0.053	0.130	0.334
0.95	U(0,1)	ϕ	0.813	0.146	0.482	0.985
		μ	-8.761	0.386	-9.497	-7.969
		σ_η	0.193	0.051	0.119	0.319
	Mixed Prior ($p = 0.95$)	ϕ	0.848	0.138	0.505	0.997
		μ	-8.752	0.391	-9.501	-7.959
		σ_η	0.191	0.049	0.119	0.312

Table 3.8: Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. *Sample Size $n=500$, Informative Priors on μ and σ_η^2 .*

ϕ	Prior Density for ϕ	Parameter	$\hat{\lambda}_{mean}$	$\hat{\lambda}_{std.err.}$	$\hat{\lambda}_{lower}$	$\hat{\lambda}_{upper}$
1	U(0,1)	ϕ	0.994	0.004	0.983	0.999
		μ	-8.569	0.699	-9.893	-7.144
		σ_η	0.293	0.039	0.224	0.377
	Mixed Prior ($p = 0.95$)	ϕ	0.994	0.004	0.983	0.999
		μ	-8.587	0.694	-9.896	-7.175
		σ_η	0.290	0.034	0.223	0.375
0.98	U(0,1)	ϕ	0.981	0.011	0.956	0.997
		μ	-8.639	0.570	-9.742	-7.498
		σ_η	0.283	0.043	0.209	0.376
	Mixed Prior ($p = 0.95$)	ϕ	0.982	0.002	0.957	0.999
		μ	-8.656	0.575	-9.758	-7.502
		σ_η	0.280	0.041	0.208	0.368
0.95	U(0,1)	ϕ	0.960	0.019	0.915	0.991
		μ	-8.814	0.375	-9.529	-8.029
		σ_η	0.265	0.047	0.185	0.367
	Mixed Prior ($p = 0.95$)	ϕ	0.964	0.019	0.920	0.997
		μ	-8.796	0.393	-9.535	-7.974
		σ_η	0.252	0.044	0.179	0.349

Table 3.9: Average Value of Posterior Means, Posterior Sampling Variances, Lower and Upper Bounds of Posterior Intervals of the SVM Parameters Over 100 Repetitions. *Sample Size $n=1000$, Informative Priors on μ and σ_η^2 .*

ϕ	Prior Density for ϕ	Parameter	$\hat{\lambda}_{mean}$	$\hat{\lambda}_{std.err.}$	$\hat{\lambda}_{lower}$	$\hat{\lambda}_{upper}$
1	U(0,1)	ϕ	0.997	0.002	0.992	0.999
		μ	-8.585	0.708	-9.917	-7.153
		σ_η	0.300	0.028	0.248	0.358
	Mixed Prior ($p = 0.95$)	ϕ	0.998	0.002	0.993	0.999
		μ	-8.583	0.723	-9.943	-7.106
		σ_η	0.300	0.029	0.248	0.360
0.98	U(0,1)	ϕ	0.981	0.008	0.965	0.995
		μ	-8.684	0.509	-9.669	-7.661
		σ_η	0.294	0.032	0.237	0.361
	Mixed Prior ($p = 0.95$)	ϕ	0.983	0.008	0.965	0.997
		μ	-8.682	0.519	-9.681	-7.636
		σ_η	0.293	0.032	0.237	0.362
0.95	U(0,1)	ϕ	0.956	0.014	0.925	0.981
		μ	-8.893	0.253	-9.375	-8.371
		σ_η	0.281	0.036	0.216	0.358
	Mixed Prior ($p = 0.95$)	ϕ	0.958	0.016	0.926	0.988
		μ	-8.897	0.265	-9.396	-8.355
		σ_η	0.282	0.037	0.217	0.359

3.7.2 Sensitivity of the posterior inference about ϕ to the prior distribution of ϕ

As we said earlier, $U(0,1)$ density has commonly been used as a prior density for the AR parameter. The densities we introduced are also reasonable prior densities for the AR parameter and serve very well to our purpose of using posterior intervals to perform a unit root test. We compare the posterior inference regarding ϕ obtained through a mixed density versus the one obtained through a $U(0,1)$. We make a similar comparison between the mixed density and a continuous distribution that assigns a positive density on the values of ϕ that are greater than one. This sensitivity analysis is carried out via Wilcoxon's two sample rank sum test. For Table 3.10, sample 1 is the posterior mean estimates of ϕ obtained from 100 Monte Carlo samples where prior density of ϕ is $U(0,1)$. Similarly, sample 2 consists of the posterior mean estimates of ϕ collected through 100 Monte Carlo samples where prior density for ϕ is mixed prior with a fixed mixing probability. For Table 3.11, sample 1 is the posterior mean estimates of ϕ obtained through the use of $U(0,1.95)$. Each sample has size 100. The reason that we perform the nonparametric test on posterior inference concerning the mean of the posterior distribution of ϕ is that posterior inference on medians of posterior distributions are generally less sensitive to the change in the model compared to the posterior inferences regarding the means. In Tables 3.10 and 3.11 we give the p-values of the test.

The posterior inference about ϕ obtained through $U(0,1)$ prior and mixed prior look the same for most of the cases shown in Table 3.10. However, when the data are generated with a unit root in the volatility, the nonparametric test picked a difference between the posterior inference obtained from the mixed prior and the one obtained from $U(0,1)$ for sample size $n = 1000$. On the other hand, although the test found a significance difference between the posterior inference obtained from the mixed prior and $U(0,1.95)$ for $n = 500$ in the presence of a unit root, it vanishes as n increases.

Table 3.10: Sensitivity of Priors Using Wilcoxon's Rank Sum Test. ($H_0 : E[\phi|\underline{x}, \phi \sim U(0, 1)] = E[\phi|\underline{x}, \phi \sim \text{Mixed Uniform with } p = 0.95]$)

Prior Density for μ and σ_η^2	ϕ	n	p-value
$\mu \sim N(0, 10^6), \sigma_\eta^2 \sim IG(2 + 10^{-10}, 0.1)$	1	100	0.467
		500	0.192
		1000	0.028
	0.98	100	0.439
		500	0.832
		1000	0.849
	0.95	500	0.274
		1000	0.509
	$\mu \sim N(0, 10), \sigma_\eta^2 \sim IG(5, 0.1)$	1	100
500			0.459
1000			0.051
0.98		100	0.293
		500	0.458
		1000	0.589
0.95		100	0.350
		500	0.703
		1000	0.645

Table 3.11: Sensitivity of Priors Using Wilcoxon's Rank Sum Test. ($H_0 : E[\phi|\underline{x}, \phi \sim U(0, 1.95)] = E[\phi|\underline{x}, \phi \sim \text{Mixed Uniform with } p = 0.95]$)

Prior Density for μ and σ_η^2	ϕ	n	p-value
$\mu \sim N(0, 10^6), \sigma_\eta^2 \sim IG(2 + 10^{-10}, 0.1)$	1	500	0.035
		1000	0.095
	0.98	500	0.853
		1000	0.757
	0.95	500	0.372
		1000	0.070

These behaviors can be explained by the following theorem.

Theorem: Let β be the parameter of interest which has a continuous distribution and it is defined on a compact set. Let β_0 be the value of β that minimizes the distance between the model distribution $f(\underline{x}|\beta)$ and the true distribution $f(\underline{x})$. Let A

be a neighborhood of β_0 with nonzero prior probability. Then, $P(\beta \in A|\underline{r}) \rightarrow 1$ as $n \rightarrow \infty$.

Proof: See Gelman et al (1995), Appendix B for the proof.

The U(0,1) prior density doesn't put a positive prior probability on $\phi = 1$ and so the result of the theorem stated above may not be working for this case. Since mixed prior we introduced assigns a positive prior probability to $\phi = 1$ the result of the theorem applies to this case. This implies that a mixed prior possibly picked up the correct inference more often than the U(0,1) prior. That is why we see a difference between the posterior inferences when sample size is large, say $n=1000$.

3.7.3 Proportion of the posterior intervals that contain the true value of ϕ when the true value of ϕ is 0.98 or 0.95

Table 3.12: The number of the posterior intervals that include the true value of ϕ . 100 Monte Carlo replications.

ϕ	Prior Density	$n = 500$	$n = 1000$
0.98	U(0,1)	91	92
	U(0,1.95)	94	94
	U(0,20)	95	94
	U(-2,2)	94	94
	Mixed with U(0,1), $p = 0.95$	96	96
0.95	U(0,1)	94	95
	U(0,1.95)	95	97
	U(0,20)	96	97
	U(-2,2)	95	97
	Mixed with U(0,1), $p = 0.95$	99	98

("MU" stands for the prior for ϕ which is a Mixed with U(0,1) where $p = 0.95$. Also, priors for nuisance parameters are $\mu \sim N(0, 10^6)$ and $\sigma_\eta^2 \sim IG(2 + 10^{-10}, 0.1)$)

In Table 3.12, we display the number of posterior intervals of ϕ that contain the true value of ϕ . We see that, when we use a mixed prior density with $p = 0.95$, U(-2,2), U(0,1.95), and U(0,20) the proportion of the posterior intervals that cover

the true value of ϕ is higher than the case where we use a $U(0,1)$ density. Also, they maintain roughly, a 95% coverage probability as we use a 95% posterior interval.

3.8 Discussion

The algorithms that are suggested to compute the Bayes Factor are not very convenient as they have computational problems for large data sets. Actually, one can estimate the Bayes Factor from the Markov Chain much more easily if a mixed prior density is used for ϕ . Notice that, Bayes Factor is the ratio of the posterior odds and the prior odds. Recall that, the posterior odds for a unit root test is $\frac{Pr(H_0|x)}{Pr(H_1|x)} = \frac{Pr(\phi=1|x)}{Pr(0<\phi<1|x)}$. As we show in Appendix A, when there is a positive mass on the point 1 apriori, posterior density of ϕ assigns a positive prior on the point 1 as well. Therefore, the numerator and the denominator of the posterior odds can be easily estimated from the Markov Chain as long as the chain is irreducible and positive recurrent when we use a mixed prior on ϕ . This estimator is simply $\frac{\sum_{i=m+1}^{m+M} I(\phi^{(i)}=1)}{\sum_{i=m+1}^{m+M} I(0<\phi^{(i)}<1)}$ where $I(\cdot)$ is the indicator function and it is 1 if the statement in the argument is satisfied, 0 otherwise. Recall that, here, m is the burn-in point, M is the number of iterations used for inference. In addition, $\phi^{(i)}$ represents the i th element of the Markov Chain for ϕ that is used for the inferential purposes after the burn-in point. The Bayesian unit root test based on the posterior odds rejects the null hypothesis if this estimator is less than 1.

One should find out how long the chain should be in order to have this estimator converge to the posterior odds. First of all, note that, if the median of the posterior distribution of ϕ is 1 then the posterior odds is greater than or equal to 1. To get an idea about the size of the chain to estimate the posterior odds well, we make use of this relationship between the median and the posterior odds, and so apply Raftery-Lewis (1992) test for the median of the posterior distribution. This test, that is applicable in CODA, determines the number of iterations in the Markov Chain after the burn-in

point such that the quantile of interest be estimated within $\pm r$ with probability s . Here, r is the degree of the precision. In CODA, we ran this test with $r = 0.005$ and $s = 0.95$ to get the number of iterations determined where our quantile of interest is 0.5. The size of the chain suggested by this test turned out to be at least 38415 iterations.

When one uses a mixed prior for ϕ as we introduced in this thesis, Bayes Factor can easily be estimated by using this method which we discussed in this section. One should remember to construct a long Markov Chain in order to estimate the posterior odds well. Constructing a long Markov Chain is not a difficult task with today's computer technology.

Chapter 4

Application

In this chapter, we apply the frequentist and Bayesian unit root testing procedures we discuss in previous chapters to the data sets that were analyzed by Harvey et al. The data consist of the following four exchange rates: Pound/Dollar, Deutschmark(DM)/Dollar, Yen/Dollar, and Swiss-Frank(SF)/Dollar. The data are close exchange rates from 1 October 1981 to 28 June 1985 on the weekdays. The sample size is $n = 946$. Let us denote the exchange rate for series i which is on day t by p_{it} . Then, for each series i , the mean corrected return data calculated as

$$r_{it} = \Delta \log p_{it} - \frac{1}{n} \sum \Delta \log p_{it} \quad t = 1, \dots, n$$

where Δ is the differencing operator. Here, i is 1, 2, 3, or 4 corresponding to Pound/Dollar, Deutschmark/Dolar, Yen/Dollar, or Swiss-Frank/Dollar respectively.

The time series plots of the returns are displayed in Figure 4. It is hard to tell from the figures whether the affect of the random shocks to volatilities are persistent or not.

For each of the four series, we fitted an ARMA(1,1) model to the data in the form of $\log r_t^2$ and checked the autocorrelations of the residuals from these fits by using PROC ARIMA in SAS. The test result is that residuals are white noise implying that ARMA(1,1) is an adequate fit to $\log r_t^2$ for each series which in turn implies that a

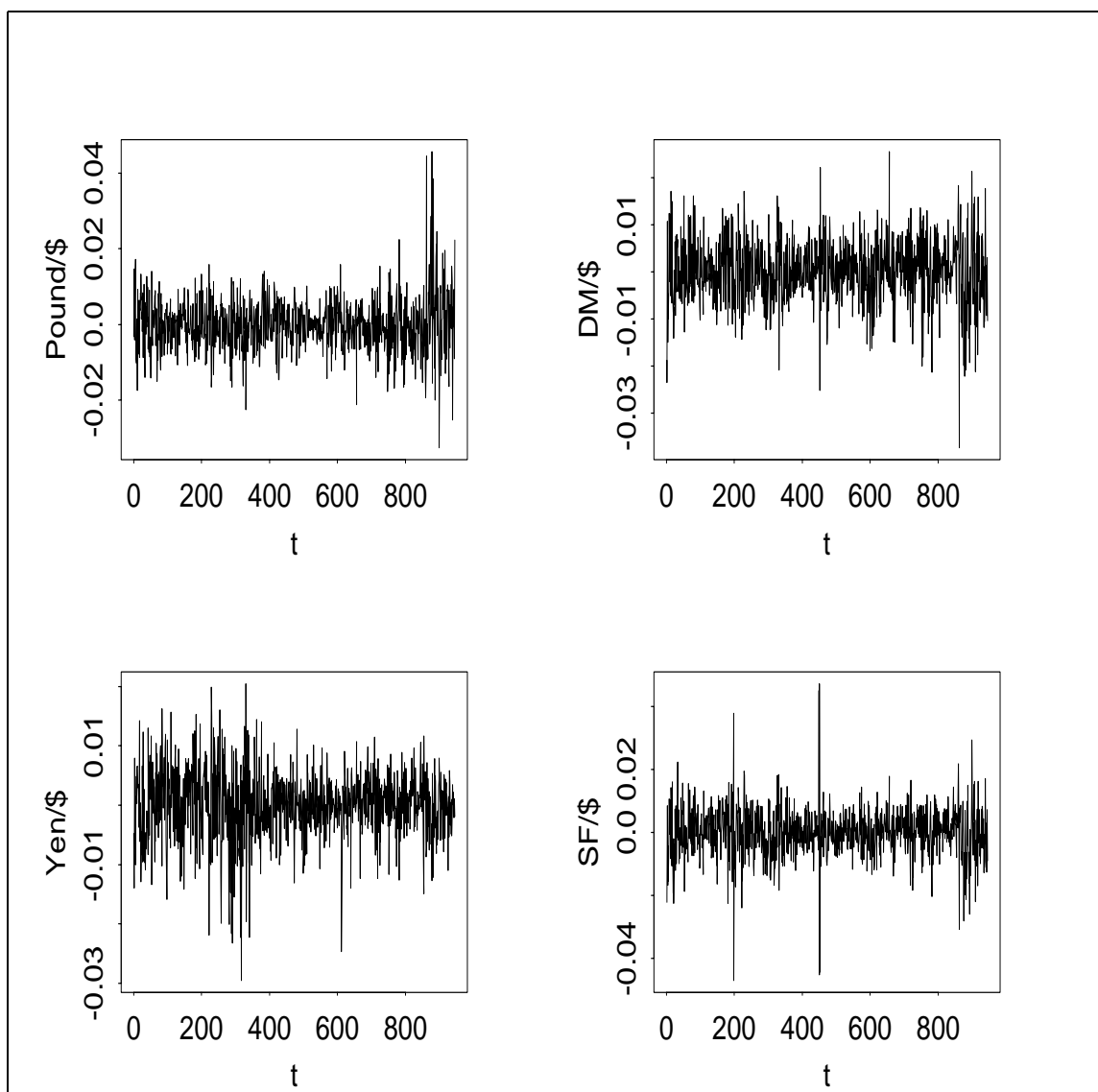


Figure 4.1: Plot of the mean corrected returns r_{it} versus time

SVM is a reasonable fit to the mean corrected exchange rate returns.

The results of the Ljung-Box test are given in Table 4.1.

Table 4.1: p-values for the Ljung-Box Test Statistic

Series	Lag	p-value
Pound/\$	6	0.81
	12	0.93
	18	0.94
	24	0.96
DM/\$	6	0.11
	12	0.31
	18	0.62
	24	0.37
Yen/\$	6	0.32
	12	0.54
	18	0.56
	24	0.77
SF/\$	6	0.12
	12	0.27
	18	0.36
	24	0.13

Harvey et al. computed the Maximum Likelihood Estimates of ϕ for each series treating the random errors in the corresponding ARMA(1,1) models as though they were $N(0, \pi^2/2)$. The maximum likelihood estimates were all close to one. They also calculated the log-likelihoods under the null and the alternative models for each series. The results implied that random walk model to the log-volatilities fit well for all four series.

As we mentioned in Section 1.3., they approximated the ARMA(1,1) model for each series by AR(9) with nonzero mean and calculated the ADF test statistics based on OLS estimators. They showed that the test rejected the hypothesis of a unit root at the 1% level for all the series as the values of the test statistics were smaller than the Dickey Fuller critical point.

4.1 Application of Frequentist Unit Root Tests

In order to investigate the unit root in volatilities of these exchange rates by using the frequentist unit root testing techniques, we transformed the squared mean corrected returns to the logarithmic scale. That is, each data set i analyzed in this section is in the form of $\log r_{it}^2$.

We ran PROC ARIMA in SAS on each of these four data sets and obtained the Unconditional Maximum Likelihood (UML) Estimates of AR parameter ϕ , MA parameter θ , and variance of the random shocks σ^2 in the corresponding ARMA(1,1) model. This estimate of ϕ and σ^2 along with the estimates obtained from the unit root testing procedures we discuss in Chapter 2 are tabulated in Table 4.2.

For example, by looking at the UML estimates of ϕ and the corresponding standard errors of the estimates, we expect to find a unit root in log-volatility process for possibly all the exchange rate series we analyze here.

Table 4.2: Different estimates of ϕ and σ^2 of the ARMA(1,1) model that corresponds to the SVM.

parameter	Estimation procedure	Pound/\$	DM/\$	Yen/\$	SF/\$
ϕ	one step GN	0.9630 (0.0124)	0.9493 (0.0128)	0.9497 (0.0142)	0.9385 (0.0143)
	iterative GN	0.9859 (0.0064)	0.9573 (0.0104)	0.9949 (0.0033)	0.9575 (0.0101)
	UMLE	0.9902 (0.0082)	0.9637 (0.0203)	0.9954 (0.0042)	0.9627 (0.0211)
σ^2	ADF/OLS, p_{min}	4.9614	5.0542	5.6041	5.6442
	ADF/OLS, p_{full}	5.2285	5.2682	5.8834	5.9807
	ADF/WS, p_{full}	4.8907	4.9681	5.5181	5.6316
	one step GN	4.8486	4.9704	5.5354	5.5961
	iterative GN	4.7679	4.9593	5.4124	5.5808
	UMLE	4.7950	5.0030	5.4380	5.6270

(standard error estimates of the AR parameter estimates are given in the parentheses.)

By exploiting the fact that $\log r_{it}^2$ has same moments as an ARMA(1,1) process with the model parameters given in (2.3) and (2.4), we apply the ADF unit root test

with OLS and WS regressions as well as Said-Dickey unit root test to the data that is in the form of $\log r_{it}^2$ for each series. Before performing the unit root tests, the sample mean of $\{\log r_{it}^2\}$ is subtracted from each such observation of each series. We take the level of the test to be 0.05. For all the exchange rate series we consider here, we computed Durbin and one step Gauss Newton estimates of θ in order to select the cutoff point from Tables 2.3, 2.5, 2.7, and 2.9. We chose not to use the Durbin adjusted estimate as it didn't perform well in general in terms of maintaining the nominal level as can be seen in the empirical power tables in Section 2.7. The one step Gauss Newton estimate of θ uses a Durbin estimate of θ as a starting point. In this section, we use two different lags to truncate the infinite sum in calculation of the Durbin estimate of θ . The first one is 10 which is chosen arbitrarily. We denote this estimate by $\hat{\theta}_{durbin,10}$ and the corresponding one step Gauss Newton estimate by $\hat{\theta}_{one\ step\ GN,10}$. The second lag we consider is $[10n^{1/4}]$ which is equal to 55 for $n = 945$. We denote this one by $\hat{\theta}_{durbin,full}$ and the following one step Gauss Newton estimate by $\hat{\theta}_{one\ step\ GN,full}$. Note that, one observation is lost when we take the difference of the logarithm of exchange rates to obtain the returns. Besides the one step Gauss-Newton estimates, we also computed iterative Gauss Newton estimate for all the series we have here. In each iteration, we take the value of ϕ to be 1 and we continue iterating until the convergence criteria are satisfied. We use the same convergence criteria as in Said and Dickey (1985); stop iterating when the change in θ and the change in v_0 are less than 0.00002. Recall that v_0 is the random shock at time $t = 0$ in an ARMA(1,1) process and Said and Dickey treat it as a parameter in performing Gauss Newton estimation procedure. We denote the estimate of θ obtained that way by $\hat{\theta}_{itr. GN,10}$ when the Durbin estimate that is found by using a 10 lag regression is used as an initial point to the iterative method. Similarly, we denote the iterative Gauss Newton estimate of θ by $\hat{\theta}_{itr. GN,full}$ when we use a Durbin estimate that is computed by using the maximum lag of regression as an initial point. In addition to the estimates of θ that we considered in our Monte Carlo Power Study, we consider

one more type of estimating procedure for θ in this chapter. Due to the ease of computing the MLE of θ in SAS, we used the unconditional MLE of θ to select a critical point. We denote the unconditional MLE of θ by $\hat{\theta}_{umle}$.

Below, we analyze each series separately. For the sake of the clarity, we list the estimates of θ in a vector in the following order;

$$\hat{\theta} = (\hat{\theta}_{durbin,10}, \hat{\theta}_{durbin,full}, \hat{\theta}_{one\ step\ GN,10}, \hat{\theta}_{one\ step\ GN,full}, \hat{\theta}_{itr.\ GN,10}, \hat{\theta}_{itr.\ GN,full}, \hat{\theta}_{umle}).$$

The empirical critical points that are selected by using these estimates are also given in an 1×7 row vector. The j th element of this vector is the critical point selected by using the j th element of the estimate vector $\hat{\theta}$. These critical points are approximated using the nonlinear models that are fit to smooth the percentiles over θ for a fixed sample size. This is Method 2 of Section 2.5. Smoothing equations are presented in Appendix C. We perform the tests by comparing the calculated test statistic values against these critical percentiles individually.

Pound/Dollar Series: For this series, $\hat{\theta}$ is (0.8886, 0.9417, 0.9532, 0.9557, 0.9542, 0.9542, 0.9579).

The minimum order ADF unit root test is applied and the order is chosen as 5 by the procedure. The value of the test statistic, $t_{\mu,(ADF/OLS),p_{min}}$, calculated for this data set is -11.61. The critical points approximated for this test that corresponds to $n = 1000$ and different estimates of θ are (-4.74, -8.80, -10.43, -10.85, -10.60, -10.60, -11.22) respectively. The ADF/OLS unit root test with minimum order rejects the unit root null.

We also applied ADF unit root test with 55 lags. The test statistic $t_{\mu,(ADF/OLS),p_{full}}$ is calculated to be -2.47. The approximated critical points that correspond to the estimates of θ given in the order above are (-3.06, -3.99, -4.44, -4.55, -4.48, -4.48, -4.66). Consequently, this test does not reject the unit root null for all the estimates of θ we examined.

The test that uses the statistic $t_{\mu,(ADF/WS),p_{full}}$ does not reject the unit root null for all type of estimates of θ we examine here as the computed test statistic for the

data is -2.38 and the critical points accompanying the estimates of θ we discuss here are (-2.56, -2.70, -2.83, -2.87, -2.84, -2.84, -2.91).

Latter two tests do not reject the unit root null when the comparison of the calculated values of test statistic is made against Dickey-Fuller critical points which are -2.86 for ADF with OLS regression and -2.52 for ADF with WS regression.

The Said-Dickey test statistic, $t_{\mu, SDD}$, for this data set is -2.99 where the critical points smoothed for the estimates of θ given above are (-3.11, -4.74, -5.83, -6.14, -5.94, -5.94, -6.43). As a result of this test, we don't have strong evidence against unit root in the volatility of Pound/Dollar. That is, we reach the conclusion that shocks to volatility in Pound/Dollar rates remain for a long time. However, the Said-Dickey test that uses Dickey-Fuller critical point finds an evidence for stationarity. The latter inference is doubtful though because e.g. unconditional mle of the AR parameter of ϕ is 0.9902 which suggests a nonstationarity in variance for the return defined for Pound/Dollar series.

DM/Dollar Series: For this series, $\hat{\theta}$ is (0.8812, 0.8953, 0.9139, 0.9129, 0.9080, 0.9080, 0.9105).

The order chosen by AIC for this series is 5 and so $p_{min} = 7$. The calculated test statistic, $t_{\mu, (ADF/OLS), p_{min}}$, is -9.19. The critical points smoothed for those estimates of θ are (-4.46, -5.04, -6.13, -6.06, -3.26, -3.26, -5.90). Accordingly, we have statistically significant evidence in favor of stationarity in variance of DM/Dollar series based on this testing procedure for all the estimates of θ we inspect here.

In addition, the value of $t_{\mu, (ADF/OLS), p_{full}}$ is -4.75. The critical points smoothed with respect to the estimates of θ then are (-3.01, -3.11, -3.34, -3.32, -3.26, -3.26, -3.29). Consequently, this test finds statistically sufficient evidence for stationarity in the volatility of DM/Dollar series.

The value of $t_{\mu, (ADF/WS), p_{full}}$ for this series is -3.73. The critical points with respect to these various estimates of θ are (-2.56, -2.56, -2.58, -2.58, -2.57, -2.57, -2.58) which results in rejection of the unit root.

The value of the Said-Dickey test statistic is -3.97. Critical points approximated with respect to the estimates of θ are (-3.05, -3.17, -3.49, -3.46, -3.36, -3.36, -3.41). Using Said-Dickey procedure to perform a unit root test results in rejection of the unit root model.

All of these four test statistics applied to DM/Dollar series also reject the null model when Dickey-Fuller critical values are used.

Yen/Dollar series: For this series, the estimates of $\hat{\theta}$ are (0.8549, 0.9594, 0.9304, 0.9702, 0.9731, 0.9731, 0.9721)

The order of the ADF fit to ARMA(1,1) is $p_{min} = 8$ as the model that is selected by AIC has 6 lags. The value of $t_{\mu,(ADF/OLS),p_{min}}$ is -8.26 whereas the approximated critical points corresponding to different estimates of θ are (-3.75, -11.48, -7.53, -13.67, -14.34, -14.34, -14.10). Accordingly, a random walk model for volatility is rejected when $\hat{\theta}_{durbin,10}$ or $\hat{\theta}_{one\ step\ GN,10}$ is employed to approximate the critical point. Nevertheless, the test based on the critical points that corresponds to the other estimates that are employed fails to reject the unit root null and this conclusion is actually in agreement with our expectations regarding the volatility of this series. On the other hand, using Dickey-Fuller critical points along with this test statistic rejects the unit root null in spite of the fact that some estimates of the AR parameter strongly suggest nonstationarity as seen in Table 4.2. It is once again demonstrated that comparing the test statistic against Dickey-Fuller critical points might produce unreliable results.

The value of the test statistic $t_{\mu,(ADF/OLS),p_{full}}$ is -2.11. The approximate critical points are (-2.89, -4.74, -3.58, -5.40, -5.61, -5.61, -5.54) so this test does not reject the unit root in the volatility of this series. Same conclusion is also obtained by comparing the test statistic against Dickey-Fuller critical points.

$t_{\mu,(ADF/WS),p_{full}}$ for this series is calculated as -2.01. Correspondingly, the smoothed critical points are (-2.55, -2.94, -2.63, -3.26, -3.38, -3.38, -3.33). The test results in not rejecting the unit root null no matter what estimate of θ is used. Using Dickey-Fuller critical points, the test fails to reject unit root null model as well.

The Said-Dickey test applied to the Yen/Dollar series produces $t_{\mu,SDD}$ equal to -3.55. The approximate critical points are (-2.95, -6.65, -4.06, -8.66, -9.35, -9.35, -9.10). The unit root model is rejected when $\hat{\theta}_{durbin,10}$ is used to pick out the corresponding critical point. However, except for this one, the Said-Dickey unit root test does not reject the unit root null implying that variation in the return related with Yen/Dollar exchange rate is not constant over time. The test that compares the calculated test statistic against the Dickey-Fuller critical value rejects the random walk model for the volatility of this series.

SF/Dollar series: For this series, $\hat{\theta}$ is (0.8624, 0.9018, 0.9024, 0.9095, 0.9093, 0.9093, 0.9130)

AIC selects 6 as the order of the ADF regression for this series and hence $p_{min} = 8$. The value of $t_{\mu,(ADF/OLS),p_{min}}$ is computed as -8.02. Critical values for this test statistic approximated by the corresponding nonlinear model are (-3.92, -5.37, -5.40, -5.83, -5.82, -5.82, -6.07). Hence, the test concludes in rejection for all the estimates of θ we try.

Also, the value of $t_{\mu,(ADF/OLS),p_{full}}$ for this series comes out to be -4.76. Critical points estimated are as follows; (-2.92, -3.18, -3.19, -3.28, -3.27, -3.27, -3.33). This test gives a strong evidence against the null hypothesis for this series.

The test statistic $t_{\mu,(ADF/WS),p_{full}}$ for this series is -3.83. Critical points approximated for the given estimates of θ are (-2.55, -2.57, -2.57, -2.57, -2.57, -2.57, -2.58). The result of these tests imply that shocks to volatility of SF/Dollar series are not persistent.

$t_{\mu,SDD}$ is -4.29 whereas the approximated critical points for various estimates of θ are (-2.97, -3.26, -3.26, -3.39, -3.39, -3.39, -3.47). This test finds an evidence against unit root in volatility of this series.

Note that, for SF/Dollar series, comparing all the four test statistics to Dickey-Fuller critical points rejects the unit root as well.

Our recommendation for the practitioners working with SVM would be to be

cautious in using the well known unit root tests. If one tends to approximate the corresponding ARMA(1,1) model by an ADF regression to perform a unit root test, we suggest them fitting a more parsimonious AR model than the one whose order is $(p_{AIC}+2)$. Then, using the smoothed empirical critical points provides reliable results regarding a unit root in volatility whereas using Dickey-Fuller critical points is likely to be unreliable. Tests based on full order ADF regression along with OLS or WS estimation can be applied confidently. Using either smoothed percentiles or Dickey-Fuller critical points as cutoff points for these tests both provide reliable results. The test that applies the Said-Dickey procedure should compare the calculated test statistic against the smoothed cutoff points in order to attain a valid conclusion regarding the persistence of volatility. The most pleasing test among the ones that are suggested above is the test that uses the Said-Dickey testing procedure and compares it with the smoothed critical points. The reason is that it is more powerful than the others as is seen in our Power Study. This is not surprising as the test is fitting the correct model rather than an approximation and the use of smoothed critical values helps to compensate for the discrepancy between limit and finite sample critical values. Based on our experience, we advise practitioners to try different estimates of θ and approximate the critical point for each of them so that they can get a better view of a possible unit root in the volatility of the series.

4.2 Application of Bayesian Posterior Interval Unit Root Test

In this section, we make use of the Posterior Interval unit root test to seek for evidence of a unit root in the volatility of the exchange rate series considered here. Notice that, unlike the previous section, Bayesian unit root tests are conducted on the data $\{r_t\}$. In the applications, we chose to use flat prior distributions for μ and σ_η^2 . The same prior distributions used in the simulation experiment are chosen for

ϕ . That is, we use a mixed prior for ϕ with a constant mixing probability p as well as with a mixing probability p having a Beta distribution as a prior distribution. We construct three Markov chains starting from three different set of initial values for the parameters. We execute the Gibbs Sampling iterations 10,000 times for each chain corresponding to different set of starting values. For each chain, the first 5,000 iterations are discarded and every 5th element of the last 5,000 iterations are used to obtain the posterior distribution of the parameters. Below, the posterior mean and the 95% posterior interval of the distribution of the parameters are given.

Table 4.3: Post. Inference on ϕ . $\phi \sim \text{Mixed } U(0, 1)$ with $p = 0.95$

Exchange rate	Estimated Posterior mean	Standard error of the estimate	Estimated 95%Posterior Interval
Pound/\$	0.9920	0.0013	[0.9528,1.0]
DM/\$	0.9881	0.0018	[0.9363,1.0]
Yen/\$	0.9898	0.0016	[0.9437,1.0]
SF/\$	0.9791	0.0033	[0.8925,1.0]

Table 4.4: Post. Inference on ϕ and p . $\phi \sim \text{Mixed } U(0, 1)$ with $p \sim \text{Beta}(0.5, 0.5)$

Exchange rate	Prm	Estimated Posterior mean	Standard error of the estimate	Estimated 95%Posterior Interval
Pound/\$	ϕ	0.9940	9.810^{-4}	[0.9658,1.0]
	p	0.8757	0.0126	[0.3596,1.0]
DM/\$	ϕ	0.9893	0.0019	[0.9330,1.0]
	p	0.8717	0.0134	[0.3507,1.0]
Yen/\$	ϕ	0.9968	5.610^{-4}	[0.9772,1.0]
	p	0.8773	0.0130	[0.3581,1.0]
SF/\$	ϕ	0.9777	0.0034	[0.8981,1.0]
	p	0.8759	0.0130	[0.3644,1.0]

When we use a mixed prior with $p = 0.95$, in addition to the posterior inference given in the tables, median of the posterior distribution of the parameter ϕ turns out to be one for each series. It appears that, for each series, there is more than 50%

Table 4.5: Post. Inference on ϕ . $\phi \sim U(0, 1.95)$

Exchange rate	Estimated Posterior mean	Standard error of the estimate	Estimated 95%Posterior Interval
Pound/\$	0.9769	7.010^{-4}	[0.9462,0.9971]
DM/\$	0.9642	0.0013	[0.9167,0.9978]
Yen/\$	0.9868	7.410^{-4}	[0.9557,1.0]
SF/\$	0.9405	0.0017	[0.8739,0.9869]

probability that ϕ is 1 given the data. The posterior intervals cover the point one implying that null hypothesis can not be rejected for any of the series.

When we use $\phi \sim U(0, 1.95)$, the test based on the posterior interval does find sufficient evidence in favor of the unit root for Yen/Dollar series. The test rejects the unit root null hypothesis for all the other series.

Chapter 5

Conclusions

In this thesis, we have developed frequentist and Bayesian unit root tests for Stochastic Volatility (SV) Models where the log-squared volatilities are assumed to have a first order autoregressive model.

We have adjusted the classical unit root tests based on the Augmented Dickey Fuller (ADF) test statistic with ordinary least squares and weighted symmetric regression, Said-Dickey test statistic, and the test statistics which are based on the instrumental variable estimators in order to test for a unit root in SV Models. In order to test for a unit root in SV Models, we use the cutoff points generated from an ARMA(1,1) model which has the same autocovariance structure as our SVM. In finite samples, unit root tests require using critical points which are selected with respect to an estimate of the moving average parameter as well as the sample size. We have shown the importance of estimating the MA parameter in determining approximate critical values for unit root test statistics in finite samples.

For the SV Model, we have derived the asymptotic distributions of various test statistics based on the instrumental variable estimators under the unit root model. We have shown that the pivotal test statistic based on the instrumental variable estimators in a SV Model multiplied by a correction factor has the same limiting distribution as the Dickey-Fuller test statistic.

Through a Monte Carlo power study, we have established the performance of the adjusted frequentist unit root tests for finite samples generated from a SV Model. We have found out that the adjusted unit root tests based on ADF and Said-Dickey test statistics maintain the nominal level for moderately large samples. In these cases, the powers of the tests are high enough, indicating that one can obtain reliable conclusions from applying these tests in SV Models. Another relevant finding is that the powers of the adjusted tests based on the ADF test statistic with weighted symmetric regression and Said-Dickey test statistic are a bit higher than the test based on the ADF test statistic with ordinary least squares regression. In this finite sample study, we see that the adjusted unit root test based on the instrumental variable regression is not as powerful as the other test statistics discussed here in SV Models.

In addition to the frequentist unit root tests, we considered Bayesian approach to perform a unit root test in our SVM. We have introduced a mixed prior density that puts a nonzero mass on the null value of the autoregressive parameter. We also suggested to use densities that include the nonstationary values of ϕ as prior densities. By using such densities on ϕ , one can use the posterior interval of ϕ in order to test for a unit root in a SVM. If one uses a continuous density defined on a space where the null value of the parameter is not included, one does not have the advantage of using a posterior interval unit root test as the posterior interval never includes the value being tested. In that case, one uses a Bayesian unit root test which is based on the Bayes Factor. However, computing the Bayes Factor requires the use of computationally intensive algorithms and therefore it is not practical. On the other hand, calculating the posterior interval is very simple and can be performed easily in WinBUGS.

Using a Monte Carlo study, we have shown that the posterior interval unit root test in SV Models is reliable and simple to compute. We found a statistically significant difference between the posterior mean of ϕ obtained through using a mixed prior density and the posterior mean of ϕ obtained through using a $U(0,1)$ prior density.

We are also interested in comparing the power of the frequentist and Bayesian unit root tests considered in this study for SV Models. As Bayesian unit root tests don't have a fixed significance level, one way to compare the Bayesian tests and frequentist tests with each other is to look at the total error rates. Below, we compare the sum of Type 1 and Type 2 errors for different frequentist and Bayesian unit root tests considered for SVM in this thesis. In order to compute the total error rates for the following frequentist tests, we use the empirical significance level and power results from Tables 2.26 and 2.27 that correspond to the column of $\hat{\theta}_{one\ step\ GN}$. For the Bayesian tests, we consider the case where flat prior densities are used for the nuisance parameters μ and σ_{η}^2 .

One can derive several conclusions from Tables 5.1-5.3. For small samples like $n = 100$ in SVM, as we see from Table 5.1, total error rates committed by the Bayesian unit root tests except for the one that is based on the posterior interval are higher than the ones committed by the frequentist unit root tests. Although the Bayesian test based on the posterior interval seems to have a small total error rate, the Type 1 error committed by this test is prominently large as seen in Table 3.2. For such small samples, among the frequentist tests, total error rate of the test based on $t_{\mu,(ADF/OLS),p_{min}}$ is smaller than the others. Note that, the total error rates of the tests based on the posterior interval and Bayes factor are both smaller, when we use a prior density like $U(-2,2)$.

Notice that, total error rates of the frequentist unit root tests based on the instrumental variable estimators are higher than all the other tests considered in SVM. Therefore, we make the following statements based on the rest of the frequentist unit root tests we consider.

We see that, for sample sizes like $n = 500$, all the Bayesian unit root tests we consider have less total error rates than the frequentist unit root tests. For larger samples, e.g. $n = 1000$, we observe the same situation between the Bayesian unit root tests based on the posterior interval and the frequentist tests, if the true value of

Table 5.1: Total Error Rates. *Sample Size* $n = 100$, *Nominal Level for the Frequentist Tests:* 0.05

Test Statistic	$\phi = 0.98$	$\phi = 0.95$
$t_{\mu,(ADF/OLS),p_{min}}$	0.88	0.74
$t_{\mu,(ADF/OLS),p_{full}}$	0.96	0.92
$t_{\mu,(ADF/WS),p_{full}}$	0.96	0.95
$t_{\mu,SDD}$	0.94	0.87
$t_{\mu,IV,Hall}$	0.99	0.99
$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.99	0.99
$t_{\mu,IV,SVM}(H_0)$	0.99	0.99
$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.94	0.93
$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.99	0.99
$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.98	0.98
$n(\hat{\phi}_{IV} - 1)$	0.94	0.82
$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.93	0.83
Posterior Interval (Mixed with $U(0,1)$, $p = 0.95$)	0.91	0.70
Posterior Interval (Mixed with $U(0,1)$, $p = 0.50$)	0.81	0.65
Posterior Interval ($\phi \sim U(-2, 2)$)	0.94	0.26
BF via Algorithm 2 (Mixed with $U(0,1)$, $p = 0.95$)	1.01	1.06
BF via Algorithm 2 (Mixed with $U(0,1)$, $p = 0.50$)	1.11	1.12
BF via Algorithm 2 ($\phi \sim U(0, 1)$)	1.01	1.08
BF via Algorithm 2 ($\phi \sim U(-2, 2)$)	0.85	0.76
Posterior Odds via Algorithm 2 (Mixed with $U(0,1)$, $p = 0.95$)	1.00	1.00
Posterior Odds via Algorithm 2 (Mixed with $U(0,1)$, $p = 0.50$)	1.11	1.12
Posterior Odds via Algorithm 2 ($\phi \sim U(0, 1)$)	1.01	1.08

the underlying persistence parameter is somewhat close to unity. We see that, when the sample size is sufficiently large, the Bayesian unit root test based on the posterior

Table 5.2: Total Error Rates. *Sample Size* $n = 500$, *Nominal Level for the Frequentist Tests:* 0.05

Test Statistic	$\phi = 0.98$	$\phi = 0.95$
$t_{\mu,(ADF/OLS),p_{min}}$	0.70	0.23
$t_{\mu,(ADF/OLS),p_{full}}$	0.86	0.62
$t_{\mu,(ADF/WS),p_{full}}$	0.77	0.58
$t_{\mu,SDD}$	0.77	0.35
$t_{\mu,IV,Hall}$	0.99	0.98
$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.99	0.98
$t_{\mu,IV,SVM}(H_0)$	1.02	1.03
$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.95	0.91
$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.98	0.97
$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.97	0.95
$n(\hat{\phi}_{IV} - 1)$	0.83	0.64
$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.84	0.65
Posterior Interval (Mixed with $U(0,1)$, $p = 0.95$)	0.65	0.17
Posterior Interval (Mixed with $U(0,1)$, $p \sim Beta(0.5, 0.5)$)	0.44	0.24
Posterior Interval ($p \sim U(0, 1.95)$)	0.64	0.13
Posterior Interval ($p \sim U(0, 20)$)	0.65	0.15
Posterior Interval ($p \sim U(-2, 2)$)	0.60	0.13
BF of So and Li ($\phi \sim U(0, 1)$)	0.68	0.22

interval along with the usage of a high prior probability on the point 1 for ϕ compete with the frequentist unit root tests in terms of the total error rates. We have the same situation between posterior interval test along with a prior density such as $U(-a,a)$ with $a > 1$ or $U(0,1+\epsilon)$ and the frequentist tests.

From Tables 5.1 and 5.2, it appears that the Bayesian test based on the posterior intervals with mixed density along with $p \sim Beta(0.5, 0.5)$ prior outperform all the tests for moderate to large samples. However, one should remember that the Type 1

Table 5.3: Total Error Rates. *Sample Size* $n = 1000$, *Nominal Level for the Frequentist Tests:* 0.05

Test Statistic	$\phi = 0.98$	$\phi = 0.95$
$t_{\mu,(ADF/OLS),p_{min}}$	0.35	0.05
$t_{\mu,(ADF/OLS),p_{full}}$	0.60	0.17
$t_{\mu,(ADF/WS),p_{full}}$	0.34	0.07
$t_{\mu,SDD}$	0.45	0.05
$t_{\mu,IV,Hall}$	0.96	0.95
$t_{\mu,IV,Hall}^{(dgp:SVM)}$	0.96	0.95
$t_{\mu,IV,SVM}(H_0)$	1.01	1.03
$t_{\mu,IV,SVM}(H_0 \text{ and } H_1)$	0.93	0.87
$t_{\mu,IV,SVM}^{(rst,SVM)}(H_0 \text{ and } H_1)$	0.94	0.92
$t_{\mu,IV,SVM}^{(dgp:SVM)}(H_0 \text{ and } H_1)$	0.93	0.90
$n(\hat{\phi}_{IV} - 1)$	0.68	0.50
$n(\hat{\phi}_{IV} - 1)^{(dgp:SVM)}$	0.66	0.50
Posterior Interval (Mixed with $U(0,1)$, $p = 0.95$)	0.25	0.04
Posterior Interval (Mixed with $U(0,1)$, $p \sim Beta(0.5, 0.5)$)	0.21	0.15
Posterior Interval ($p \sim U(0, 1.95)$)	0.21	0.08
Posterior Interval ($p \sim U(0, 20)$)	0.22	0.08
Posterior Interval ($p \sim U(-2, 2)$)	0.19	0.07
BF of So and Li ($\phi \sim U(0, 1)$)	0.46	0.12

error committed by this test is higher than the others.

As a future study, we plan to work on various unit root testing procedures in a SVM having one or more of the following properties;

- A nonzero correlation between η_t and u_t ,
- A more general autoregressive process for h_t ,
- A model that includes covariates via μ .

Bibliography

- Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions*. Dover Publications.
- Best, N. G., Cowles, M. K. & Vines, S. K. (1995). *CODA Convergence Diagnosis and Output Analysis Software for Gibbs Sampler Output*.
- Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons, Inc.
- Bollerslev, T. & Engle, R. (1993). Common persistence in conditional variance. *Econometrica* **61**(1), 167–186.
- Box, G. & Jenkins, G. M. (1976). *Time Series Analysis forecasting and control*. Holden-Day Series in Time Series Analysis and Digital Processing.
- Carter, C. & Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika* **81**, 541–553.
- Casella, G. & George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician* **46**(3), 167–174.
- Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association* **90**(432), 1137–1504.
- Chou, R. (1988). Volatility persistence and stock valuations: Some empirical evidence using garch. *Journal of Applied Econometrics* **3**, 279–294.

- Fuller, W. A. (1996). *Introduction to Statistical Time Series*. Wiley Series in Probability and Statistics. second edition edn. John Wiley & Sons, Inc.
- Gelman, A., C., J.B., S., H.S. & Rubin, D. (1995). *Bayesian Data Analysis*. Chapman and Hall, Texts in Statistical Sciences.
- Gilks, W., Richardson, S. & Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall.
- Gilks, W., Thomas, A. & Spiegelhalter, D. (1994). A language and a program for complex bayesian modeling. *The Statistician* **43**, 169–177.
- Harvey, A., Ruiz, E. & Shephard, N. (1994). Multivariate stochastic models. *Review of Economic Studies* **61**, 247–264.
- Jacquier, E., Polson, N. & Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* **12**(4), 371–388.
- Kim, S., Shephard, N. & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *Review of Economic Studies* **65**, 361–393.
- Kopka, H. & Daly, P. (1999). *A Guide to L^AT_EX*. Addison - Wesley.
- Mahieu, R. J. & Schotman, P. C. (1998). An empirical application of stochastic volatility models. *Journal of Applied Econometrics* **13**, 333–360.
- Meyer, R. & Yu, J. (2000). Bugs for a bayesian analysis of stochastic volatility models. *Econometrics Journal* **3**, 198–215.
- Pantula, S. G. (1991). Asymptotic distributions of unit-root tests when the process is nearly stationary. *Journal of Business & Economic Statistics* **9**(1), 63–71.
- Pindyck, R. (1984). Risk, inflation, and the stock market. *The American Economic Review* **74**(3), 335–351.

- Pindyck, R. S. (1988). Risk aversion and determinants of stock market behavior. *The Review of Economics and Statistics* **70**(2), 183–190.
- Poterba, J. M. & Summers, L. H. (1986). The persistence of volatility and stock market fluctuations. *The American Economic Review*.
- Said, S. E. & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **71**(3), 599–607.
- Said, S. E. & Dickey, D. A. (1985). Hypothesis testing in arima(p,1,q) models. *Journal of the American Statistical Association* **80**(390), 369–374. Theory and Methods.
- Sandmann, G. & Koopman, S. J. (1998). Estimation of stochastic volatility models via monte carlo maximum likelihood. *Journal of Econometrics* **87**, 271–301.
- Schwert, G. W. (1989). Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics* **7**(2), 147–159.
- So, M. K. P. & Li, W. K. (1999). Bayesian unit-root testing in stochastic volatility models. *Journal of Business & Economic Statistics* **17**, 491–496.
- Spiegelhalter, D., Thomas, A. & Best, N. (1999). *WinBUGS Version 1.2. User Manual*.
- Spiegelhalter, D., Thomas, A., Best, N. & Gilks, W. (1996a). *BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.5, (version ii)*.
- Spiegelhalter, D., Thomas, A., Best, N. & Gilks, W. (1996b). *BUGS Examples Volume 1, Version 0.5, (version ii)*.
- Spiegelhalter, D., Thomas, A., Best, N. & Gilks, W. (1996c). *BUGS Examples Volume 2, Version 0.5, (version ii)*.

APPENDIX

Appendix A

Properties of the Joint Posterior Density of the Parameters in SVM

In this part, we show that

1. The joint posterior density of the parameters is proper even when the joint prior density of them is improper,
2. Posterior density of ϕ has a nonzero mass on the point 1, when we use a mixed prior for ϕ .

First of all, as it is a common case in the SVM literature, we assume a priori that the parameters of our SVM are independent. We consider improper prior densities for μ and σ_η^2 as follows;

Prior density assigned for μ ; $f(\mu) \propto 1, \mu \in (-\infty, \infty)$

Prior density assigned for σ_η^2 ; $f(\sigma_\eta^2) \propto 1, \sigma_\eta^2 \in (0, \infty)$

We also assume that $\phi = b + (1 - b)U$ where $U \sim U(0, 1)$. That is, the prior density assigned for ϕ ;

$$f_\phi(\phi) \propto \begin{cases} p & \text{if } \phi = 1 \\ 1 - p & \text{if } 0 < \phi < 1 \\ 0 & \text{if elsewhere} \end{cases}$$

Note that at $\phi = 1$,

$$\begin{aligned}
\int f(\underline{h}|\mu, \sigma_\eta^2) d\mu d\sigma_\eta^2 &\propto \int \left(\frac{1}{\sigma_\eta^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma_\eta^2} \left[\sum_{t=2}^n (h_t - h_{t-1})^2 + (h_1 - \mu)^2 \right]} d\mu d\sigma_\eta^2 \\
&\propto \int \left(\frac{1}{\sigma_\eta^2}\right)^{\frac{n-1}{2}} e^{-\frac{1}{2\sigma_\eta^2} \left[\sum_{t=2}^n (h_t - h_{t-1})^2 \right]} d\sigma_\eta^2 d\underline{h} \\
&\propto \left(\sum_{t=2}^n (h_t - h_{t-1})^2 \right)^{-\frac{n-3}{2}}.
\end{aligned}$$

Now, first of all, let

$$\begin{aligned}
w(\phi) &= (1 - \phi^2) + (n - 1)(1 - \phi)^2 \\
B_1 &= \frac{\left[h_1(1 - \phi^2) + (1 - \phi) \sum_{t=2}^n (h_t - \phi h_{t-1}) \right]^2}{w(\phi)^2} \\
B_2 &= \frac{(1 - \phi^2)h_1^2 + \sum_{t=2}^n (h_t - \phi h_{t-1})^2}{w(\phi)} \\
B_3 &= (B_1)^{\frac{1}{2}}
\end{aligned}$$

Then, for $0 < \phi < 1$,

$$\begin{aligned}
\int f(\underline{h}|\mu, \sigma_\eta^2, \phi) d\mu d\sigma_\eta^2 &\propto \int (1 - \phi^2)^{\frac{1}{2}} \left(\frac{1}{\sigma_\eta^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2} \left[(1 - \phi^2)(h_1 - \mu)^2 + \sum_{t=2}^n (h_t - \mu - \phi(h_{t-1} - \mu))^2 \right]} \\
&\quad d\mu d\sigma_\eta^2 \\
&\propto \int (1 - \phi^2)^{\frac{1}{2}} \left(\frac{1}{\sigma_\eta^2}\right)^{\frac{n}{2}} e^{-\frac{w(\phi)}{2\sigma_\eta^2} (\mu - B_3)^2} e^{-\frac{w(\phi)}{2\sigma_\eta^2} (-B_1 + B_2)} d\mu d\sigma_\eta^2 \\
&\propto \int (1 - \phi^2)^{\frac{1}{2}} \left(\frac{1}{\sigma_\eta^2}\right)^{\frac{n-1}{2}} w(\phi)^{-\frac{1}{2}} e^{-\frac{w(\phi)}{2\sigma_\eta^2} (B_2 - B_1)} d\sigma_\eta^2 \\
&\propto (1 - \phi^2)^{\frac{1}{2}} (w(\phi))^{\frac{n-4}{2}} \\
&\quad \left\{ w(\phi) \left((1 - \phi^2)h_1^2 + \sum_{t=2}^n (h_t - \phi h_{t-1})^2 \right) - \left(h_1(1 - \phi^2) + (1 - \phi) \sum_{t=2}^n (h_t - \phi h_{t-1}) \right)^2 \right\}^{-\frac{n-3}{2}}.
\end{aligned}$$

Now, the integral of the joint posterior density of the parameters over the parameter space is,

$$\begin{aligned}
\int f(\mu, \sigma_\eta^2, \phi | \underline{r}) d\mu d\sigma_\eta^2 d\phi &\propto \int \int f(\mu, \sigma_\eta^2, \phi, \underline{h} | \underline{r}) d\mu d\sigma_\eta^2 d\phi d\underline{h} \\
&\propto \int \int f(\underline{r} | \underline{h}) f(\underline{h} | \mu, \sigma_\eta^2, \phi) f(\mu) f(\sigma_\eta^2) f(\phi) d\mu d\sigma_\eta^2 d\phi d\underline{h} \\
&\propto p \int f(\underline{r} | \underline{h}) \left(\int f(\underline{h} | \mu, \sigma_\eta^2) d\mu d\sigma_\eta^2 \right) d\underline{h} \\
&\quad + (1-p) \int f(\underline{r} | \underline{h}) \left(\int f(\underline{h} | \mu, \sigma_\eta^2, \phi) d\mu d\sigma_\eta^2 \right) d\underline{h} d\phi \\
&\propto p \int f(\underline{r} | \underline{h}) \left(\sum_{t=2}^n (h_t - h_{t-1})^2 \right)^{-\frac{n-3}{2}} d\underline{h} \\
&\quad + (1-p) \int (1 - \phi^2)^{\frac{1}{2}} (w(\phi))^{\frac{n-4}{2}} f(\underline{r} | \underline{h}) \\
&\quad \left\{ w(\phi) \left((1 - \phi^2) h_1^2 + \sum_{t=2}^n (h_t - \phi h_{t-1})^2 \right) - \left(h_1 (1 - \phi^2) + (1 - \phi) \sum_{t=2}^n (h_t - \phi h_{t-1}) \right)^2 \right\}^{-\frac{n-3}{2}} \\
&\quad d\underline{h} d\phi
\end{aligned}$$

Note that,

(i) The likelihood of the mean corrected returns given the log-volatilities is,

$$f(\underline{r} | \underline{h}) = e^{-\frac{1}{2} \sum_{t=1}^n (h_t + r_t^2 e^{-h_t})}$$

(ii) Let $g(\cdot)$ be a quadratic function of h_t . Then,

$$\begin{aligned}
\lim_{h_t \rightarrow \infty} e^{-\frac{1}{2}(h_t + r_t^2 e^{-h_t})} \frac{1}{(g(h_t))^\alpha} &= \lim_{h_t \rightarrow -\infty} e^{-\frac{1}{2}(h_t + r_t^2 e^{-h_t})} \frac{1}{(g(h_t))^\alpha} \\
&= 0 \quad \forall \alpha
\end{aligned}$$

(iii) $\sum_{t=2}^n (h_t - h_{t-1})^2$ is positive almost surely,

(iv) $w(\phi) \left((1 - \phi^2) h_1^2 + \sum_{t=2}^n (h_t - \phi h_{t-1})^2 \right) - \left(h_1 (1 - \phi^2) + (1 - \phi) \sum_{t=2}^n (h_t - \phi h_{t-1}) \right)^2$ is positive almost surely.

Proof of (iii): Note that,

$$\sum_{t=2}^n (h_t - h_{t-1})^2 \geq 2(n-2) \min_{2 \leq t \leq n} (h_t - h_{t-1})^2$$

and,

$$P(\min_{2 \leq t \leq n} (h_t - h_{t-1})^2 = 0) = 0$$

as $\{h_t\}$ are continuous random variables. Hence, the result follows.

Proof of (iv): Let

$$\begin{aligned} a_1 &= h_1(1 - \phi^2)^{\frac{1}{2}} & a_t &= h_t - \phi h_{t-1} \text{ for } t = 2, \dots, n \\ b_1 &= (1 - \phi^2)^{\frac{1}{2}} & b_t &= 1 - \phi \text{ for } t = 2, \dots, n \end{aligned}$$

Then,

$$\begin{aligned} w(\phi) &\left((1 - \phi^2)h_1^2 + \sum_{t=2}^n (h_t - \phi h_{t-1})^2 \right) - \left(h_1(1 - \phi^2) + (1 - \phi) \sum_{t=2}^n (h_t - \phi h_{t-1}) \right)^2 \\ &= \left(\sum_{t=1}^n a_t^2 \right) \left(\sum_{t=1}^n b_t^2 \right) - \left(\sum_{t=1}^n a_t b_t \right)^2 \\ &= \sum_{1 \leq t < j \leq n}^1 (a_t b_j - a_j b_t)^2 \\ &\geq (a_1 b_2 - a_2 b_1)^2 \end{aligned}$$

Note that,

$$P((a_1 b_2 - a_2 b_1)^2 = 0) = P(h_1 - h_2 = 0) = 0$$

as $\{h_t\}$ are continuous random variables. Hence, the result follows.

The results in (i), (ii), (iii), and (iv) altogether imply that

$$\int f(\mu, \sigma_\eta^2, \phi | \mathcal{L}) d\mu d\sigma_\eta^2 d\phi < \infty$$

That is, the joint posterior density of the parameters are proper even one uses improper prior densities for the parameters μ and σ_η^2 .

¹ By the Lagrange's identity for real numbers; $(\sum_{t=1}^n a_t b_t)^2 = (\sum_{t=1}^n a_t^2) (\sum_{t=1}^n b_t^2) - \sum_{1 \leq t < j \leq n} (a_t b_j - a_j b_t)^2$

Note that, the marginal posterior density of ϕ is,

$$f(\phi|\underline{r}) \propto f(\underline{r}|\phi)f(\phi)$$

and

$$f(\underline{r}|\phi) \propto \int f(\underline{r}|\underline{h}) \left(\int f(\underline{h}|\mu, \sigma_\eta^2, \phi) d\mu d\sigma_\eta^2 \right) d\underline{h}$$

Then, using the results derived previously, the marginal posterior density of ϕ is,

$$f(\phi|\underline{r}) \propto \begin{cases} p \int e^{-\frac{1}{2} \sum_{t=1}^n (h_t + r_t^2 e^{-h_t})} \left(\sum_{t=2}^n (h_t - h_{t-1})^2 \right)^{-\left(\frac{n-3}{2}\right)} d\underline{h} & \text{if } \phi = 1 \\ (1-p) \left(\frac{1-\phi^2}{w(\phi)} \right)^{\frac{1}{2}} \int e^{-\frac{1}{2} \sum_{t=1}^n (h_t + r_t^2 e^{-h_t})} B^{-\left(\frac{n-3}{2}\right)} d\underline{h} & \text{if } 0 < \phi < 1 \\ 0 & \text{if elsewhere} \end{cases}$$

where

$$B = w(\phi) \left((1-\phi^2)h_1^2 + \sum_{t=2}^n (h_t - \phi h_{t-1})^2 \right) - \left(h_1(1-\phi^2) + (1-\phi) \sum_{t=2}^n (h_t - \phi h_{t-1}) \right)^2$$

Appendix B

Smoothing the Percentiles

We fit the model

$$t_i = \lambda_1 - \lambda_2 \theta^{\lambda_3} + \xi_i \quad (\text{B.1})$$

where $\lambda_2 > 0$, $\lambda_3 > 0$ and $\xi_i \sim NID(0, \sigma_\xi^2)$ to smooth the percentiles over the values of θ . In this model, t stands for the percentile of the unit root test statistic of interest. We fit this nonlinear regression to the empirical percentiles we computed by using SAS/PROC NLIN. Least square estimates of the parameters and their standard errors are given in the following tables.

Table B.1: Estimates of the smoothing parameters $(\lambda_1, \lambda_2, \lambda_3)$. ($n=100$)

nominal level	test statistic	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$
0.05	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{min})$	-2.76 (0.08)	4.63 (0.130)	3.31 (0.24)
	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{full})$	-2.00 (0.01)	1.60 (0.03)	5.60 (0.28)
	$\tau_{\mu,ADF/WS,n,\theta}, (p_{full})$	-2.70 (0.01)	0.51 (0.04)	23.76 (4.45)
	$\tau_{\mu,SDD,n,\theta}$	-2.81 (0.08)	4.08 (0.19)	6.31 (0.65)
	$\tau_{\mu,IV,Hall,n,\theta}$	-2.62 (0.02)	1.62 (0.06)	7.25 (0.59)
0.10	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{min})$	-2.46 (0.09)	4.67 (0.14)	3.57 (0.29)
	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{full})$	-1.80 (0.01)	1.46 (0.03)	6.21 (0.32)
	$\tau_{\mu,ADF/WS,n,\theta}, (p_{full})$	-2.35 (0.007)	0.55 (0.03)	21.68 (2.58)
	$\tau_{\mu,SDD,n,\theta}$	-2.52 (0.05)	3.28 (0.14)	7.14 (0.67)
	$\tau_{\mu,IV,Hall,n,\theta}$	-2.39 (0.02)	0.87 (0.08)	9.66 (1.85)

Approximate Standard Errors of the estimators of the parameters are given in the parantheses.

Table B.2: Estimates of the smoothing parameters $(\lambda_1, \lambda_2, \lambda_3)$. ($n=500$)

nominal level	test statistic	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$
0.05	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{min})$	-2.83 (0.29)	13.22 (0.89)	10.75 (1.47)
	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{full})$	-2.70 (0.01)	4.19 (0.04)	15.70 (0.35)
	$\tau_{\mu,ADF/WS,n,\theta}, (p_{full})$	-2.54 (0.002)	2.26 (0.02)	36.75 (0.62)
	$\tau_{\mu,SDD,n,\theta}$	-2.87 (0.007)	11.72 (0.03)	25.29 (0.15)
	$\tau_{\mu,IV,Hall,n,\theta}$	-2.66 (0.20)	2.04 (0.23)	2.01 (0.66)
0.10	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{min})$	-2.54 (0.31)	13.39 (1.02)	11.88 (1.82)
	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{full})$	-2.00 (0.03)	1.60 (0.04)	5.60 (0.37)
	$\tau_{\mu,ADF/WS,n,\theta}, (p_{full})$	-2.26 (0.003)	2.34 (0.02)	39.27 (0.90)
	$\tau_{\mu,SDD,n,\theta}$	-2.60 (0.007)	9.84 (0.04)	26.96 (0.21)
	$\tau_{\mu,IV,Hall,n,\theta}$	-2.38 (0.30)	1.57 (0.31)	1.78 (1.04)

Approximate Standard Errors of the estimators of the parameters are given in the parantheses.

Table B.3: Estimates of the smoothing parameters $(\lambda_1, \lambda_2, \lambda_3)$. ($n=1000$)

nominal level	test statistic	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$
0.05	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{min})$	-2.93 (0.09)	19.82 (0.41)	20.27 (0.84)
	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{full})$	-2.80 (0.01)	5.75 (0.08)	26.20 (0.78)
	$\tau_{\mu,ADF/WS,n,\theta}, (p_{full})$	-2.55 (0.001)	3.54 (0.01)	53.35 (0.41)
	$\tau_{\mu,SDD,n,\theta}$	-2.91 (0.03)	18.36 (0.25)	38.38 (1.00)
	$\tau_{\mu,IV,Hall,n,\theta}$	-2.63 (0.39)	2.91 (0.47)	2.30 (1.07)
0.10	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{min})$	-2.71 (0.05)	20.88 (0.30)	26.54 (0.76)
	$\tau_{\mu,ADF/OLS,n,\theta}, (p_{full})$	-2.51 (0.01)	5.43 (0.07)	30.31 (0.87)
	$\tau_{\mu,ADF/WS,n,\theta}, (p_{full})$	-2.24 (0.001)	3.56 (0.01)	52.79 (0.45)
	$\tau_{\mu,SDD,n,\theta}$	-2.60 (0.02)	14.61 (0.17)	36.57 (0.83)
	$\tau_{\mu,IV,Hall,n,\theta}$	-2.34 (0.47)	2.45 (0.56)	2.21 (1.45)

Approximate Standard Errors of the estimators of the parameters are given in the parantheses.

One can use these nonlinear regression equations in order to get the corresponding cutoff value that corresponds to an estimate of θ .

Figures B.1-B.13 display the plots of the calculated empirical percentiles as well as the prediction curves.

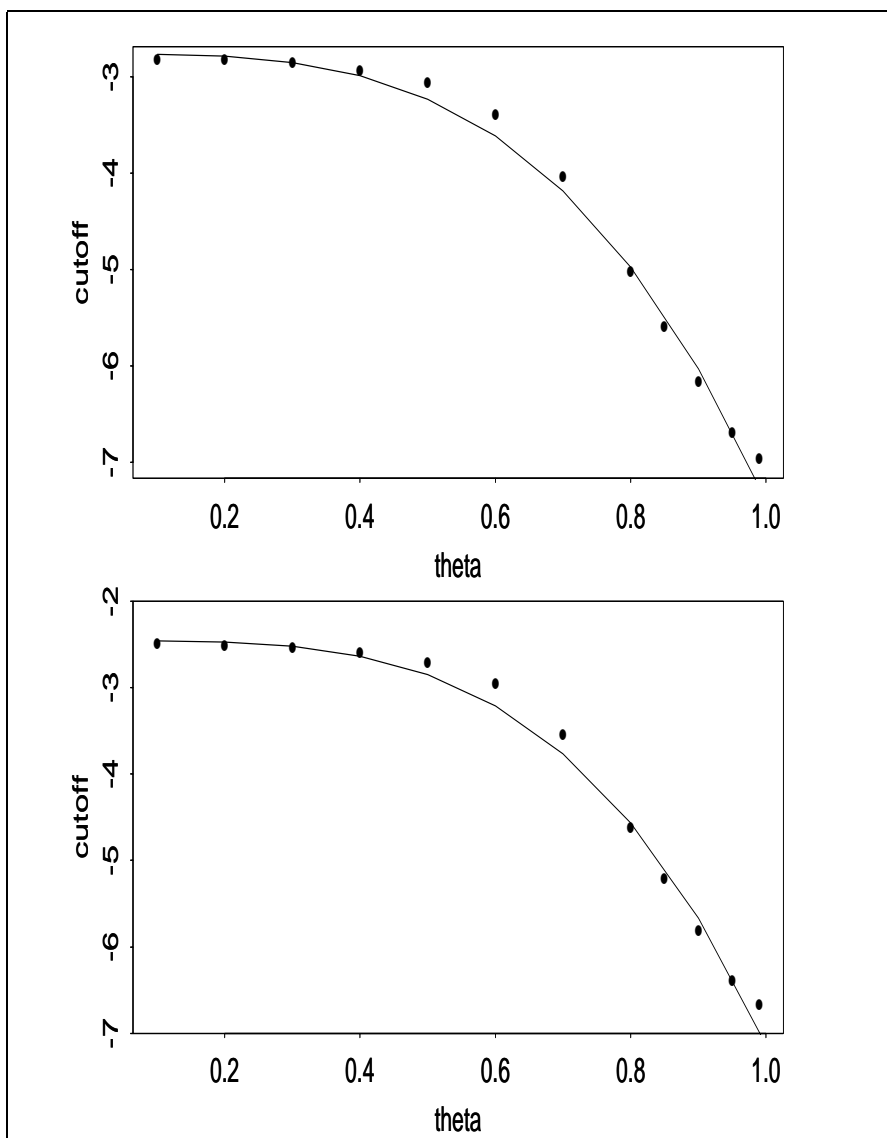


Figure B.1: Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}(p_{min}), n = 100$. *Top figure*: level of the test=0.05, *Bottom figure*: level of the test=0.10

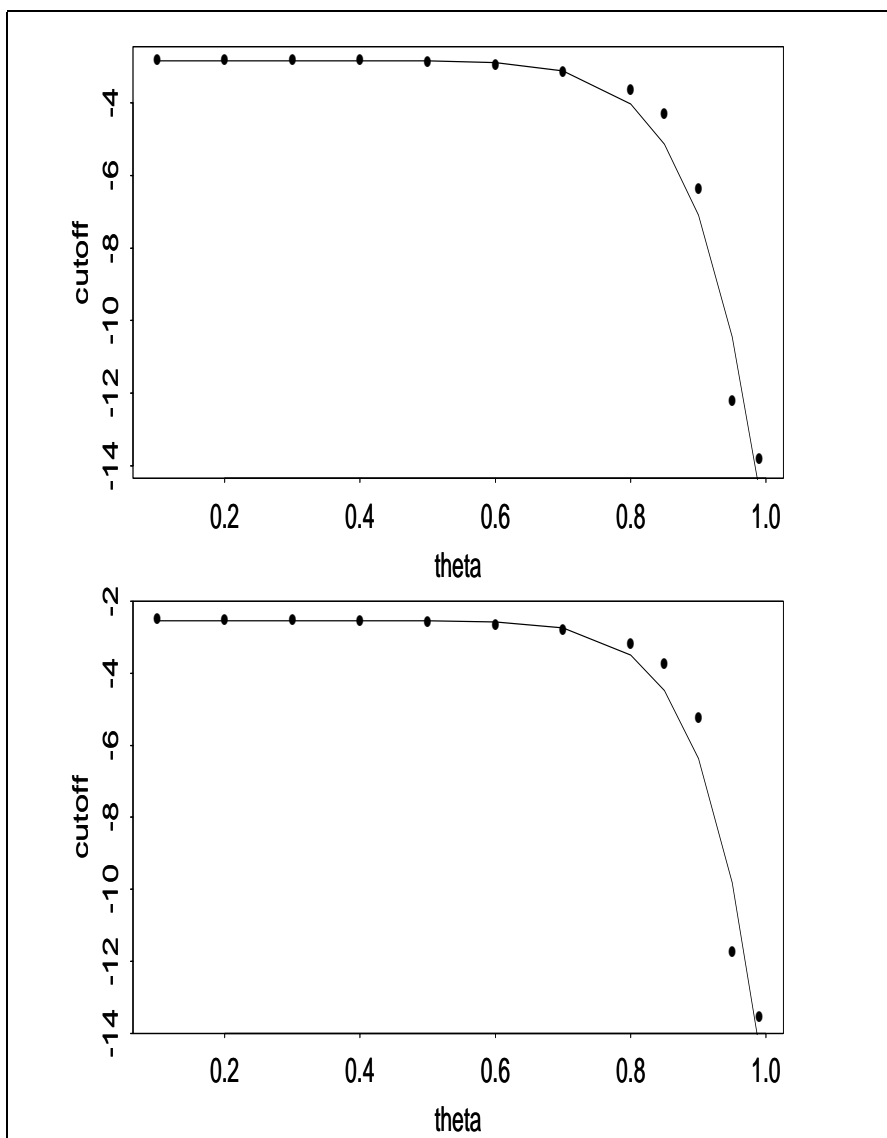


Figure B.2: Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}(p_{min}), n = 500$. *Top figure*: level of the test=0.05, *Bottom figure*: level of the test=0.10

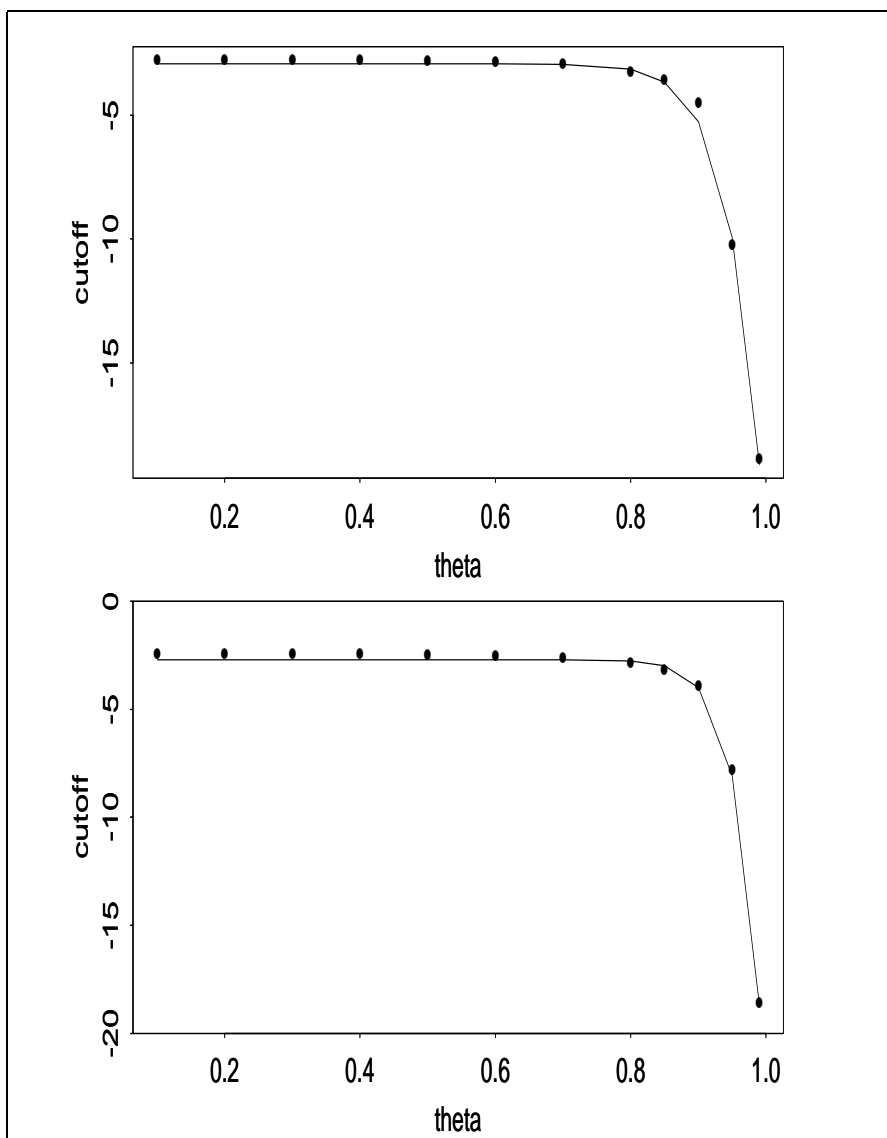


Figure B.3: Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}(p_{min}), n = 1000$. *Top figure:* level of the test=0.05, *Bottom figure:* level of the test=0.10

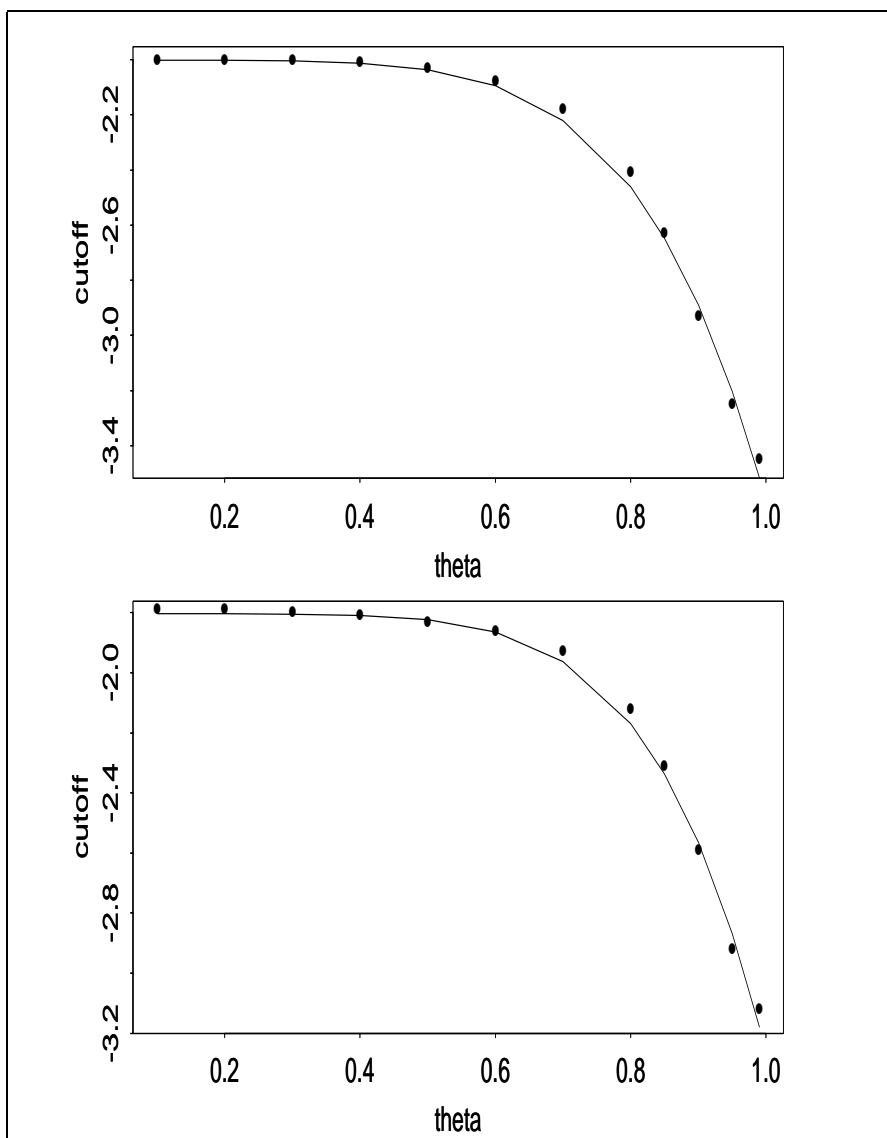


Figure B.4: Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}, (p_{full}), n = 100$. *Top figure*: level of the test=0.05, *Bottom figure*: level of the test=0.10

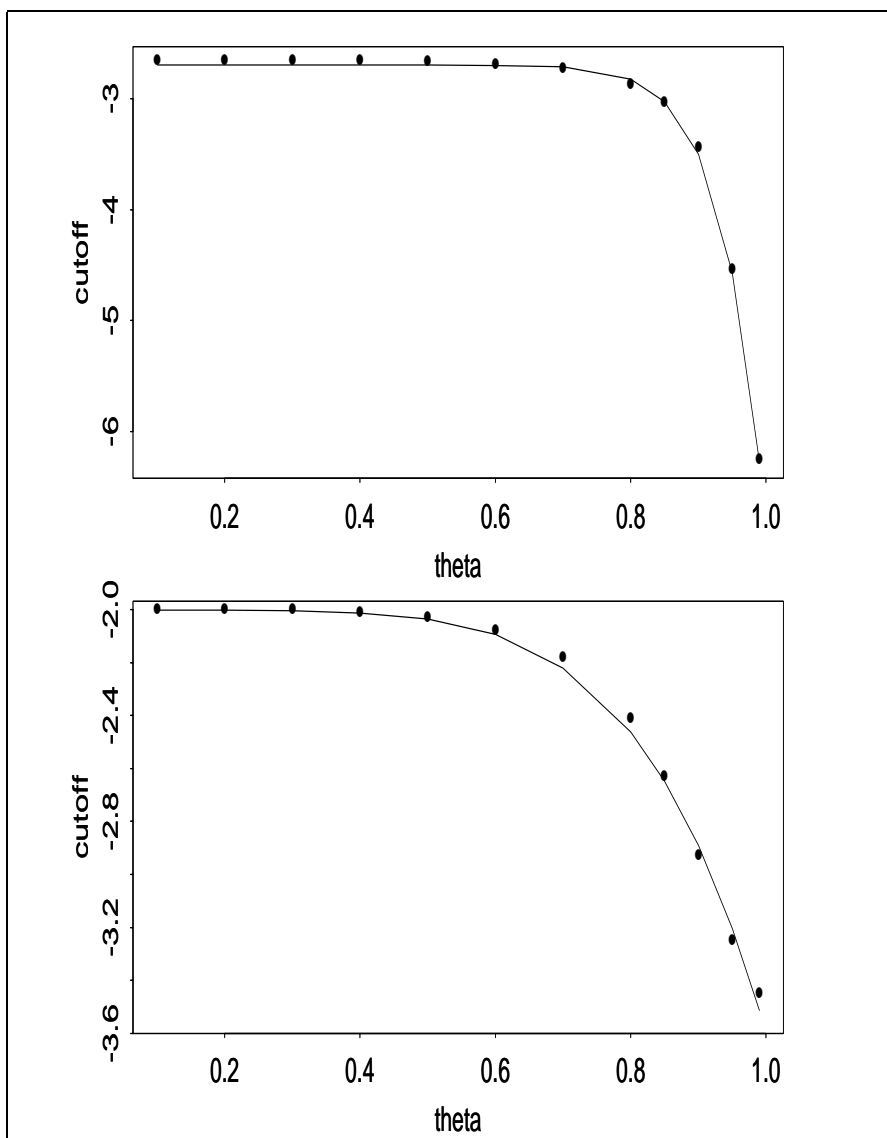


Figure B.5: Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}, (p_{full}), n = 500$. *Top figure*: level of the test=0.05, *Bottom figure*: level of the test=0.10

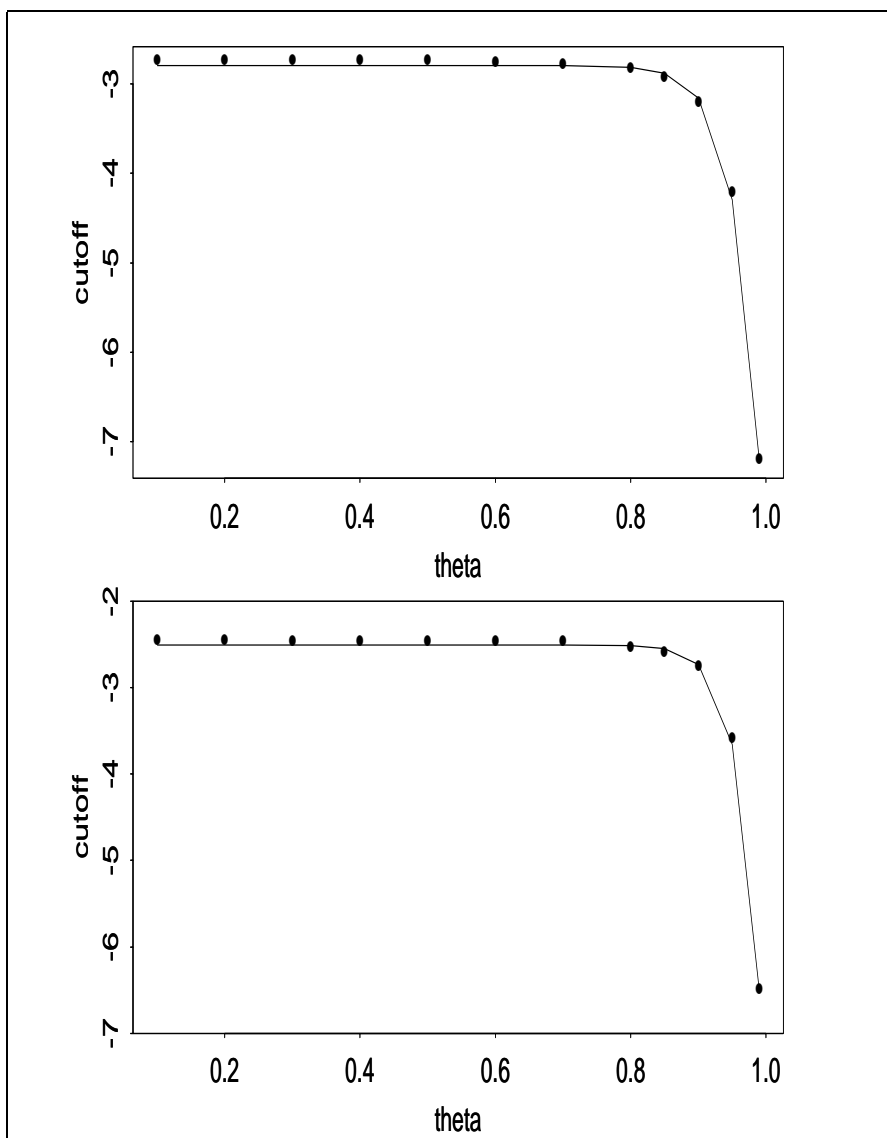


Figure B.6: Prediction Curve for $\tau_{\mu,ADF/OLS,n,\theta}, (p_{full}), n = 1000$. *Top figure:* level of the test=0.05, *Bottom figure:* level of the test=0.10

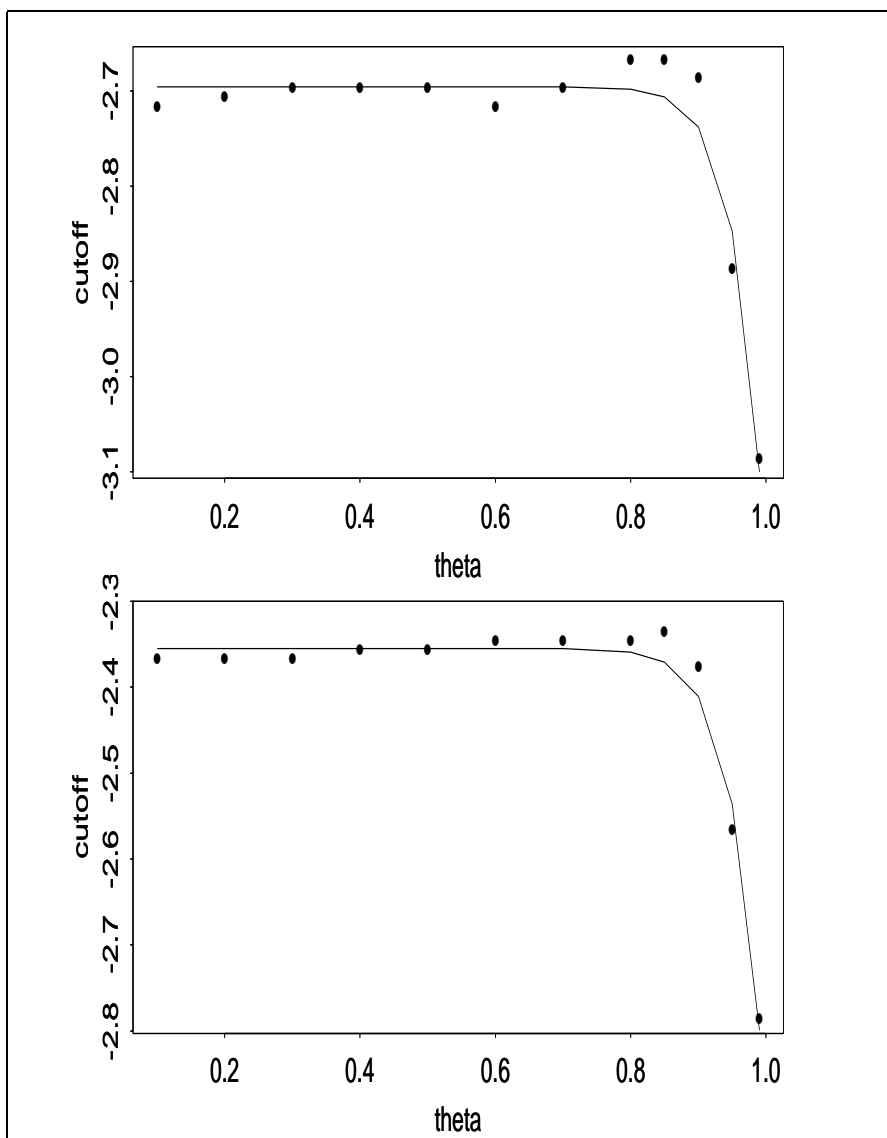


Figure B.7: Prediction Curve for $\tau_{\mu,ADF/WS,n,\theta}, (p_{full}), n = 100$. *Top figure*: level of the test=0.05, *Bottom figure*: level of the test=0.10

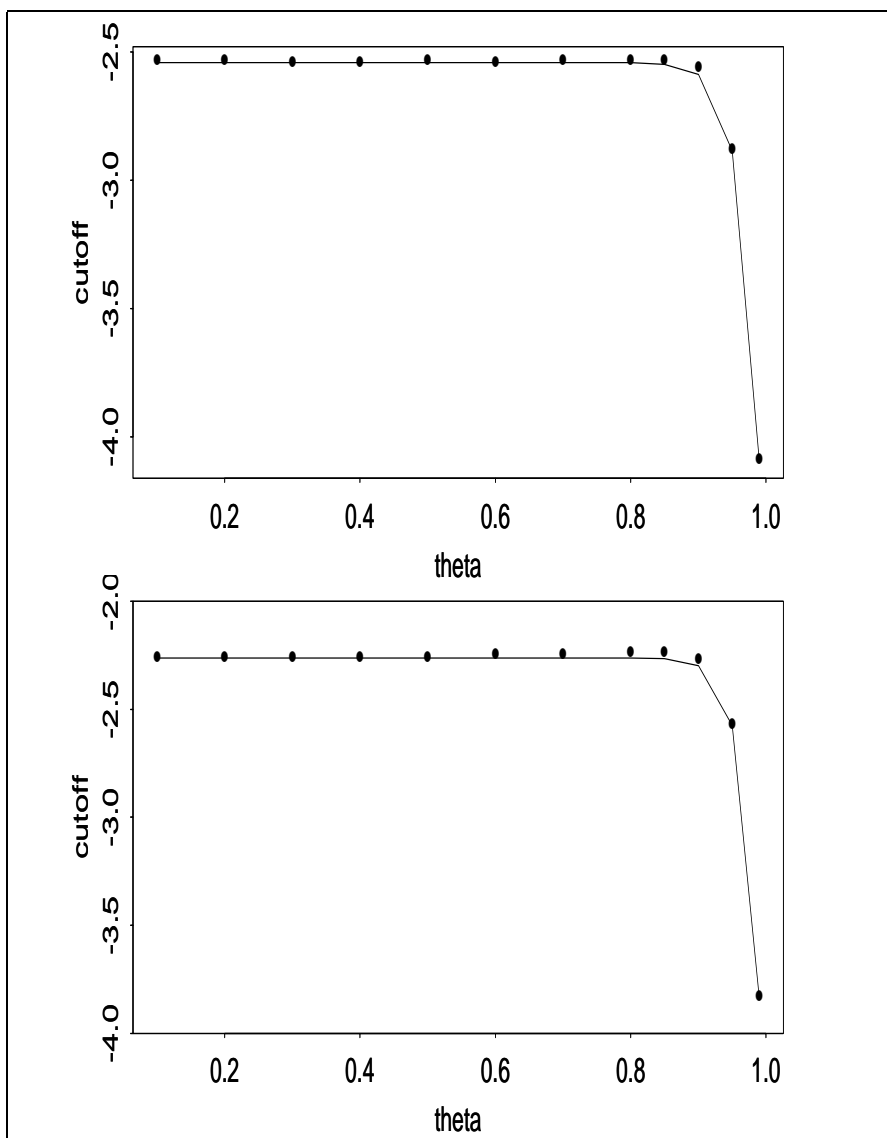


Figure B.8: Prediction Curve for $\tau_{\mu,ADF/WS,n,\theta,(p_{full}),n = 500}$. *Top figure:* level of the test=0.05, *Bottom figure:* level of the test=0.10

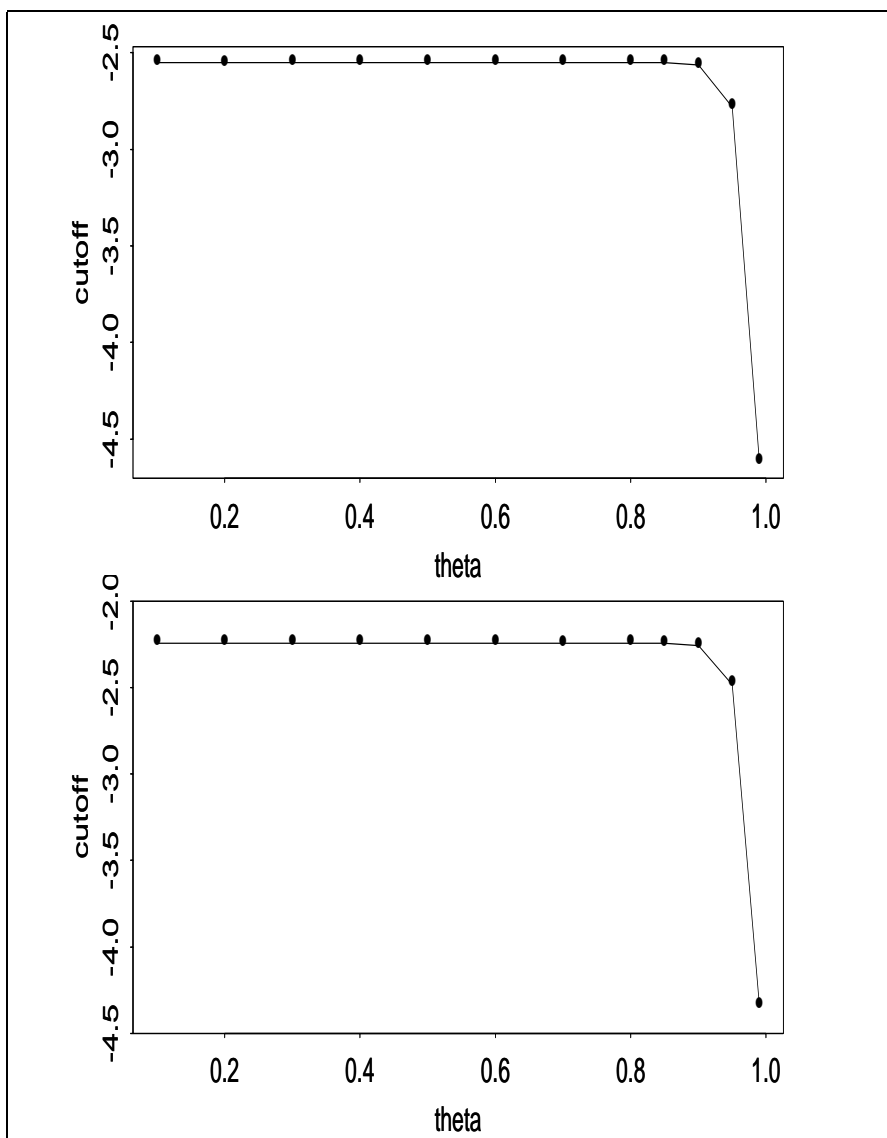


Figure B.9: Prediction Curve for $\tau_{\mu,ADF/WS,n,\theta}, (p_{full}), n = 1000$. *Top figure:* level of the test=0.05, *Bottom figure:* level of the test=0.10

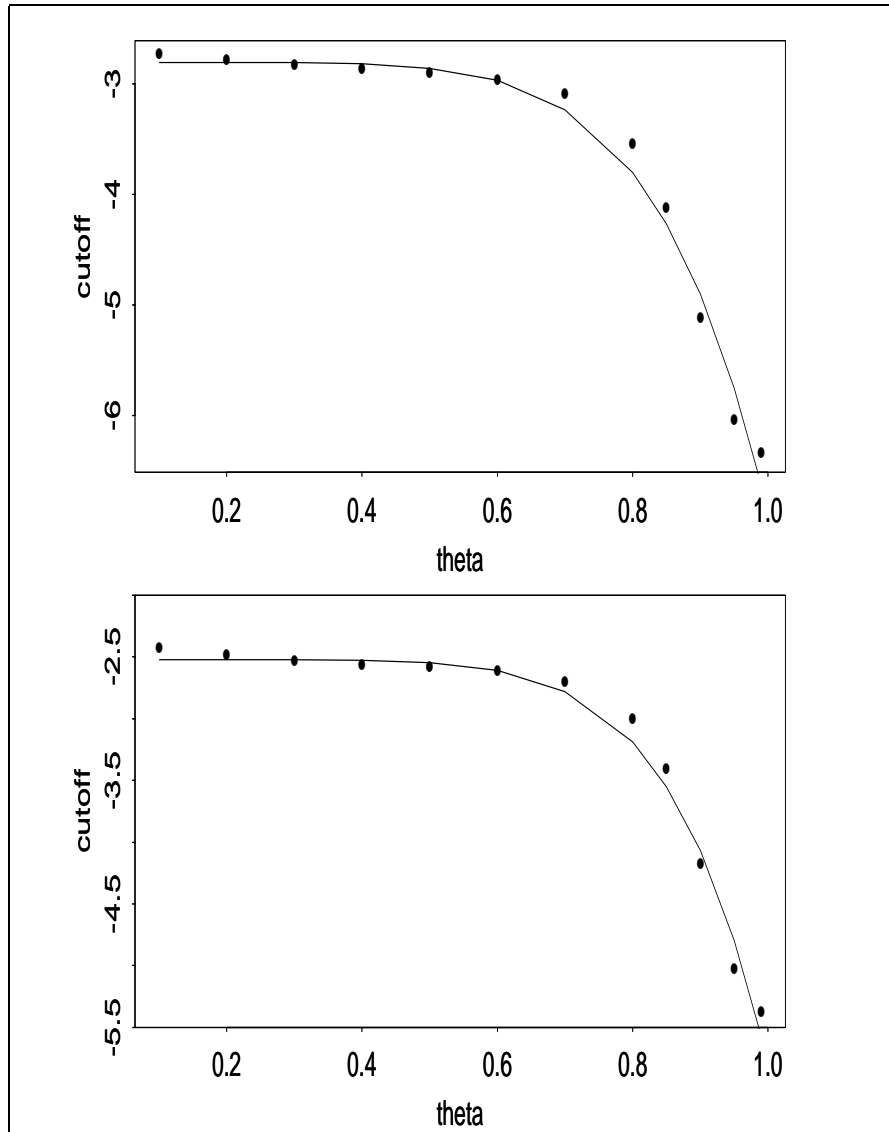


Figure B.10: Prediction Curve for $\tau_{\mu,SDD,n,\theta} = 100$. *Top figure:* level of the test=0.05, *Bottom figure:* level of the test=0.10

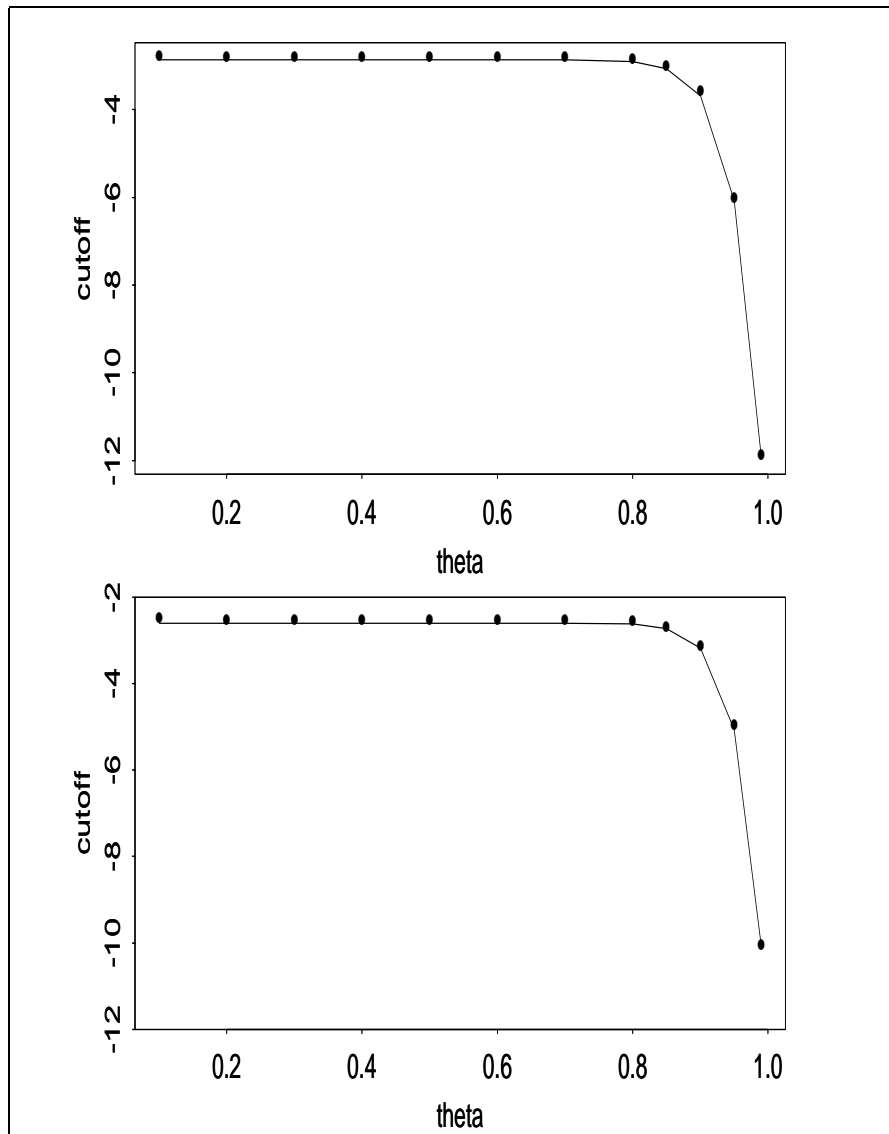


Figure B.11: Prediction Curve for $\tau_{\mu, SDD, n, \theta, n} = 500$. *Top figure:* level of the test=0.05, *Bottom figure:* level of the test=0.10

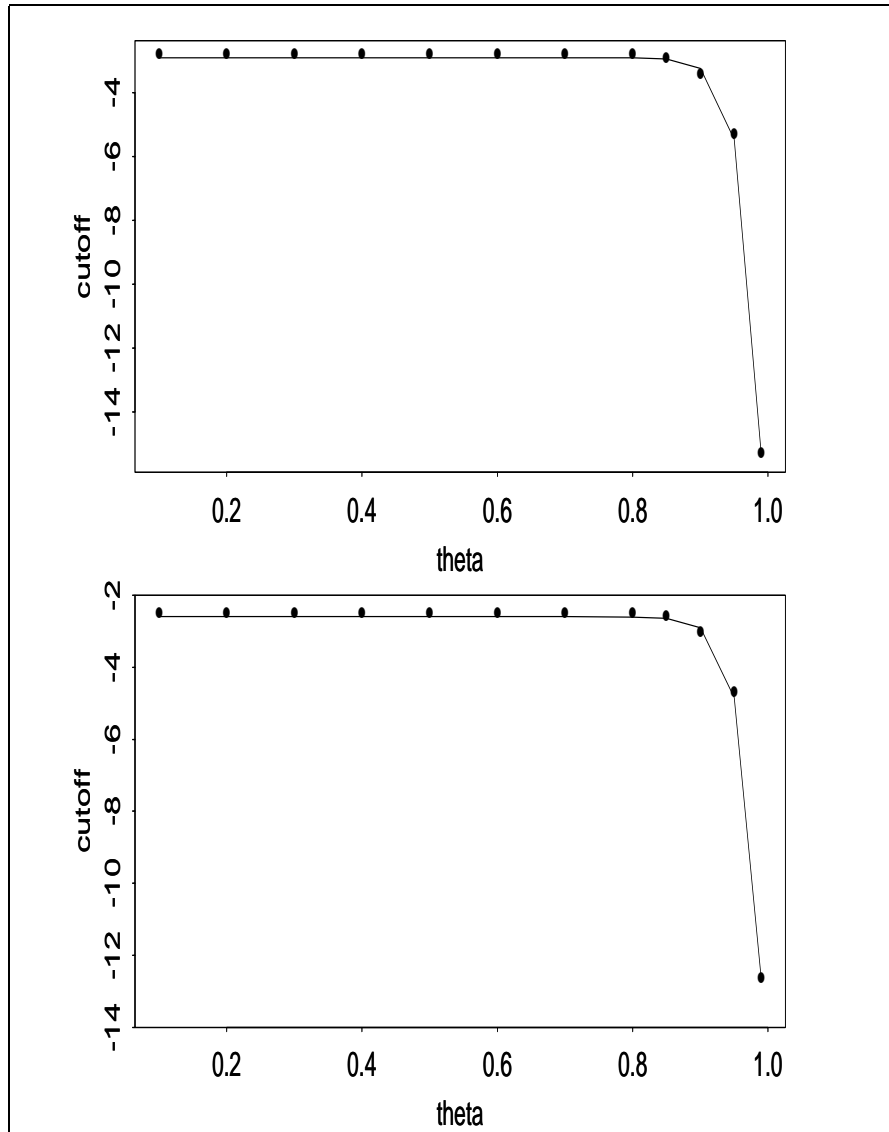


Figure B.12: Prediction Curve for $\tau_{\mu, SDD, n, \theta, n} = 1000$. *Top figure:* level of the test=0.05, *Bottom figure:* level of the test=0.10

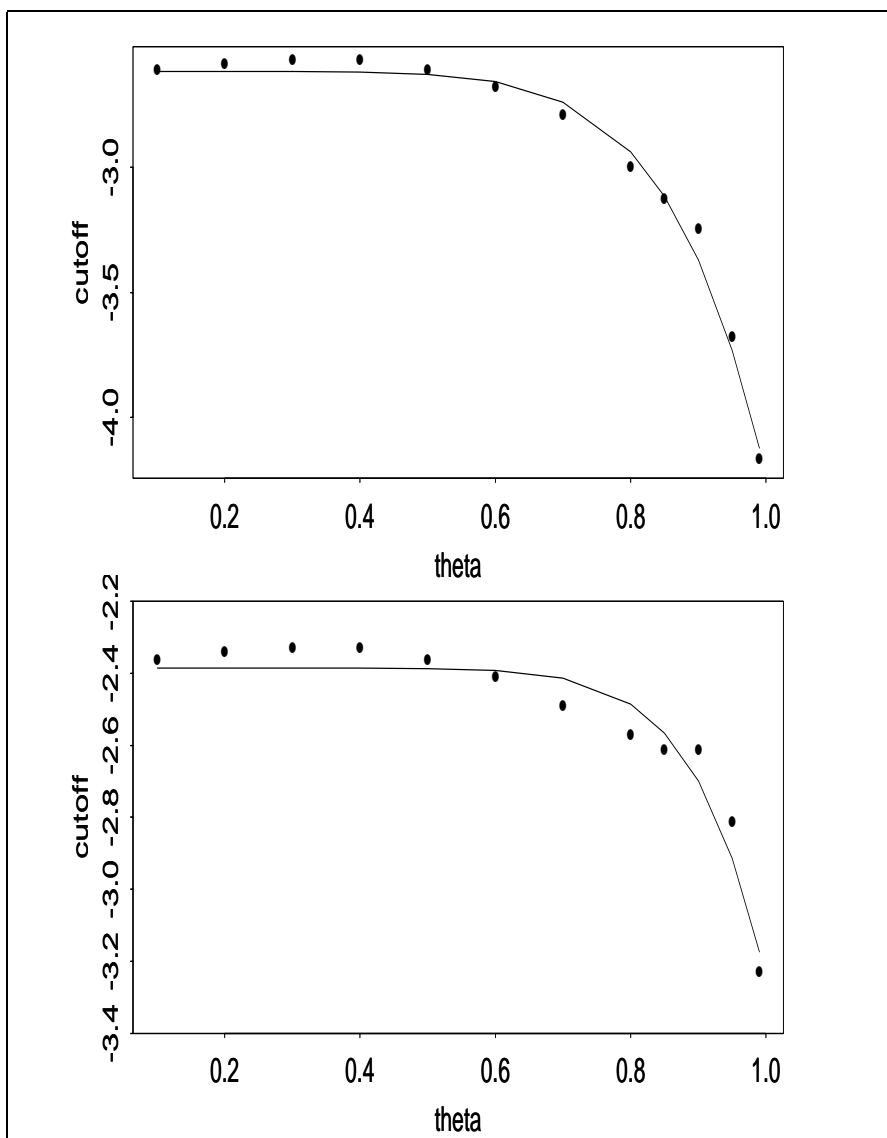


Figure B.13: Prediction Curve for $\tau_{\mu,IV,Hall,n,\theta}, n = 100$. *Top figure:* level of the test=0.05, *Bottom figure:* level of the test=0.10

Appendix C

The Sampling Distribution of $\hat{\theta}$

When we apply frequentist unit root tests in SVM, we need to use an estimate of θ to determine the critical point. In Chapter 2, we considered three types of estimate of θ ; $\hat{\theta}_{durbin}$, $\hat{\theta}_{adj.durbin}$, and $\hat{\theta}_{one\ step\ gn}$. Below, we list the sampling distribution of these estimators. Data are generated from SVM with $\sigma_{\eta}^2 = 0.1$ and the value of ϕ given in the first column of the tables below.

Note that if the true value of ϕ is 1, corresponding true value of θ is 0.87. Similarly, it is 0.86 if the true value of ϕ is 0.98 and 0.95. These are obtained by the relationships given in equations (2.3) and (2.4).

Table C.1: Sampling Distribution of $\hat{\theta}_{durbin}$. (1,000 Monte Carlo Replications)

ϕ	n	Monte Carlo Average	Monte Carlo Standard Error
1	100	0.73	0.15
	500	0.82	0.03
	1000	0.82	0.02
0.98	100	0.71	0.18
	500	0.81	0.04
	1000	0.81	0.02
0.95	100	0.70	0.19
	500	0.79	0.06
	1000	0.80	0.03

Table C.2: Sampling Distribution of $\hat{\theta}_{adj.durbin}$. (1,000 Monte Carlo Replications)

ϕ	n	Monte Carlo Average	Monte Carlo Standard Error
1	100	0.96	0.12
	500	0.88	0.03
	1000	0.87	0.02
0.98	100	0.97	0.14
	500	0.88	0.04
	1000	0.87	0.02
0.95	100	0.99	0.15
	500	0.89	0.05
	1000	0.87	0.03

Table C.3: Sampling Distribution of $\hat{\theta}_{one\ step\ gn}$. (1,000 Monte Carlo Replications)

ϕ	n	Monte Carlo Average	Monte Carlo Standard Error
1	100	0.75	0.16
	500	0.86	0.03
	1000	0.86	0.02
0.98	100	0.73	0.19
	500	0.84	0.04
	1000	0.85	0.03
0.95	100	0.70	0.20
	500	0.81	0.07
	1000	0.83	0.04