DISSERTATION

Titel der Dissertation

„Stochastic Unit-Root Models in Economics –
Essays on Testing, Estimating and Forecasting“

Verfasser

Mag. Jürgen Holl

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 094 140
Dissertationsgebiet lt. Studienblatt:	Volkswirtschaftslehre (Economics)
Betreuer:	Univ.-Prof. Dipl.-Ing. Dr. Robert M. Kunst

Contents

List of Figures

List of Tables

Acknowledgments

I am grateful to my first supervisor Robert M. Kunst who initiated my interest in time series analysis. His comments were not only restricted to statistical issues. Carefully reading through my manuscripts he also gave feedback concerning language and presentation. I would like to thank my second supervisor Erhard Reschenhofer for valuable comments. His lectures provided me with technical skills essential for studying contemporary literature in time series analysis.

Financial support by the University of Vienna is gratefully acknowledged. My PhD-Studies were financed as part of the initiative group *Issues in the Global Economy: Dynamics, Governance and Information*. Initiative groups are structured doctoral study programs offered by the University of Vienna to promote early stage researchers.

Introduction

The present dissertation contains three essays aimed at studying the relevance of so-called stochastic unit-root (STUR) models as introduced in Granger and Swanson (1997) for economics. What makes a certain statistical model relevant for modeling economic data? In fact, this depends on the data. To reformulate the question: Do we believe that real economic data are generated by the underlying statistical process? The usual way to answer such questions is to rely on the results of statistical testing devices prepared to discriminate between different model classes. That is, we check whether adequate hypothesis tests are available.

In economics it is common sense that lots of time series have a nonstationary nature. Trending behavior either following deterministic or stochastic patterns, seasonal cycles or changing volatilities are observed in various economic areas. Traditionally, nonstationary behavior resulting in a stochastic trend is captured by fixed-coefficient unit-root processes. Several tests have been developed to distinguish that kind of nonstationary processes from stationary ones. Particularly, fixed-coefficient unit-root processes can be differenced to stationarity. In contrast, stochastic unit-root processes are unit-root processes driven by random coefficients allowing for changes between stationary and explosive regimes. Thus, they are nonstationary, though not difference-stationary. As conventional cointegration and error-correction techniques assume difference stationarity, the presence of STUR processes may have serious consequences. The special case where the STUR model in Granger and Swanson (1997) has a coefficient mean equal to one is known as a stochastic unit root. In what follows, STUR models or STUR processes having a stochastic unit root are referred to as such explicitly.

Certain fixed-coefficient models are nested in the STUR model. This is where the *first essay* comes in. There, a new test of the null hypothesis of a unit root is introduced to distinguish stochastic unit-root models from conventional time series models. The test is based on a deviance statistic calculated from pseudo-likelihood functions. Under the alternative hypothesis stochastic unit-root models as well as stationary models are allowed. The asymptotic distribution of the test statistic is derived and critical values are retrieved from the empirical distribution function. Size and power simulations are presented to assess the performance of the new STUR test against previously developed competitors. Competing tests also assume a unit root under the null hypothesis. Whether they have power against stochastic unit-root processes and/or stationary processes is studied in detail. Finally, the STUR test and its competitors are applied to unemployment rates of ten countries. Some evidence is provided in favor of nonstationary processes for most of the countries. For some countries STUR seems to be an attractive option.

In course of the *second essay* we fit STUR models to unemployment rates to evaluate the results of the tests concerning STUR alternatives. Whether it is reasonable from an economic point of view to model unemployment rates using STUR is discussed as well. Different estimation techniques are considered. As the coefficient process is unobserved, we apply a Bayesian procedure which allows us not only to estimate the STUR parameters but also to simulate the distribution of the coefficient process. Looking at the coefficient distribution enables us to measure how close the STUR model comes to a fixed-coefficient model. All STUR estimates are significantly different from zero. Coefficient distributions are in line with the test results for six out of ten countries in favor of nonstationary models. For four of them STUR models are suggested. The Bayesian procedure applied is very time-consuming. An alternative estimator to obtain estimates of coefficient mean and variance is successfully evaluated in simulations.

Testing for the null hypothesis of a unit root in unemployent rates using tests having power against STUR processes serves as a first indicator whether STUR might be a relevant model. Fitting STUR models to the respective unemployment rates using Bayesian techniques we get a more sophisticated picture of relevance. However, after testing for STUR models as well as fitting them there are still ambiguous decisions left for certain countries. Thus, in the *third essay*, the parameters of competing models are estimated for each country. The fitted equations are then used to generate a large number of replications to which we apply the different test statistics again. That is, we simulate the empirical distribution functions of the test statistics to draw final conclusions. The results of estimations and simulations correspond to the test results concerning the rejection of the null hypothesis. For six out of ten countries STUR is the model of choice.

To gain a more differentiated idea of which process is relevant under the alternative we calculate conditional mean forecasts. Unfortunately, the random walk model is the winner of the forecast competition. The STUR model performs poorly on average, however, showing some strength for certain countries. As the stationary model performs bad as well, forecasting seems not to be adequate to discriminate between different alternatives of the unit-root tests considered. Unfortunately, forecast results do not correspond to test results. For most countries and different forecast horizons using forecast combinations does not result in better forecasts. Cumulative forecast errors show strong similarities among the rival models. We may conlude that STUR models even if found in the data are not good forecast models at all. Further evidence is provided in 1-step forecasts from simulated data.

To sum up, STUR is a relevant model for economics. Nonstationary economic data at the frontier between stationarity and difference-stationarity can be modeled in a more sophisticated way. To test whether STUR is relevant to model a certain economic time series, a new test is introduced having power against stationary as well as STUR processes competing successfully against different adequate tests. Estimation by Bayesian techniques is very time-consuming and thus calls for alternative procedures to make STUR models more attractive to applied economists. Results from testing and estimating including simulation experiments confirm the relevance of STUR models for unemployment rates. Except for forecasting, where STUR performs worse even if the data are generated by STUR, results of the present dissertation provide evidence for the relevance of STUR models in economics.

1 Testing for Unit Roots Using the AR-ARCH Structure of STUR Models

1.1 Introduction

Unit-root tests as presented by Dickey and Fuller (1979) and Said and Dickey (1984) test for the null hypothesis of a unit root versus the alternative of a stationary autoregressive process where the coefficients of the autoregressive terms are assumed to be constant over time. We refer to fixed-coefficient processes. Particularly, fixed unit-root processes like the random walk can be differenced to stationarity. Granger and Swanson (1997) introduce the class of so-called stochastic unit-root (STUR) processes which do not share this feature, i.e. they are nonstationary, though not difference-stationary. More general, STUR models have random coefficients allowing for a change between stationary and explosive regimes with the mean of the coefficient of the autoregressive term equal to one in case of a stochastic unit root. Furthermore, the coefficient process may be correlated over time. As the coefficient is drawn from a contin-

Not submitted.

uous distribution, the probability of the occurence of a unit root at a certain point in time is equal to zero. In presence of STUR alternatives, Granger and Swanson (1997) report a weak power performance of the Dickey-Fuller (DF) and augmented Dickey-Fuller (ADF) unit-root tests. So they apply the score test by Leybourne et al. (1996) of the null hypothesis of a fixed unit root versus the alternative of a stochastic unit root. The Leybourne-McCabe-Tremayne (LMT) test really has power against STUR processes, however it substantially drops down with decreasing coefficient variances. As the DF test, it can be augmented by lagged differences. We refer to the augmented LMT (ALMT) test. In addition, it assumes any kind of unit root to be present. Consequently, the LMT test is not expected to have power against stationary alternatives.

We suggest a new likelihood-ratio-type test of the null of a fixed unit root which has power against stationary alternatives as the DF test, against STUR alternatives as the LMT test and it also outperforms the latter in case of small coefficient variation. The setup derives from the STUR model nesting stationary as well as fixed and stochastic unit-root processes. In section 1.2, we introduce the STUR model by Granger and Swanson (1997) which may be represented by an autoregressive model of order one (AR(1)) with autoregressive conditional heteroskedastic errors of order one (ARCH(1)) as initiated by Tsay (1987). To test for the reduction to a random walk in AR(1)-ARCH(1) models, Klüppelberg et al. (2002) provide a theorem to obtain the asymptotic distribution of the deviance statistic. We discuss the present testing problem in Section 1.3, define the test statistic and derive its limiting distribution by verifying the assumptions of Lemma B.1 in Klüppelberg et al. (2002). To evaluate the characteristics of the test, we provide some Monte Carlo evidence in Section 1.4 where we calculate critical values and compare the size and power performance of the STUR test to the DF/ADF test and the LMT/ALMT test. In Section 1.5 we apply the test to unemployment rates and compare the results to DF and LMT. Finally, we may draw some conclusions.

1.2 STUR Model

The test is based on the stochastic unit-root (STUR) model introduced by Granger and Swanson (1997) where the coefficient mean is not restricted to one. There, a series x_t is generated by

$$x_t = a_t x_{t-1} + \varepsilon_t \tag{1.1}$$

where ε_t is independent identically distributed (i.i.d.) with mean zero and variance σ_ε^2 and a_t follows a random coefficient process

$$a_t = e^{\alpha_t} \tag{1.2}$$

where α_t is normally distributed with mean m, variance σ_α^2 and finite power spectrum $g_\alpha(\omega)$.

The coefficient process is allowed to be correlated over time according to the stationary AR(1) process with intercept

$$\alpha_t = \mu + \rho\alpha_{t-1} + \eta_t \tag{1.3}$$

where $|\rho| < 1$ and η_t is independent normally distributed with mean zero and variance σ_η^2 and thus, $m = \mu/(1-\rho)$ and $\sigma_\alpha^2 = \sigma_\eta^2/(1-\rho^2)$. Moreover, η_t is assumed to be independent of ε_t. Consequently, the random coefficient a_t is log-normally distributed with mean

$$E[a_t] = e^{m+\sigma_\alpha^2/2} \tag{1.4}$$

and variance

$$Var[a_t] = (e^{\sigma_\alpha^2} - 1)e^{2m+\sigma_\alpha^2}. \tag{1.5}$$

The random coefficient process a_t may also be considered as a random variable fluctuating around a constant mean. This is implemented by substituting the infinite-order Taylor-series approximation of a_t around $m + \sigma_\alpha^2/2$

$$a_t = E[a_t] + E[a_t]\sum_{i=1}^{\infty}\frac{(\alpha_t - \ln E[a_t])^i}{i!} \tag{1.6}$$

for a_t in equation (1.1), to obtain

$$x_t = \phi x_{t-1} + u_t \tag{1.7}$$

where $\phi = E[a_t]$ and $u_t = b_t x_{t-1} + \varepsilon_t$ with

$$b_t = E[a_t]\sum_{i=1}^{\infty}\frac{(\alpha_t - \ln E[a_t])^i}{i!}. \tag{1.8}$$

Approximating a_t by Taylor series should work reasonably well for moderate coefficient variances. In equation (1.7), the mean of x_t conditional on x_{t-1} is given by

$$\begin{aligned} E[x_t|x_{t-1}] = & \quad \phi x_{t-1} + E[u_t|x_{t-1}] \\ = & \quad \phi x_{t-1} + E[b_t|x_{t-1}]x_{t-1} + E[\varepsilon_t|x_{t-1}] \\ = & \quad \phi x_{t-1} + E[b_t]x_{t-1} + E[\varepsilon_t] \\ = & \quad \phi x_{t-1} \end{aligned} \tag{1.9}$$

where b_t does not depend on x_{t-1}, having unconditional mean $E[b_t] = 0$ to be checked from equations (1.6) and (1.8) for all $E[a_t] > 0$ and ε_t is i.i.d. with mean zero. The variance of x_t conditional on x_{t-1} is calculated from

$$\begin{aligned} Var[x_t|x_{t-1}] = & \quad Var[u_t|x_{t-1}] \\ = & \quad Var[b_t|x_{t-1}]x_{t-1}^2 + Var[\varepsilon_t|x_{t-1}] \\ = & \quad Var[b_t]x_{t-1}^2 + Var[\varepsilon_t] \\ = & \quad Var[a_t]x_{t-1}^2 + \sigma_\varepsilon^2 \end{aligned} \tag{1.10}$$

as b_t does not depend on x_{t-1}, ε_t is i.i.d. with variance σ_ε^2 and independent of η_t which is driving the coefficient process and from equations (1.6) and (1.8), we have $Var[b_t] = Var[a_t]$. In fact, the STUR process in equation (1.1) may be well represented in first and second conditional moment by the fixed-coefficient AR(1) model with errors generated by an ARCH(1) model in equation (1.7) within some neighborhood of the mean of a_t. As in Tsay (1987), studying more general random coefficient models with independent coefficients, we may derive moment equivalence between an AR-ARCH model and the random coefficient model with correlated coefficients in equation (1.1).

Several models are nested in the STUR model and therefore in the AR(1)-

ARCH(1) model as well. For $\sigma_\alpha^2 = 0$, the STUR model may reduce to a stationary AR(1) ($m < 0$) or a random walk ($m = 0$) model with a fixed unit root $a_t = 1$ for all t while for $\sigma_\alpha^2 > 0$ and $m = -\sigma_\alpha^2/2$, we obtain a so-called stochastic unit root with $E[a_t] = 1$. And one may also think of less explosive random coefficient processes with $\sigma_\alpha^2 > 0$ and $m < -\sigma_\alpha^2/2$. Correspondingly, from equations (1.6) and (1.8) we have for $\sigma_\alpha^2 = 0$ that $a_t = E[a_t]$ for all t and thus $b_t = 0$. That is, conditional heteroskedasticity vanishes in this case and we obtain a stationary or a difference-stationary AR(1) process according to the value of $\phi = E[a_t]$. Clearly, $E[a_t] = e^{m+\sigma_\alpha^2/2}$ in the STUR process is restricted to positive numbers whereas ϕ in the AR(1)-ARCH(1) process is not. As a consequence, oscillating data cannot be modeled by STUR.

1.3 AR-ARCH Test

We may now reformulate equation (1.7) by using the conditional variance of x_t in equation (1.10) to obtain the following AR(1)-ARCH(1) model with true parameter values indicated by subscript 0

$$x_t = \phi_0 x_{t-1} + \sigma_{t,0} e_t \tag{1.11}$$

for $t = 1, \ldots, n$ where $x_0 = 0$, e_t is i.i.d. with mean zero, variance one and finite eight moment, moments of e_t are denoted μ_k of order $k = 3, 4, 5, 6, 8$, respectively, $\sigma_{t,0} = \sqrt{\beta_0 + \lambda_0 x_{t-1}^2}$ with $\beta_0 = \sigma_\varepsilon^2$ and $\lambda_0 = Var[a_t]$. Note that the ARCH part is driven by previous observations as introduced in Weiss (1984) rather than by previous innovations. e_t denotes the value of $\hat{e}_t = (x_t - e^{\bar{\alpha}} x_{t-1})/\sigma_t$ when the parameters take their true values. Particularly, for the autoregressive coefficient we have

$$\phi_0 = E[a_t] = e^{\bar{\alpha}_0} \tag{1.12}$$

where $\bar{\alpha}_0 = m + \sigma_\alpha^2/2$ is defined as the value of α_0 associated with the mean of a_t. The vector of parameters $\theta = (\bar{\alpha}, \beta, \lambda)'$ assumes its values in the parameter space $\Theta = (-\infty, \infty) \times (0, \infty) \times [0, \infty)$. To test for a unit root, we consider the following null hypothesis

$$H_0 : \theta_0 \in \Theta_0 := \{0\} \times (0, \infty) \times \{0\} \tag{1.13}$$

i.e. $\bar{\alpha}_0 = 0$ ($\phi_0 = 1$), $\beta_0 > 0$, $\lambda_0 = 0$ where the process in equation (1.11) reduces to a mean-zero random walk. Under the alternative hypothesis

$$H_1 : \theta_0 \in \Theta_1 := (-\infty, \infty) \times (0, \infty) \times [0, \infty) \backslash \Theta_0 \tag{1.14}$$

i.e. $\bar{\alpha}_0 \in \mathbb{R}$ ($\phi_0 > 0$), $\beta_0 > 0$, $\lambda_0 \geq 0$, remaining processes nested in the STUR model as discussed in Section 1.2 may occur. Thus, a test of H_0 against H_1 is expected to have power against i.a. stationary AR(1) and STUR with stochastic unit root. We suggest a pseudo-likelihood ratio test using the deviance statistic

$$d_n := -2 \left[\mathcal{L}_n(\hat{\theta}_{n,0}) - \mathcal{L}_n(\hat{\theta}_{n,1}) \right] \tag{1.15}$$

where $\mathcal{L}_n(\hat{\theta}_{n,0})$ and $\mathcal{L}_n(\hat{\theta}_{n,1})$ denote the maximum values of the log-pseudo-likelihood functions conditional on x_0 under the null and under the alternative, respectively. Klüppelberg et al. (2002) prove the existence of maximizers $\hat{\theta}_{n,0}$ in Θ_0 and $\hat{\theta}_{n,1}$ in Θ_1 of the log-pseudo-likelihood functions with probability approaching one as $n \to \infty$. The resulting estimators are consistent for θ_0. Furthermore, in their Lemma B.1 they provide a rather general theorem to derive the limiting distribution of a certain deviance statistic which is illustrated by testing for the reduction of an AR(1)-ARCH(1) model to a random walk. There, Klüppelberg et al. (2002) assume with $H_1 : \theta_0 \in \Theta_{1'} := (-\infty, \infty) \times (0, \infty) \times [0, \infty)$ where $\theta_0 = (\phi_0, \beta_0, \lambda_0)'$ a different alternative hypothesis. The theorem may be reduced to Lemma 1 where $\mathcal{S}_n(\theta_0)$ and $\mathcal{F}_n(\theta_0)$ are the vector of the positive first and the matrix of the negative second derivatives evaluated at the true values of parameters, $M^{1/2}$ denotes the left Cholesky square root of a positive definite matrix M, $O_P(1)$ and $o_P(1)$ indicate random variables bounded in probability and converging to zero in probability, respectively, as $n \to \infty$. Klüppelberg et al. (2002) refer to the neighborhood of θ_0 defined by

$$N_n(A) = \{\theta : (\theta - \theta_0)' G_n (\theta - \theta_0) \leq A^2\} \tag{1.16}$$

where G_n is a positive definite scaling matrix for $n \geq 1$ and $A > 0$.

Lemma 1. *First, suppose there is a deterministic, diagonal, non-singular matrix G_n with minimum eigenvalue $\lambda_{min}(G_n) \to \infty$ as $n \to \infty$ such that*

$$G_n^{-1/2} \mathcal{F}_n^{1/2}(\theta_0) = Y_n + o_P(1) \tag{1.17}$$

where Y_n is a lower triangular matrix with positive diagonal elements and non-diagonal elements being zero in the first column. Second, suppose that

$$\left[G_n^{-1/2} \mathcal{S}_n(\theta_0), Y_n\right] \xrightarrow{D} (S, Y) \tag{1.18}$$

as $n \to \infty$ for some almost surely (a.s.) finite random vector S and a.s. finite, *nonsingular matrix Y. Particularly, joint convergence is required. And third, suppose that*

$$\sup_{\theta \in N_n(A)} \left| G_n^{-1/2} \left[\mathcal{F}_n(\theta) - \mathcal{F}_n(\theta_0) \right] G_n^{-1/2} \right| \xrightarrow{P} 0 \tag{1.19}$$

as $n \to \infty$ for each $A > 0$. Then, there exist pseudo-maximum-likelihood estimators $\hat{\theta}_{n,0}$ and $\hat{\theta}_{n,1}$ uniquely maximizing $\mathcal{L}_n(\theta)$ on $N_n(A) \cap \Theta_0$ and $N_n(A) \cap \Theta_1$, respectively, for each $A > 0$, on an event with probability approaching one as $n \to \infty$ and $A \to \infty$. The resulting estimators are consistent for θ_0. Furthermore, it holds that

$$\mathcal{F}_n^{-1/2}(\theta_0) \mathcal{S}_n(\theta_0) \xrightarrow{D} Y^{-1} S =: Z \tag{1.20}$$

for a finite random vector $Z = (Z^1, Z^2, Z^3)'$ and

$$d_n = -2 \left[\mathcal{L}_n(\hat{\theta}_{n,0}) - \mathcal{L}_n(\hat{\theta}_{n,1}) \right] \xrightarrow{D} (Z^1)^2 + (Z^3)^2 \mathbb{1}(Z^3 \geq 0). \tag{1.21}$$

as $n \to \infty$ where $\mathbb{1}$ denotes the indicator function.

Proof of Lemma 1. See Klüppelberg et al. (2002) for the proof and a discussion of the catalogue of assumptions. □

As a result, we may derive the limiting distribution of the test statistic in equation (1.15) applied to the model in equations (1.11) and (1.12) under the testing problem in (1.13) and (1.14) by verifying assumptions (1.17), (1.18), and (1.19). Subsequently, the matrices of derivatives are scaled by

$$G_n = diag(n^2, n, n^3) \tag{1.22}$$

to allow for convergence. To verify assumption (1.17), Lemma 2 will be useful:

Lemma 2. *We suppose that*

$$\frac{1}{n^{5/2}}\sum_{t=1}^{n} x_{t-1}^3 e_t \xrightarrow{P} 0, \tag{1.23}$$

$$\frac{1}{n^2}\sum_{t=1}^{n} x_{t-1}^2 (e_t^2 - 1) \xrightarrow{P} 0, \tag{1.24}$$

$$\frac{1}{n^3}\sum_{t=1}^{n} x_{t-1}^4 (e_t^2 - 1) \xrightarrow{P} 0 \tag{1.25}$$

as $n \to \infty$*. Then, the symmetric matrix* $G_n^{-1/2}\mathcal{F}_n(\theta_0)G_n^{-1/2}$ *has the asymptotic representation*

$$\begin{aligned} &G_n^{-1/2}\mathcal{F}_n(\theta_0)G_n^{-1/2} \\ &= \begin{bmatrix} \frac{1}{\beta_0 n^2}\sum_{t=1}^{n} x_{t-1}^2 & 0 & 0 \\ & \frac{1}{2\beta_0^2} & \frac{1}{2\beta_0^2 n^2}\sum_{t=1}^{n} x_{t-1}^2 \\ & & \frac{1}{2\beta_0^2 n^3}\sum_{t=1}^{n} x_{t-1}^4 \end{bmatrix} + o_P(1). \end{aligned} \tag{1.26}$$

Proof of Lemma 2. $\mathcal{S}_n(\theta) = [\mathcal{S}_n^i(\theta)]$ and $\mathcal{F}_n(\theta) = [\mathcal{F}_n^{ij}(\theta)]$ denote the vector of positive first and the matrix of negative second partial derivatives, respectively, of the log-likelihood function conditional on x_0 which is equal to

$$\mathcal{L}_n(\theta) = -\frac{1}{2}\sum_{t=1}^{n}\ln(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{n}\hat{e}_t^2 - \frac{1}{2}(n-1)\ln(2\pi) \tag{1.27}$$

where

$$\sigma_t^2 = \beta + \lambda x_{t-1}^2 \tag{1.28}$$

does not depend on $\bar{\alpha}$ and

$$\hat{e}_t = \frac{x_t - e^{\bar{\alpha}} x_{t-1}}{\sigma_t}. \tag{1.29}$$

In calculating the first and second derivatives of $\mathcal{L}_n(\theta)$ with respect to the parameters $\bar{\alpha}$, β, and λ, we use the following expressions:

$$\frac{\partial \sigma_t^2}{\partial \beta} = 1, \qquad \frac{\partial \sigma_t^2}{\partial \lambda} = x_{t-1}^2 \tag{1.30}$$

and

$$\frac{\partial \hat{e}_t^2}{\partial \bar{\alpha}} = -\frac{2e^{\bar{\alpha}} x_{t-1} \hat{e}_t}{\sigma_t}, \qquad \frac{\partial \hat{e}_t^2}{\partial \beta} = -\frac{\hat{e}_t^2}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \beta}, \qquad \frac{\partial \hat{e}_t^2}{\partial \lambda} = -\frac{\hat{e}_t^2}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \lambda} \tag{1.31}$$

resulting from equations (1.28) and (1.29), respectively. Then, we get for the first partial derivatives

$$\mathcal{S}_n^1 = \frac{\partial \mathcal{L}_n}{\partial \bar{\alpha}} = -\frac{1}{2} \sum_{t=1}^{n} \frac{\partial \hat{e}_t^2}{\partial \bar{\alpha}} = \sum_{t=1}^{n} \frac{e^{\bar{\alpha}} x_{t-1} \hat{e}_t}{\sigma_t}, \tag{1.32}$$

$$\mathcal{S}_n^2 = \frac{\partial \mathcal{L}_n}{\partial \beta} = -\frac{1}{2} \sum_{t=1}^{n} \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \beta} - \frac{1}{2} \sum_{t=1}^{n} \frac{\partial \hat{e}_t^2}{\partial \beta} = \frac{1}{2} \sum_{t=1}^{n} \frac{\hat{e}_t^2 - 1}{\sigma_t^2}, \tag{1.33}$$

$$\mathcal{S}_n^3 = \frac{\partial \mathcal{L}_n}{\partial \lambda} = -\frac{1}{2} \sum_{t=1}^{n} \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \lambda} - \frac{1}{2} \sum_{t=1}^{n} \frac{\partial \hat{e}_t^2}{\partial \lambda} = \frac{1}{2} \sum_{t=1}^{n} \frac{\hat{e}_t^2 - 1}{\sigma_t^2} x_{t-1}^2 \tag{1.34}$$

and thus for the negative second partial derivatives

$$\begin{aligned}\mathcal{F}_n^{11} = -\frac{\partial^2 \mathcal{L}_n}{\partial \bar{\alpha}^2} &= -\sum_{t=1}^{n} \frac{x_{t-1}}{\sigma_t}\left[e^{\bar{\alpha}}\hat{e}_t + e^{\bar{\alpha}}\frac{1}{2}\left(\hat{e}_t^2\right)^{-1/2}\frac{\partial\ \hat{e}_t^2}{\partial \bar{\alpha}}\right] \\ &= \sum_{t=1}^{n} \frac{e^{\bar{\alpha}} x_{t-1}}{\sigma_t}\left[\frac{e^{\bar{\alpha}} x_{t-1}}{\sigma_t} - \hat{e}_t\right], \end{aligned} \tag{1.35}$$

$$\begin{aligned}\mathcal{F}_n^{12} = \mathcal{F}_n^{21} = -\frac{\partial^2 \mathcal{L}_n}{\partial \bar{\alpha}\partial\beta} &= -\frac{1}{2}\sum_{t=1}^{n} \frac{1}{\sigma_t^2}\frac{\partial \hat{e}_t^2}{\partial \bar{\alpha}} \\ &= \sum_{t=1}^{n} \frac{e^{\bar{\alpha}} x_{t-1}\hat{e}_t}{\sigma_t^3}, \end{aligned} \tag{1.36}$$

$$\begin{aligned}\mathcal{F}_n^{13} = \mathcal{F}_n^{31} = -\frac{\partial^2 \mathcal{L}_n}{\partial \bar{\alpha}\partial\lambda} &= -\frac{1}{2}\sum_{t=1}^{n} \frac{x_{t-1}^2}{\sigma_t^2}\frac{\partial \hat{e}_t^2}{\partial \bar{\alpha}} \\ &= \sum_{t=1}^{n} \frac{e^{\bar{\alpha}} x_{t-1}^3\hat{e}_t}{\sigma_t^3}, \end{aligned} \tag{1.37}$$

$$\begin{aligned}\mathcal{F}_n^{22} = -\frac{\partial^2 \mathcal{L}_n}{\partial \beta^2} &= -\frac{1}{2}\sum_{t=1}^{n} \frac{\frac{\partial \hat{e}_t^2}{\partial \beta}\sigma_t^2 - \frac{\partial \sigma_t^2}{\partial \beta}(\hat{e}_t^2 - 1)}{\sigma_t^4} \\ &= \frac{1}{2}\sum_{t=1}^{n} \frac{2\hat{e}_t^2 - 1}{\sigma_t^4}, \end{aligned} \tag{1.38}$$

$$\begin{aligned}\mathcal{F}_n^{23} = \mathcal{F}_n^{32} = -\frac{\partial^2 \mathcal{L}_n}{\partial \beta\partial\lambda} &= -\frac{1}{2}\sum_{t=1}^{n} \frac{\frac{\partial \hat{e}_t^2}{\partial \lambda}\sigma_t^2 - \frac{\partial \sigma_t^2}{\partial \lambda}(\hat{e}_t^2 - 1)}{\sigma_t^4} \\ &= \frac{1}{2}\sum_{t=1}^{n} x_{t-1}^2\frac{2\hat{e}_t^2 - 1}{\sigma_t^4}, \end{aligned} \tag{1.39}$$

$$\mathcal{F}_n^{33} = -\frac{\partial^2 \mathcal{L}_n}{\partial \lambda^2} = -\frac{1}{2}\sum_{t=1}^{n} x_{t-1}^2 \frac{\frac{\partial \hat{e}_t^2}{\partial \lambda}\sigma_t^2 - \frac{\partial \sigma_t^2}{\partial \lambda}(\hat{e}_t^2 - 1)}{\sigma_t^4}$$
$$= \frac{1}{2}\sum_{t=1}^{n} x_{t-1}^4 \frac{2\hat{e}_t^2 - 1}{\sigma_t^4}. \tag{1.40}$$

By pre- and post-multiplying $\mathcal{F}_n(\theta)$ evaluated at θ_0 by G_n given in equation (1.22), we obtain the symmetric matrix

$$G_n^{-1/2}\mathcal{F}_n(\theta_0)G_n^{-1/2}$$
$$= \begin{bmatrix} \frac{\mathcal{F}_n^{11}(\theta_0)}{n^2} & \frac{\mathcal{F}_n^{12}(\theta_0)}{n^{3/2}} & \frac{\mathcal{F}_n^{13}(\theta_0)}{n^{5/2}} \\ & \frac{\mathcal{F}_n^{22}(\theta_0)}{n} & \frac{\mathcal{F}_n^{23}(\theta_0)}{n^2} \\ & & \frac{\mathcal{F}_n^{33}(\theta_0)}{n^3} \end{bmatrix} \tag{1.41}$$

from where we show equation (1.26) element by element. For $\theta_0 = (0, \beta_0, 0)'$, we have

$$\frac{\mathcal{F}_n^{11}(\theta_0)}{n^2} = \frac{1}{\beta_0 n^2}\sum_{t=1}^{n} x_{t-1}^2 - \frac{1}{\sqrt{\beta_0}n^2}\sum_{t=1}^{n} x_{t-1}e_t \tag{1.42}$$

where $-\beta_0^{-1/2}n^{-1}\sum_{t=1}^n x_{t-1}e_t$ weakly converges to $-[W^2(1) - 1]/2$ by the functional central limit theorem with $W(1)$ a standard Brownian motion with variance one (Phillips, 1987). The mean of $-\beta_0^{-1/2}n^{-2}\sum_{t=1}^n x_{t-1}e_t$ is equal to zero. As $W(1) \sim N(0,1)$ and hence $W^2(1) \sim \chi^2(1)$, the variance converges to zero as $n \to \infty$, i.e. $-\beta_0^{-1/2}n^{-2}\sum_{t=1}^n x_{t-1}e_t$ converges to zero in probability and we arrive at

$$\frac{\mathcal{F}_n^{11}(\theta_0)}{n^2} = \frac{1}{\beta_0 n^2}\sum_{t=1}^{n} x_{t-1}^2 + o_P(1). \tag{1.43}$$

By the same arguments, we derive

$$\frac{\mathcal{F}_n^{12}(\theta_0)}{n^{3/2}} = \frac{1}{\beta_0^{3/2} n^{3/2}} \sum_{t=1}^{n} x_{t-1} e_t \tag{1.44}$$

where $\beta_0^{-3/2} n^{-1} \sum_{t=1}^{n} x_{t-1} e_t$ weakly converges to $\beta_0^{-1}[W^2(1) - 1]/2$ by the functional central limit theorem. Again, the mean of $\beta_0^{-3/2} n^{-3/2} \sum_{t=1}^{n} x_{t-1} e_t$ is equal to zero and the variance converges zero as $n \to \infty$, i.e. $\beta_0^{-3/2} n^{-3/2} \sum_{t=1}^{n} x_{t-1} e_t$ converges to zero in probability resulting in

$$\frac{\mathcal{F}_n^{12}(\theta_0)}{n^{3/2}} = 0 + o_P(1). \tag{1.45}$$

Now, by assumption (1.23) we see that

$$\begin{aligned} \frac{\mathcal{F}_n^{13}(\theta_0)}{n^{5/2}} &= \frac{1}{\beta_0^{3/2} n^{5/2}} \sum_{t=1}^{n} x_{t-1}^3 e_t \\ &= 0 + o_P(1). \end{aligned} \tag{1.46}$$

Then, we deduce

$$\begin{aligned} \frac{\mathcal{F}_n^{22}(\theta_0)}{n} &= \frac{1}{\beta_0^2 n} \sum_{t=1}^{n} e_t^2 - \frac{1}{2\beta_0^2} \\ &= \frac{1}{2\beta_0^2} + \left(\frac{1}{\beta_0^2 n} \sum_{t=1}^{n} e_t^2 - \frac{1}{\beta_0^2} \right) \\ &= \frac{1}{2\beta_0^2} + o_P(1) \end{aligned} \tag{1.47}$$

as the term in parentheses converges to zero in probability by the weak law of large numbers. By using assumption (1.24), we obtain

$$\begin{aligned}
\frac{\mathcal{F}_n^{23}(\theta_0)}{n^2} &= \frac{1}{\beta_0^2 n^2}\sum_{t=1}^n x_{t-1}^2 e_t^2 - \frac{1}{2\beta_0^2 n^2}\sum_{t=1}^n x_{t-1}^2 \\
&= \frac{1}{2\beta_0^2 n^2}\sum_{t=1}^n x_{t-1}^2 + \frac{1}{\beta_0^2 n^2}\sum_{t=1}^n x_{t-1}^2 e_t^2 - \frac{1}{\beta_0^2 n^2}\sum_{t=1}^n x_{t-1}^2 \\
&= \frac{1}{2\beta_0^2 n^2}\sum_{t=1}^n x_{t-1}^2 + \frac{1}{\beta_0^2 n^2}\sum_{t=1}^n x_{t-1}^2 (e_t^2 - 1) \\
&= \frac{1}{2\beta_0^2 n^2}\sum_{t=1}^n x_{t-1}^2 + o_P(1). \qquad (1.48)
\end{aligned}$$

Finally, by assumption (1.25) we end up with

$$\begin{aligned}
\frac{\mathcal{F}_n^{33}(\theta_0)}{n^3} &= \frac{1}{\beta_0^2 n^3}\sum_{t=1}^n x_{t-1}^4 e_t^2 - \frac{1}{2\beta_0^2 n^3}\sum_{t=1}^n x_{t-1}^4 \\
&= \frac{1}{2\beta_0^2 n^3}\sum_{t=1}^n x_{t-1}^4 + \frac{1}{\beta_0^2 n^3}\sum_{t=1}^n x_{t-1}^4 e_t^2 - \frac{1}{\beta_0^2 n^3}\sum_{t=1}^n x_{t-1}^4 \\
&= \frac{1}{2\beta_0^2 n^3}\sum_{t=1}^n x_{t-1}^4 + \frac{1}{\beta_0^2 n^3}\sum_{t=1}^n x_{t-1}^4 (e_t^2 - 1) \\
&= \frac{1}{2\beta_0^2 n^3}\sum_{t=1}^n x_{t-1}^4 + o_P(1) \qquad (1.49)
\end{aligned}$$

and Lemma 2 is proved. □

We are now ready to state Theorem 1 and prove it by verifying assumptions (1.17), (1.18), and (1.19):

Theorem 1. *Suppose x_t satisfies the model outlined in (1.11) and (1.12). Then, under the null hypothesis (1.13) the following convergence result holds*

$$d_n \xrightarrow{D} \left(Z^1\right)^2 + \left(Z^3\right)^2 \mathbb{1}\left(Z^3 \geq 0\right) \qquad (1.50)$$

as $n \to \infty$ where

$$Z^1 = \frac{\frac{1}{2}\left[W_2^1(1)\right]^2 - \frac{1}{2}}{\sqrt{\int_0^1 \left[W_2^1(s)\right]^2 ds}} \tag{1.51}$$

and

$$\begin{aligned} Z^3 =& \frac{1}{\sqrt{2}\sqrt{\int_0^1 \left[W_2^1(s)\right]^4 ds - \left[\int_0^1 \left[W_2^1(s)\right]^2 ds\right]^2}} \\ &\times \left\{\mu_3 \int_0^1 \left[W_2^1(s)\right]^2 dW_2^1(s) + \sqrt{\mu_4 - \mu_3^2 - 1} \int_0^1 \left[W_2^1(s)\right]^2 dW_2^2(s) \right. \\ &\left. - \int_0^1 \left[W_2^1(s)\right]^2 ds \left[\mu_3 W_2^1(1) + \sqrt{\mu_4 - \mu_3^2 - 1} W_2^2(1)\right]\right\}. \end{aligned} \tag{1.52}$$

with $W_2(s) = [W_2^1(s), W_2^2(s)]'$ standard Brownian motion in two dimensions.

Note that the asymptotic distribution of d_n depends on the third and fourth moment of e_t, namely μ_3 and μ_4, which have to be estimated in advance. Furthermore, d_n in the model (1.11) and (1.12) converges under the same testing problem to the same distribution as the deviance statistic in $x_t = \phi_0 x_{t-1} + \sigma_{t,0} e_t$ where $\phi_0 \in \mathbb{R}$ as derived in Klüppelberg et al. (2002). At the same time, the specification $\phi_0 = e^{\bar{\alpha}_0}$ solves the problem where ϕ_0 is restricted to the interval $(0, \infty)$. Then there exists a cone C_{Θ_0} (C_{Θ_1}) with vertex at $\theta_0 \in \Theta_0$ ($\theta_0 \in \Theta_1$) coinciding with Θ_0 (Θ_1) on the neigborhood $N_n(A)$ of θ_0 which is required in the proof of Lemma B.1 in Klüppelberg et al. (2002).

Proof of Theorem 1. By showing assumptions (1.23), (1.24), and (1.25) of Lemma 2 to hold, we prepare the verification of assumption (1.17). So for equation (1.23), we have

$$\frac{1}{n^{5/2}}\sum_{t=1}^{n} x_{t-1}^3 e_t = \frac{1}{\sqrt{n}\sqrt{n}}\sum_{t=1}^{n}\frac{x_{t-1}^3 e_t}{n^{3/2}} \tag{1.53}$$

where $x_{t-1}^3 e_t$ for $t = 1, \ldots, n$ form a martingale difference sequence (m.d.s.). That is,

$$\begin{aligned} &E[x_{t-1}^3 e_t | x_{t-2}^3 e_{t-1}, \ldots, x_0^3 e_1] \\ &= E[x_{t-1}^3 | x_{t-2}^3 e_{t-1}, \ldots, x_0^3 e_1] E[e_t] \\ &= 0 \end{aligned} \tag{1.54}$$

and as $E[x_{t-1}^3 e_t] = 0$ and $E[e_t^2] = 1$, for the variance

$$\begin{aligned} Var[x_{t-1}^3 e_t] &= E[x_{t-1}^6 e_t^2] - E^2[x_{t-1}^3 e_t] \\ &= E[x_{t-1}^6]. \end{aligned} \tag{1.55}$$

Under the null hypothesis, x_{t-1} is a random walk, i.e. a cumulative sum of i.i.d. mean zero random variables $\sqrt{\beta_0} e_i$

$$\begin{aligned} E[x_{t-1}^6] &= E\left[\left(\sum_{i=1}^{t-1}\sqrt{\beta_0} e_i\right)^6\right] \\ &= \beta_0^3 E\left[\sum_{i\neq j\neq k} e_i^2 e_j^2 e_k^2\right] + \beta_0^3 E\left[\sum_{i\neq j} e_i^2 e_j^4\right] \\ &\quad + \beta_0^3 E\left[\sum_{i\neq j} e_i^3 e_j^3\right] + \beta_0^3 E\left[\sum_{i} e_i^6\right] \\ &= \beta_0^3 \sum_{i\neq j\neq k} E[e_i^2]E[e_j^2]E[e_k^2] + \beta_0^3 \sum_{i\neq j} E[e_i^2]E[e_j^4] \\ &\quad + \beta_0^3 \sum_{i\neq j} E[e_i^3]E[e_j^3] + \beta_0^3 \sum_{i} E[e_i^6] \end{aligned} \tag{1.56}$$

and by the multinomial theorem we obtain

$$\begin{aligned}
E[x_{t-1}^6] = \beta_0^3 &\binom{6}{2,2,2,0,\dots,0}\binom{t-1}{3}\cdot 1 \\
&+ \beta_0^3\binom{6}{2,4,0,\dots,0}\binom{t-1}{1}\binom{t-2}{1}\mu_4 \\
&+ \beta_0^3\binom{6}{3,3,0,\dots,0}\binom{t-1}{2}\mu_3^2 \\
&+ \beta_0^3\binom{6}{6,0,\dots,0}\binom{t-1}{1}\mu_6 \\
&= 15\beta_0^3(t-1)(t-2)(t-3) + 15\beta_0^3(t-1)(t-2)\mu_4 \\
&+ 10\beta_0^3(t-1)(t-2)\mu_3^2 + \beta_0^3(t-1)\mu_6. \qquad (1.57)
\end{aligned}$$

Thus, the variance of $x_{t-1}^3 e_t$ is growing at the rate t^3. As a consequence, scaling by $n^{3/2}$ may result in a finite variance. We apply the functional central limit theorem for martingale difference sequences (Billingsley, 1968, 206), to obtain

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_{t-1}^3 e_t}{n^{3/2}} \xrightarrow{D} N\left(0, Var\left[\frac{x_{t-1}^3 e_t}{n^{3/2}}\right]\right) \qquad (1.58)$$

and thus

$$\frac{1}{\sqrt{n}\sqrt{n}}\sum_{t=1}^{n}\frac{x_{t-1}^3 e_t}{n^{3/2}} \xrightarrow{P} 0. \qquad (1.59)$$

Next, we verify assumption (1.24)

$$\frac{1}{n^2}\sum_{t=1}^{n} x_{t-1}^2(e_t^2-1) = \frac{1}{\sqrt{n}\sqrt{n}}\sum_{t=1}^{n}\frac{x_{t-1}^2(e_t^2-1)}{n} \qquad (1.60)$$

where now $x_{t-1}^2(e_t^2-1)$ for $t=1,\dots,n$ is a m.d.s. as

$$\begin{aligned}
&E[x_{t-1}^2(e_t^2-1)|x_{t-2}^2(e_{t-1}^2-1),\dots,x_0^2(e_1^2-1)]\\
&=E[x_{t-1}^2|x_{t-2}^2(e_{t-1}^2-1),\dots,x_0^2(e_1^2-1)]E[e_t^2-1]\\
&=0 \qquad (1.61)
\end{aligned}$$

and as $E[x_{t-1}^2(e_t^2-1)]=0$ and $E[e_t^2]=1$, for the variance

$$\begin{aligned}
Var[x_{t-1}^2(e_t^2-1)] &= E[x_{t-1}^4(e_t^2-1)^2]-E^2[x_{t-1}^2(e_t^2-1)]\\
&=E[x_{t-1}^4]E[(e_t^2-1)^2]\\
&=E[x_{t-1}^4]E[e_t^4-2e_t^2+1]\\
&=(\mu_4-1)E[x_{t-1}^4]. \qquad (1.62)
\end{aligned}$$

Again, under the null hypothesis we have a cumulative sum of i.i.d. mean zero random variables $\sqrt{\beta_0}e_i$

$$\begin{aligned}
E[x_{t-1}^4] &= E\left[\left(\sum_{i=1}^{t-1}\beta_0^{1/2}e_i\right)^4\right]\\
&=\beta_0^2E\left[\sum_{i\neq j}e_i^2e_j^2\right]+\beta_0^2E\left[\sum_i e_i^4\right]\\
&=\beta_0^2\sum_{i\neq j}E[e_i^2]E[e_j^2]+\beta_0^2\sum_i E[e_i^4] \qquad (1.63)
\end{aligned}$$

where we obtain by the multinomial theorem

$$\begin{aligned}
E[x_{t-1}^4] &= \beta_0^2\binom{4}{2,2,0,\dots,0}\binom{t-1}{2}1+\beta_0^2\binom{4}{4,0\dots,0}\binom{t-1}{1}\mu_4\\
&=3\beta_0^2(t-1)(t-2)+\beta_0^2(t-1)\mu_4. \qquad (1.64)
\end{aligned}$$

That is, the variance of $x_{t-1}^2(e_t^2-1)$ is growing at the rate t^2 and we may scale by n. Applying the functional central limit theorem for martingale difference sequences results in

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_{t-1}^2(e_t^2-1)}{n}\xrightarrow{D}N\left(0,Var\left[\frac{x_{t-1}^2(e_t^2-1)}{n}\right]\right) \tag{1.65}$$

and therefore,

$$\frac{1}{\sqrt{n}\sqrt{n}}\sum_{t=1}^{n}\frac{x_{t-1}^2(e_t^2-1)}{n}\xrightarrow{P}0. \tag{1.66}$$

Finally, we show that assumption (1.25) is satisfied

$$\frac{1}{n^3}\sum_{t=1}^{n}x_{t-1}^4(e_t^2-1)=\frac{1}{\sqrt{n}\sqrt{n}}\sum_{t=1}^{n}\frac{x_{t-1}^4(e_t^2-1)}{n^2} \tag{1.67}$$

where $x_{t-1}^4(e_t^2-1)$ for $t=1,\ldots,n$ is a m.d.s. as

$$\begin{aligned}
&E[x_{t-1}^4(e_t^2-1)|x_{t-2}^4(e_{t-1}^2-1),\ldots,x_0^4(e_1^2-1)]\\
&=E[x_{t-1}^4|x_{t-2}^4(e_{t-1}^2-1),\ldots,x_0^4(e_1^2-1)]E[e_t^2-1]\\
&=0
\end{aligned} \tag{1.68}$$

and as $E[x_{t-1}^4(e_t^2-1)]=0$ and $E[e_t^2]=1$, for the variance

$$\begin{aligned}
Var[x_{t-1}^4(e_t^2-1)]&=E[x_{t-1}^8(e_t^2-1)^2]-E^2[x_{t-1}^4(e_t^2-1)]\\
&=E[x_{t-1}^8]E[(e_t^2-1)^2]\\
&=E[x_{t-1}^8]E[e_t^4-2e_t^2+1]\\
&=(\mu_4-1)E[x_{t-1}^8].
\end{aligned} \tag{1.69}$$

For the mean of x_{t-1}^8 under the null, we derive

$$
\begin{aligned}
E[x_{t-1}^8] &= E\left[\left(\sum_{i=1}^{t-1}\sqrt{\beta_0}e_i\right)^8\right] \\
&= \beta_0^4 E\left[\sum_{i\neq j\neq k\neq l} e_i^2 e_j^2 e_k^2 e_l^2\right] + \beta_0^4 E\left[\sum_{i\neq j\neq k} e_i^2 e_j^2 e_k^4\right] \\
&\quad + \beta_0^4 E\left[\sum_{i\neq j\neq k} e_i^2 e_j^3 e_k^3\right] + \beta_0^4 E\left[\sum_{i\neq j} e_i^2 e_j^6\right] \\
&\quad + \beta_0^4 E\left[\sum_{i\neq j} e_i^3 e_j^5\right] + \beta_0^4 E\left[\sum_{i\neq j} e_i^4 e_j^4\right] \\
&\quad + \beta_0^4 E\left[\sum_{i} e_i^8\right] \\
&= \beta_0^4 \sum_{i\neq j\neq k\neq l} E[e_i^2]E[e_j^2]E[e_k^2]E[e_l^2] + \beta_0^4 \sum_{i\neq j\neq k} E[e_i^2]E[e_j^2]E[e_k^4] \\
&\quad + \beta_0^4 \sum_{i\neq j\neq k} E[e_i^2]E[e_j^3]E[e_k^3] + \beta_0^4 \sum_{i\neq j} E[e_i^2]E[e_j^6] \\
&\quad + \beta_0^4 \sum_{i\neq j} E[e_i^3]E[e_j^5] + \beta_0^4 \sum_{i\neq j} E[e_i^4]E[e_j^4] \\
&\quad + \beta_0^4 \sum_{i} E[e_i^8] \qquad (1.70)
\end{aligned}
$$

and we obtain by the multinomial theorem

$$\begin{aligned}
E[x_{t-1}^8] = \beta_0^4 &\binom{8}{2,2,2,2,0,\ldots,0}\binom{t-1}{4}\cdot 1 \\
&+ \beta_0^4\binom{8}{2,2,4,0\ldots,0}\binom{t-1}{2}\binom{t-3}{1}\mu_4 \\
&+ \beta_0^4\binom{8}{2,3,3,0\ldots,0}\binom{t-1}{2}\binom{t-3}{1}\mu_3^2 \\
&+ \beta_0^4\binom{8}{2,6,0\ldots,0}\binom{t-1}{1}\binom{t-2}{1}\mu_6 \\
&+ \beta_0^4\binom{8}{3,5,0\ldots,0}\binom{t-1}{1}\binom{t-2}{1}\mu_3\mu_5 \\
&+ \beta_0^4\binom{8}{4,4,0\ldots,0}\binom{t-1}{2}\mu_4^2 \\
&+ \beta_0^4\binom{8}{8,0\ldots,0}\binom{t-1}{1}\mu_8 \\
= 105&\beta_0^4(t-1)(t-2)(t-3)(t-4)\cdot 1 \\
&+ 210\beta_0^4(t-1)(t-2)(t-3)\mu_4 \\
&+ 280\beta_0^4(t-1)(t-2)(t-3)\mu_3^2 \\
&+ 28\beta_0^4(t-1)(t-2)\mu_6 \\
&+ 56\beta_0^4(t-1)(t-2)\mu_3\mu_5 \\
&+ 35\beta_0^4(t-1)(t-2)\mu_4^2 \\
&+ (t-1)\beta_0^4\mu_8. \qquad (1.71)
\end{aligned}$$

Consequently, the variance of $x_{t-1}^4(e_t^2-1)$ is growing at the rate t^4 and we may scale by n^2 to make the functional central limit theorem for martingale difference sequences applicable, yielding

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_{t-1}^4(e_t^2-1)}{n^2} \xrightarrow{D} N\left(0, Var\left[\frac{x_{t-1}^4(e_t^2-1)}{n^2}\right]\right) \qquad (1.72)$$

and thus

$$\frac{1}{\sqrt{n}\sqrt{n}}\sum_{t=1}^{n}\frac{x_{t-1}^4(e_t^2-1)}{n^2}\xrightarrow{P}0. \tag{1.73}$$

That is, by Lemma 2 the following representation holds

$$\begin{aligned} & G_n^{-1/2}\mathcal{F}_n(\theta_0)G_n^{-1/2} \\ &= \begin{bmatrix} \frac{1}{\beta_0 n^2}\sum_{t=1}^n x_{t-1}^2 & 0 & 0 \\ & \frac{1}{2\beta_0^2} & \frac{1}{2\beta_0^2 n^2}\sum_{t=1}^n x_{t-1}^2 \\ & & \frac{1}{2\beta_0^2 n^3}\sum_{t=1}^n x_{t-1}^4 \end{bmatrix} + o_P(1) \end{aligned} \tag{1.74}$$

where we may take the Cholesky square root to obtain

$$G_n^{-1/2}\mathcal{F}_n^{1/2}(\theta_0) = Y_n + o_P(1) \tag{1.75}$$

with $Y_n = [Y_n^{ij}]$ having the only non-zero elements

$$Y_n^{11} = \sqrt{\frac{1}{\beta_0 n^2}\sum_{t=1}^n x_{t-1}^2}, \tag{1.76}$$

$$Y_n^{22} = \frac{1}{\sqrt{2}\beta_0}, \tag{1.77}$$

$$Y_n^{32} = \frac{1}{\sqrt{2}\beta_0 n^2}\sum_{t=1}^n x_{t-1}^2, \tag{1.78}$$

$$Y_n^{33} = \sqrt{\frac{1}{2\beta_0^2 n^3}\sum_{t=1}^n x_{t-1}^4 - \left(\frac{1}{\sqrt{2}\beta_0 n^2}\sum_{t=1}^n x_{t-1}^2\right)^2}. \tag{1.79}$$

Obviously, Y_n is a lower triangular matrix with positive diagonal elements and non-diagonal elements equal to zero in the first column, hence, assumption (1.17) is verified. From equations (1.76) to (1.79) and

$$G_n^{-1/2}\mathcal{S}_n(\theta_0) = \begin{bmatrix} \frac{1}{\sqrt{\beta_0}n}\sum_{t=1}^n x_{t-1}e_t \\ \frac{1}{2\beta_0\sqrt{n}}\sum_{t=1}^n (e_t^2-1) \\ \frac{1}{2\beta_0 n^{3/2}}\sum_{t=1}^n (e_t^2-1)x_{t-1}^2 \end{bmatrix} \tag{1.80}$$

we see that we have to find the joint limiting distribution of

$$\Bigg(\frac{1}{n^2}\sum_{t=1}^n x_{t-1}^2, \frac{1}{n^3}\sum_{t=1}^n x_{t-1}^4, \\ \frac{1}{n}\sum_{t=1}^n x_{t-1}e_t, \frac{1}{\sqrt{n}}\sum_{t=1}^n (e_t^2-1), \frac{1}{n^{3/2}}\sum_{t=1}^n (e_t^2-1)x_{t-1}^2\Bigg) \tag{1.81}$$

implying the weak convergence of $[G_n^{-1/2}\mathcal{S}_n(\theta_0), Y_n]$ as required in assumption (1.18). Alternatively, we may write equation (1.80) as

$$G_n^{-1/2}\mathcal{S}_n(\theta_0) = \sum_{t=1}^n A_{tn}B_{tn} \tag{1.82}$$

where

$$A_{tn} = \left[A_{tn}^{ij}\right] = \begin{bmatrix} \frac{x_{t-1}}{\sqrt{\beta_0}\sqrt{n}} & 0 \\ 0 & \frac{1}{2\beta_0} \\ 0 & \frac{x_{t-1}^2}{2\beta_0 n} \end{bmatrix}, \qquad B_{tn} = \begin{bmatrix} \frac{e_t}{\sqrt{n}} \\ \frac{e_t^2-1}{\sqrt{n}} \end{bmatrix}. \tag{1.83}$$

We define the continuous-time processes

$$A_n(s) = A_{\lfloor ns\rfloor,n} \qquad B_n(s) = \sum_{t=1}^{\lfloor ns\rfloor} B_{tn} \tag{1.84}$$

where $\lfloor\cdot\rfloor$ denotes the integer part and $s \in [1/n, 1]$. Consequently, equation (1.82) may be represented by

$$
\begin{aligned}
G_n^{-1/2}\mathcal{S}_n(\theta_0) &= \int_{1/n}^{1+1/n} A_n(s)dB_n(s) \\
&= \int_0^1 A_n\left(s+\frac{1}{n}\right)dB_n\left(s+\frac{1}{n}\right) \qquad (1.85)
\end{aligned}
$$

while the respective elements of Y_n can be written as

$$
\begin{aligned}
\frac{1}{\beta_0 n^2}\sum_{t=1}^{n} x_{t-1}^2 &= \int_{1/n}^{1+1/n} \left[A_n^{11}(s)\right]^2 ds \\
&= \int_0^1 \left[A_n^{11}\left(s+\frac{1}{n}\right)\right]^2 d\left(s+\frac{1}{n}\right) \\
&= \int_0^1 \left[A_n^{11}\left(s+\frac{1}{n}\right)\right]^2 ds, \qquad (1.86)
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{\beta_0^2 n^3}\sum_{t=1}^{n} x_{t-1}^4 &= \int_{1/n}^{1+1/n} \left[A_n^{11}(s)\right]^4 ds \\
&= \int_0^1 \left[A_n^{11}\left(s+\frac{1}{n}\right)\right]^4 d\left(s+\frac{1}{n}\right) \\
&= \int_0^1 \left[A_n^{11}\left(s+\frac{1}{n}\right)\right]^4 ds \qquad (1.87)
\end{aligned}
$$

and thus joint convergence in (1.81) is implied by joint convergence of the elements in equation (1.80) which results from the application of Theorem 2.2 and Remark 2.3 in Kurtz and Protter (1991) adapted to this case as summarized in Lemma 3. Note that joint convergence of $(A_n(s+1/n), B_n(s+1/n))$ implies joint convergence of $(A_n(s), B_n(s))$ and so we may skip $1/n$ in asymptotics.

Lemma 3. *For each n, let (A_n, B_n) be an $\{\mathcal{H}_s^n\}$-adapted process with sample paths in $D_{\mathbb{M}^{3\times 2}\times\mathbb{R}^2}[0,1]$ where $\mathbb{M}^{3\times 2}$ denotes the set of real-valued 3×2 matrices and let B_n be an $\{\mathcal{H}_s^n\}$-semimartingale. Fix $\delta \in (0,\infty]$ and define*

$B_n^\delta = B_n - J_\delta(B_n)$ where

$$J_\delta(B_n)(s) = \sum_{r \leq s} \left(\frac{|B_n(r) - B_n(r-)| - \delta}{|B_n(r) - B_n(r-)|} \right)^+ [B_n(r) - B_n(r-)]. \qquad (1.88)$$

with $B_n(r-) = \lim_{q \uparrow r} B_n(q)$ the left-hand limit at r. Let $B_n^\delta = M_n^\delta + N_n^\delta$ be a decomposition of B_n^δ with M_n^δ an $\{\mathcal{H}_s^n\}$-local martingale and N_n^δ a process with finite variation. First, suppose that $T_s(N_n^\delta)$ is stochastically bounded for each $s > 0$ where $T_s(\cdot)$ denotes total variation and second, suppose that for each $c > 0$, $sup_n E[(M_n^\delta(s \wedge \tau_n^c))^2 + T_{s \wedge \tau_n^c}(A_n^\delta)] < \infty$ where $\{\tau_n^c\}$ is a sequence of stopping times and $s \wedge \tau_n^c = \min(s, \tau_n^c)$. If $(A_n, B_n) \xrightarrow{D} (A, B)$ in the Skorohod topology on $D_{\mathbb{M}^{3\times2}\times\mathbb{R}^2}[0,1]$, then $(A_n, B_n, \int_0^1 A_n dB_n) \xrightarrow{D} (A, B, \int_0^1 AdB)$ in the Skorohod topology on $D_{\mathbb{M}^{3\times2}\times\mathbb{R}^2\times\mathbb{R}^3}[0,1]$.

Proof of Lemma 3. See Kurtz and Protter (1991) for the proof and a discussion of alternative assumptions. □

As $|B_n(r) - B(r-)|$ is finite for all $r \in [1/n, 1]$, we choose $\delta = \infty$ to obtain $J_\delta(B_n)(s) = 0$ for all $s \in [1/n, 1]$. Thus, $B_n^\delta = B_n$ which is a right-continuous martingale with left-hand limits, i.e. càdlàg martingale and hence also an $\{H_s^n\}$-local martingale (see e.g. Protter, 2005, 37) and we have $B_n^\delta = M_n^\delta$ and $N_n^\delta = 0$. Clearly, $T_s(0)$ is stochastically bounded for all $s \in [1/n, 1]$. Finally, as

$$\begin{aligned} E\left[\left(M_n^\delta(s)\right)^2\right] &= E\left[B_n^2(s)\right] \\ &= \sum_{t=1}^{\lfloor ns \rfloor} \begin{bmatrix} \frac{1}{n} \\ \frac{\mu_4 - 1}{n} \end{bmatrix} < \infty \end{aligned} \qquad (1.89)$$

for all $s \in [1/n, 1]$ and all n we have

$$\sup_n E\left[\left(M_n^\delta(s \wedge \tau_n^c)\right)^2 + T_{s\wedge\tau_n^c}(N_n^\delta)\right] < \infty \tag{1.90}$$

for all $c > 0$ and Lemma 3 applies, if $(A_n, B_n) \xrightarrow{D} (A, B)$ in the Skorohod topology on $D_{\mathbb{M}^{3\times 2}\times\mathbb{R}^2}[0,1]$. As $x_{t-1} = \sqrt{\beta_0}\sum_{k=1}^{t-1} e_k$, joint convergence of A_n and B_n is implied by joint convergence of the elements of B_n. That is, we ask for the joint distribution of

$$\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor} e_t, \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor}(e_t^2 - 1)\right) \tag{1.91}$$

in $D^2[0,1]$. We define the linear combination of the summands in (1.91)

$$y_t = u_1 e_t + u_2(e_t^2 - 1) \tag{1.92}$$

where $u = (u_1, u_2)' \in \mathbb{R}^2$ and $u \neq 0$. As y_t is an ergodic, stationary sequence of square-integrable martingale differences, by Theorem 7.5 in Durrett (1991, 375) we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor} y_t^2 = u_1\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor} e_t + u_2\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor}(e_t^2 - 1) \xrightarrow{D} \sigma_y W(s) \tag{1.93}$$

where

$$
\begin{aligned}
\sigma_y^2 &= \frac{1}{n}\sum_{t=1}^{n} E\left[y_t^2|\mathcal{H}_{t-1}\right] \\
&= \frac{1}{n}\sum_{t=1}^{n} E\left[u_1^2 e_t^2 + 2u_1u_2e_t(e_t^2-1) + u_2^2(e_t^2-1)^2\right] \\
&= \frac{1}{n}\sum_{t=1}^{n} \left[u_1^2 + 2u_1u_2\mu_3 + u_2^2(\mu_4-1)\right] \\
&= u_1^2 + 2u_1u_2\mu_3 + u_2^2(\mu_4-1) \\
&= u'M_2u
\end{aligned}
\tag{1.94}
$$

with

$$
M_2 = \begin{bmatrix} 1 & \mu_3 \\ \mu_3 & \mu_4 - 1 \end{bmatrix}
\tag{1.95}
$$

and $W(s)$ standard Brownian motion. That is, an arbitrary linear combination of the elements in (1.91) is normally distributed with mean zero and variance $\sigma_y^2 s$ and therefore the vector in (1.91) is multivariate normally distributed with mean vector zero and variance matrix $M_2 s$ or alternatively, $B_n(s)$ has the asymptotic distribution of $M_2^{1/2}W_2(s)$ where $W_2(s) = [W_2^1(s), W_2^2(s)]'$ is standard Brownian motion in two dimensions. Consequently, as $x_{t-1} = \sqrt{\beta_0}\sum_{k=1}^{\lfloor ns\rfloor-1} e_k$ for $t/n \le s < (t+1)/n$

$$
A_n^{11}(s) = \frac{x_{t-1}}{\sqrt{\beta_0}\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{k=1}^{\lfloor ns\rfloor-1} e_k \xrightarrow{D} W_2^1(s)
\tag{1.96}
$$

and then by the continuous mapping theorem

$$A_n^{32}(s) = \frac{x_{t-1}^2}{2\beta_0 n} = \frac{1}{2}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{\lfloor ns\rfloor - 1} e_k\right)^2 \xrightarrow{D} \frac{1}{2}\left[W_2^1(s)\right]^2. \tag{1.97}$$

Indeed, $(A_n, B_n) \xrightarrow{D} (A, B)$ and we may apply Lemma 3 to obtain the joint limiting distribution of $(A_n, B_n, \int_0^1 A_n dB_n)$ where $G_n^{-1/2} S_n(\theta_0)$ converges in distribution to

$$\begin{aligned}
S &= \int_0^1 A(s)dB(s) = \int_0^1 A(s)M_2^{1/2} dW_2(s) \\
&= \begin{bmatrix} \int_0^1 W_2^1(s)dW_2^1(s) \\ \frac{\mu_3}{2\beta_0}\int_0^1 dW_2^1(s) + \frac{\sqrt{\mu_4-\mu_3^2-1}}{2\beta_0}\int_0^1 dW_2^2(s) \\ \frac{\mu_3}{2}\int_0^1 \left[W_2^1(s)\right]^2 dW_2^1(s) + \frac{\sqrt{\mu_4-\mu_3^2-1}}{2}\int_0^1 \left[W_2^1(s)\right]^2 dW_2^2(s) \end{bmatrix} \\
&= \begin{bmatrix} \int_0^1 W_2^1(s)dW_2^1(s) \\ \frac{\mu_3}{2\beta_0} W_2^1(1) + \frac{\sqrt{\mu_4-\mu_3^2-1}}{2\beta_0} W_2^2(1) \\ \frac{\mu_3}{2}\int_0^1 \left[W_2^1(s)\right]^2 dW_2^1(s) + \frac{\sqrt{\mu_4-\mu_3^2-1}}{2}\int_0^1 \left[W_2^1(s)\right]^2 dW_2^2(s) \end{bmatrix}.
\end{aligned} \tag{1.98}$$

as $n \to \infty$. Finally, from equations (1.76), (1.78), (1.79), (1.86), (1.87), and (1.96) we see that Y_n^{11}, Y_n^{32}, and Y_n^{33} converge in distribution to

$$Y^{11} = \sqrt{\int_0^1 \left[W_2^1(s)\right]^2 ds} \tag{1.99}$$

$$Y^{32} = \frac{1}{\sqrt{2}}\int_0^1 \left[W_2^1(s)\right]^2 ds \tag{1.100}$$

$$Y^{33} = \sqrt{\frac{1}{2}\int_0^1 \left[W_2^1(s)\right]^4 ds - \frac{1}{2}\left[\int_0^1 \left[W_2^1(s)\right]^2 ds\right]^2}, \tag{1.101}$$

respectively and hence, Y_n weakly converges to

$$Y = \begin{bmatrix} \sqrt{J_2} & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}\beta_0} & 0 \\ 0 & \frac{1}{\sqrt{2}} J_2 & \sqrt{\frac{1}{2}\left(J_4 - J_2^2\right)} \end{bmatrix} \tag{1.102}$$

where $J_k = \int_0^1 \left[W_2^1(s)\right]^k ds$. We have shown that $[G_n^{-1/2}\mathcal{S}_n(\theta_0), Y_n]$ jointly converge to (S, Y) as $n \to \infty$ where S and Y are given above and therefore, assumption (1.18) of Lemma 1 is verified. Finally, to show that assumption (1.19) is satisfied we require several partial results concerning convergence and stochastic boundedness which are derived in advance. In assumption (1.19), the supremum is taken over $\theta \in N_n(A)$ where from equation (1.16) for the neigborhood

$$N_n(A) = \{\theta : \bar{\alpha}^2 n^2 + (\beta - \beta_0)^2 n + \lambda^2 n^3 \leq A^2\} \tag{1.103}$$

with $A > 0$. Hence, $\bar{\alpha}^2 n^2 \leq A^2$, $(\beta - \beta_0)^2 n \leq A^2$, and $\lambda^2 n^3 \leq A^2$ implying $|\bar{\alpha}| \leq A/n$, $|\beta - \beta_0| \leq A/\sqrt{n}$, and $\lambda \leq A/n^{3/2}$, respectively. Translated into Landau notation, we have $\bar{\alpha} = O_P(1/n)$, $\beta - \beta_0 = O_P(1/\sqrt{n})$, and $\lambda = O_P(1/n^{3/2})$. We choose n large enough such that $A/\sqrt{n} \leq \beta_0/2$. Then, $|\beta - \beta_0| \leq A/\sqrt{n}$ implying $\beta - \beta_0 \geq -A/\sqrt{n}$ results in $\beta \geq \beta_0/2$. As $\sigma_t^2 = \beta + \lambda x_{t-1}^2 \geq \beta$, we obtain $\sigma_t^2 \geq \beta_0/2$. The variances of x_{t-1}, x_{t-1}^2, x_{t-1}^3, and x_{t-1}^4 are growing at the rates t, t^2, t^3, and t^4, respectively. Consequently, x_{t-1}, x_{t-1}^2, x_{t-1}^3, and x_{t-1}^4 are bounded in probability with $\max_{1 \leq t \leq n} |x_{t-1}| = O_P(\sqrt{n})$, $\max_{1 \leq t \leq n} x_{t-1}^2 = O_P(n)$, $\max_{1 \leq t \leq n} |x_{t-1}^3| = O_P(n^{3/2})$, and $\max_{1 \leq t \leq n} x_{t-1}^4 =$

$O_P(n^2)$. Furthermore,

$$\begin{aligned}
\sigma_t^2 - \beta_0 &= \beta - \beta_0 + \lambda x_{t-1}^2 \\
&= O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{n^{3/2}}\right) O_P(n) \\
&= O_P\left(\frac{1}{\sqrt{n}}\right) \\
&= o_P(1) \qquad (1.104)
\end{aligned}$$

implying $\sigma_t^4 - \beta_0^2 = o_P(1)$ and $\sigma_t^6 - \beta_0^3 = o_P(1)$ by the rules of limit calculation and we also use

$$\begin{aligned}
\beta - e^{2\bar{\alpha}}\beta_0 &= \beta_0 - e^{2\bar{\alpha}}\beta_0 + \beta - \beta_0 \\
&= \left(1 - e^{2\bar{\alpha}}\right)\beta_0 + (\beta - \beta_0) \\
&= O_P(1 - C^{1/n}) + O_P(1/\sqrt{n}) \\
&= o_P(1) \qquad (1.105)
\end{aligned}$$

where $C > 0$ does not depend on n and $\bar{\alpha} = O_P(1/n)$ to derive

$$\begin{aligned}
\sigma_t^2 - e^{2\bar{\alpha}}\beta_0 &= \beta - e^{2\bar{\alpha}}\beta_0 + \lambda x_{t-1}^2 \\
&= o_P(1) + O_P\left(\frac{1}{n^{3/2}}\right) O_P(n) \\
&= o_P(1) + O_P\left(\frac{1}{\sqrt{n}}\right) \\
&= o_P(1). \qquad (1.106)
\end{aligned}$$

By the same reasoning, we obtain $\beta - e^{\bar{\alpha}}\beta_0 = o_P(1)$, $\beta - e^{\bar{\alpha}/2}\beta_0 = o_P(1)$, $\beta - e^{\bar{\alpha}/3}\beta_0 = o_P(1)$, and $\beta - e^{2\bar{\alpha}/3}\beta_0 = o_P(1)$ and then $\sigma_t^2 - e^{\bar{\alpha}}\beta_0 = o_P(1)$ implying $\sigma_t^4 - e^{2\bar{\alpha}}\beta_0^2 = o_P(1)$ by the rules of limit calculation, $\sigma_t^4 - e^{\bar{\alpha}}\beta_0^2 = o_P(1)$, $\sigma_t^6 - e^{\bar{\alpha}}\beta_0^3 = o_P(1)$, and $\sigma_t^6 - e^{2\bar{\alpha}}\beta_0^3 = o_P(1)$, respectively. The argument of the supremum in assumption (1.19) is a symmetric matrix equal to

$$\begin{aligned}
&\left|G_n^{-1/2}\left[\mathcal{F}_n(\theta)-\mathcal{F}_n(\theta_0)\right]G_n^{-1/2}\right| \\
&= \begin{bmatrix} \frac{|\mathcal{F}_n^{11}(\theta)-\mathcal{F}_n^{11}(\theta_0)|}{n^2} & \frac{|\mathcal{F}_n^{12}(\theta)-\mathcal{F}_n^{12}(\theta_0)|}{n^2} & \frac{|\mathcal{F}_n^{13}(\theta)-\mathcal{F}_n^{13}(\theta_0)|}{n^2} \\ & \frac{|\mathcal{F}_n^{21}(\theta)-\mathcal{F}_n^{21}(\theta_0)|}{n^2} & \frac{|\mathcal{F}_n^{22}(\theta)-\mathcal{F}_n^{22}(\theta_0)|}{n^2} \\ & & \frac{|\mathcal{F}_n^{33}(\theta)-\mathcal{F}_n^{33}(\theta_0)|}{n^3} \end{bmatrix}
\end{aligned} \tag{1.107}$$

where we calculate the limit element by element. So for the $(1,1)$ element, we have

$$\begin{aligned}
&\frac{1}{n^2}\left|\mathcal{F}_n^{11}(\theta)-\mathcal{F}_n^{11}(\theta_0)\right| \\
&= \frac{1}{n^2}\left|\sum_{t=1}^{n}\left(\frac{e^{2\bar{\alpha}}}{\sigma_t^2}-\frac{1}{\beta_0}\right)x_{t-1}^2+\sum_{t=1}^{n}\left(\frac{e_t}{\sqrt{\beta_0}}-\frac{e^{\bar{\alpha}}\hat{e}_t}{\sigma_t}\right)x_{t-1}\right| \\
&\leq \frac{1}{n^2}\sum_{t=1}^{n}\left|\frac{e^{2\bar{\alpha}}}{\sigma_t^2}-\frac{1}{\beta_0}\right|\left|x_{t-1}^2\right|+\frac{1}{n^2}\sum_{t=1}^{n}\left|\frac{e_t}{\sqrt{\beta_0}}-\frac{e^{\bar{\alpha}}\hat{e}_t}{\sigma_t}\right|\left|x_{t-1}\right|
\end{aligned} \tag{1.108}$$

where in the first sum

$$\begin{aligned}
\left|\frac{e^{2\bar{\alpha}}}{\sigma_t^2}-\frac{1}{\beta_0}\right| &= \frac{\left|e^{2\bar{\alpha}}\beta_0-\sigma_t^2\right|}{\sigma_t^2\beta_0} \\
&\leq \frac{2\left|e^{2\bar{\alpha}}\beta_0-\sigma_t^2\right|}{\beta_0^2} = o_P(1)
\end{aligned} \tag{1.109}$$

uniformly in $1 \le t \le n$ and in the second sum

$$\begin{aligned}
\left|\frac{e_t}{\sqrt{\beta_0}} - \frac{e^{\bar{\alpha}}\hat{e}_t}{\sigma_t}\right| &= \frac{\left|\sigma_t e_t - \sqrt{\beta_0}e^{\bar{\alpha}}\hat{e}_t\right|}{\sqrt{\beta_0}\sigma_t} \\
&= \frac{\left|\sigma_t^2\sqrt{\beta_0}e_t - \beta_0 e^{\bar{\alpha}}\sigma_t\hat{e}_t\right|}{\beta_0\sigma_t^2} \\
&= \frac{\left|\sigma_t^2(x_t - x_{t-1}) - \beta_0 e^{\bar{\alpha}}(x_t - e^{\bar{\alpha}}x_{t-1})\right|}{\beta_0\sigma_t^2} \\
&= \frac{\left|(\sigma_t^2 - \beta_0 e^{\bar{\alpha}})x_t + (\beta_0 e^{2\bar{\alpha}} - \sigma_t^2)x_{t-1}\right|}{\beta_0\sigma_t^2} \\
&\le \frac{\left|\sigma_t^2 - \beta_0 e^{\bar{\alpha}}\right|}{\beta_0\sigma_t^2}|x_t| + \frac{\left|\beta_0 e^{2\bar{\alpha}} - \sigma_t^2\right|}{\beta_0\sigma_t^2}|x_{t-1}| \\
&\le \frac{2\left|\sigma_t^2 - \beta_0 e^{\bar{\alpha}}\right|}{\beta_0^2}|x_t| + \frac{2\left|\beta_0 e^{2\bar{\alpha}} - \sigma_t^2\right|}{\beta_0^2}|x_{t-1}| = o_P(\sqrt{n}) \qquad (1.110)
\end{aligned}$$

uniformly in $1 \le t \le n$. Therefore, we obtain

$$\frac{1}{n^2}\left|\mathcal{F}_n^{11}(\theta) - \mathcal{F}_n^{11}(\theta_0)\right| \le \frac{1}{n}O_P(1) + \frac{1}{n}O_P(1) = o_P(1) \qquad (1.111)$$

and move on to the $(1,2)$ element

$$\begin{aligned}
&\frac{1}{n^{3/2}}\left|\mathcal{F}_n^{12}(\theta) - \mathcal{F}_n^{12}(\theta_0)\right| \\
&= \frac{1}{n^{3/2}}\left|\sum_{t=1}^{n}\left(\frac{e^{\bar{\alpha}}\hat{e}_t}{\sigma_t^3} - \frac{e_t}{\beta_0^{3/2}}\right)x_{t-1}\right| \\
&\le \frac{1}{n^{3/2}}\sum_{t=1}^{n}\left|\frac{e^{\bar{\alpha}}\hat{e}_t}{\sigma_t^3} - \frac{e_t}{\beta_0^{3/2}}\right||x_{t-1}| \qquad (1.112)
\end{aligned}$$

where

$$
\begin{aligned}
\left|\frac{e^{\bar{\alpha}}\hat{e}_t}{\sigma_t^3} - \frac{e_t}{\beta_0^{3/2}}\right| &= \frac{\left|\beta_0^{3/2} e^{\bar{\alpha}}\hat{e}_t - \sigma_t^3 e_t\right|}{\sigma_t^3 \beta_0^{3/2}} \\
&= \frac{\left|\beta_0^2 e^{\bar{\alpha}} \sigma_t \hat{e}_t - \sigma_t^4 \sqrt{\beta_0} e_t\right|}{\sigma_t^4 \beta_0^2} \\
&= \frac{\left|\beta_0^2 e^{\bar{\alpha}} (x_t - e^{\bar{\alpha}} x_{t-1}) - \sigma_t^4 (x_t - x_{t-1})\right|}{\sigma_t^4 \beta_0^2} \\
&= \frac{\left|(\beta_0^2 e^{\bar{\alpha}} - \sigma_t^4) x_t + (\sigma_t^4 - \beta_0^2 e^{2\bar{\alpha}}) x_{t-1}\right|}{\sigma_t^4 \beta_0^2} \\
&\leq \frac{\left|\beta_0^2 e^{\bar{\alpha}} - \sigma_t^4\right|}{\sigma_t^4 \beta_0^2} |x_t| + \frac{\left|\sigma_t^4 - \beta_0^2 e^{2\bar{\alpha}}\right|}{\sigma_t^4 \beta_0^2} |x_{t-1}| \\
&\leq \frac{4\left|\beta_0^2 e^{\bar{\alpha}} - \sigma_t^4\right|}{\beta_0^4} |x_t| + \frac{4\left|\sigma_t^4 - \beta_0^2 e^{2\bar{\alpha}}\right|}{\beta_0^4} |x_{t-1}| = o_P(\sqrt{n}) \quad (1.113)
\end{aligned}
$$

uniformly in $1 \leq t \leq n$ and thus

$$
\frac{1}{n^{3/2}} \left|\mathcal{F}_n^{12}(\theta) - \mathcal{F}_n^{12}(\theta_0)\right| \leq \frac{1}{\sqrt{n}} O_P(1) = o_P(1). \quad (1.114)
$$

For the $(1,3)$ element, we have

$$
\begin{aligned}
&\frac{1}{n^{5/2}} \left|\mathcal{F}_n^{13}(\theta) - \mathcal{F}_n^{13}(\theta_0)\right| \\
&= \frac{1}{n^{5/2}} \left|\sum_{t=1}^{n} \left(\frac{e^{\bar{\alpha}}\hat{e}_t}{\sigma_t^3} - \frac{e_t}{\beta_0^{3/2}}\right) x_{t-1}^3\right| \\
&\leq \frac{1}{n^{5/2}} \sum_{t=1}^{n} \left|\frac{e^{\bar{\alpha}}\hat{e}_t}{\sigma_t^3} - \frac{e_t}{\beta_0^{3/2}}\right| |x_{t-1}^3| \quad (1.115)
\end{aligned}
$$

and the result follows immediately

$$
\frac{1}{n^{5/2}} \left|\mathcal{F}_n^{13}(\theta) - \mathcal{F}_n^{13}(\theta_0)\right| \leq \frac{1}{\sqrt{n}} O_P(1) = o_P(1) \quad (1.116)
$$

by using equation (1.113). The $(2,2)$ element is equal to

$$\begin{aligned}
&\frac{1}{n}\left|\mathcal{F}_n^{22}(\theta)-\mathcal{F}_n^{22}(\theta_0)\right| \\
&=\frac{1}{n}\left|\sum_{t=1}^{n}\left(\frac{\hat{e}_t^2}{\sigma_t^4}-\frac{e_t^2}{\beta_0^2}\right)+\frac{1}{2}\sum_{t=1}^{n}\left(\frac{1}{\beta_0^2}-\frac{1}{\sigma_t^4}\right)\right| \\
&\leq \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{e}_t^2}{\sigma_t^4}-\frac{e_t^2}{\beta_0^2}\right|+\frac{1}{2n}\sum_{t=1}^{n}\left|\frac{1}{\beta_0^2}-\frac{1}{\sigma_t^4}\right|
\end{aligned} \tag{1.117}$$

where in the first sum

$$\begin{aligned}
\left|\frac{\hat{e}_t^2}{\sigma_t^4}-\frac{e_t^2}{\beta_0^2}\right| &= \frac{\left|\beta_0^2\hat{e}_t^2-\sigma_t^4 e_t^2\right|}{\sigma_t^4\beta_0^2} \\
&= \frac{\left|\beta_0^3\sigma_t^2\hat{e}_t^2-\sigma_t^6\beta_0 e_t^2\right|}{\sigma_t^6\beta_0^3} \\
&= \frac{\left|\beta_0^3(x_t-e^{\bar{\alpha}}x_{t-1})^2-\sigma_t^6(x_t-x_{t-1})^2\right|}{\sigma_t^6\beta_0^3} \\
&= \frac{\left|\beta_0^3(x_t^2-2e^{\bar{\alpha}}x_t x_{t-1}+e^{2\bar{\alpha}}x_{t-1}^2)-\sigma_t^6(x_t^2-2x_t x_{t-1}+x_{t-1}^2)\right|}{\sigma_t^6\beta_0^3} \\
&= \frac{\left|(\beta_0^3-\sigma_t^6)x_t^2+2(\sigma_t^6-e^{\bar{\alpha}}\beta_0^3)x_t x_{t-1}+(\beta_0^3 e^{2\bar{\alpha}}-\sigma_t^6)x_{t-1}^2\right|}{\sigma_t^6\beta_0^3} \\
&\leq \frac{\left|\beta_0^3-\sigma_t^6\right|}{\sigma_t^6\beta_0^3}x_t^2+\frac{2\left|\sigma_t^6-e^{\bar{\alpha}}\beta_0^3\right|}{\sigma_t^6\beta_0^3}|x_t||x_{t-1}|+\frac{\left|\beta_0^3 e^{2\bar{\alpha}}-\sigma_t^6\right|}{\sigma_t^6\beta_0^3}x_{t-1}^2 \\
&\leq \frac{8\left|\beta_0^3-\sigma_t^6\right|}{\beta_0^6}x_t^2+\frac{16\left|\sigma_t^6-e^{\bar{\alpha}}\beta_0^3\right|}{\beta_0^6}|x_t||x_{t-1}|+\frac{8\left|\beta_0^3 e^{2\bar{\alpha}}-\sigma_t^6\right|}{\beta_0^6}x_{t-1}^2 \\
&= O_P(\sqrt{n})
\end{aligned} \tag{1.118}$$

uniformly in $1 \leq t \leq n$ where in the penultimate row, we utilize some information about the convergence rates of the numerators, starting with $\sigma_t^2 = \beta_0 + O_P(1/\sqrt{n})$ from equation (1.104). Raising both sides to the third power yields $\sigma_t^6 - \beta_0^3 = O_P(1/\sqrt{n})$. $\sigma_t^6 - e^{\bar{\alpha}}\beta_0^3 = o_P(1)$ is calculated using $\beta - e^{\bar{\alpha}/3}\beta_0 = o_P(1)$ as in the similar case elaborated in equations (1.105) and (1.106). For all C, there exists some n' such that for all

$n \geq n'$, $\beta - \beta_0$ converges more slowly to zero than the exponential part, i.e. $\beta - e^{\bar{\alpha}/3}\beta_0 = O_P(1/\sqrt{n})$. That is, $|1/\sqrt{n}| > |1 - C^{1/n}|$ holds for all values of $C > 0$ from a certain n' on. By the same reasoning, $\beta_0^3 e^{2\bar{\alpha}} - \sigma_t^6 = O_P(1/\sqrt{n})$. And in the second sum

$$\begin{aligned} \left| \frac{1}{\beta_0^2} - \frac{1}{\sigma_t^4} \right| &= \frac{\left|\sigma_t^4 - \beta_0^2\right|}{\beta_0^2 \sigma_t^4} \\ &\leq \frac{4\left|\sigma_t^4 - \beta_0^2\right|}{\beta_0^4} = o_P(1) \end{aligned} \tag{1.119}$$

uniformly in $1 \leq t \leq n$ and therefore

$$\frac{1}{n}\left|\mathcal{F}_n^{22}(\theta) - \mathcal{F}_n^{22}(\theta_0)\right| \leq \frac{1}{n^{1/4}} O_P(1) + \frac{1}{2n} O_P(1) = o_P(1). \tag{1.120}$$

Then, we have for the $(2,3)$ element

$$\begin{aligned} &\frac{1}{n^2}\left|\mathcal{F}_n^{23}(\theta) - \mathcal{F}_n^{23}(\theta_0)\right| \\ &= \frac{1}{n^2}\left| \sum_{t=1}^{n} \left(\frac{\hat{e}_t^2}{\sigma_t^4} - \frac{e_t^2}{\beta_0^2} \right) x_{t-1}^2 + \frac{1}{2}\sum_{t=1}^{n} \left(\frac{1}{\beta_0^2} - \frac{1}{\sigma_t^4} \right) x_{t-1}^2 \right| \\ &\leq \frac{1}{n^2} \sum_{t=1}^{n} \left| \frac{\hat{e}_t^2}{\sigma_t^4} - \frac{e_t^2}{\beta_0^2} \right| \left|x_{t-1}^2\right| + \frac{1}{2n^2} \sum_{t=1}^{n} \left| \frac{1}{\beta_0^2} - \frac{1}{\sigma_t^4} \right| \left|x_{t-1}^2\right| \\ &= \frac{1}{n^{1/4}} O_P(1) + \frac{1}{2n} O_P(1) = o_P(1) \end{aligned} \tag{1.121}$$

where we make use of equations (1.118) and (1.119). And finally, for the $(3,3)$

element

$$
\begin{aligned}
&\frac{1}{n^3}\left|\mathcal{F}_n^{33}(\theta)-\mathcal{F}_n^{33}(\theta_0)\right| \\
&=\frac{1}{n^3}\left|\sum_{t=1}^{n}\left(\frac{\hat{e}_t^2}{\sigma_t^4}-\frac{e_t^2}{\beta_0^2}\right)x_{t-1}^4+\frac{1}{2}\sum_{t=1}^{n}\left(\frac{1}{\beta_0^2}-\frac{1}{\sigma_t^4}\right)x_{t-1}^4\right| \\
&\leq \frac{1}{n^3}\sum_{t=1}^{n}\left|\frac{\hat{e}_t^2}{\sigma_t^4}-\frac{e_t^2}{\beta_0^2}\right|\left|x_{t-1}^4\right|+\frac{1}{2n^3}\sum_{t=1}^{n}\left|\frac{1}{\beta_0^2}-\frac{1}{\sigma_t^4}\right|\left|x_{t-1}^4\right| \\
&=\frac{1}{n^{1/4}}O_P(1)+\frac{1}{2n}O_P(1)=o_P(1) \qquad (1.122)
\end{aligned}
$$

where we apply the results of equations (1.118) and (1.119) again such that assumption (1.19) of Lemma 1 is verified. We may now apply Lemma 1 to prove Theorem 1 by pre-multiplying $S=(S^1,S^2,S^3)'$ in equation (1.98) by the inverse of Y in equation (1.102)

$$
Z=Y^{-1}S=\begin{bmatrix}\frac{1}{\sqrt{J_2}} & 0 & 0\\ 0 & \sqrt{2}\beta_0 & 0\\ 0 & -\frac{\sqrt{2}\beta_0 J_2}{\sqrt{J_4-J_2^2}} & \frac{\sqrt{2}}{\sqrt{J_4-J_2^2}}\end{bmatrix}\begin{bmatrix}S^1\\ S^2\\ S^3\end{bmatrix}. \qquad (1.123)
$$

By the conclusion in (1.21), we see that

$$
d_n=-2\left[\mathcal{L}_n(\hat{\theta}_n^0)-\mathcal{L}_n(\hat{\theta}_n^1)\right]\xrightarrow{D}\left(Z^1\right)^2+\left(Z^3\right)^2\mathbb{1}\left(Z^3\geq 0\right). \qquad (1.124)
$$

as $n\to\infty$ where

$$
\begin{aligned}
Z^1=\frac{S^1}{\sqrt{J_2}}&=\frac{\int_0^1 W_2^1(s)dW_2^1(s)}{\sqrt{\int_0^1\left[W_2^1(s)\right]^2 ds}} \\
&=\frac{\frac{1}{2}\left[W_2^1(1)\right]^2-\frac{1}{2}}{\sqrt{\int_0^1\left[W_2^1(s)\right]^2 ds}} \qquad (1.125)
\end{aligned}
$$

by using Itô's formula in the numerator and

$$
\begin{aligned}
Z^3 =& \frac{\sqrt{2}}{\sqrt{J_4 - J_2^2}} \left(S^3 - \beta_0 J_2 S^2 \right) \\
=& \frac{\sqrt{2}}{\sqrt{\int_0^1 \left[W_2^1(s)\right]^4 ds - \left[\int_0^1 \left[W_2^1(s)\right]^2 ds\right]^2}} \\
& \times \left\{ \frac{\mu_3}{2} \int_0^1 \left[W_2^1(s)\right]^2 dW_2^1(s) + \frac{\sqrt{\mu_4 - \mu_3^2 - 1}}{2} \int_0^1 \left[W_2^1(s)\right]^2 dW_2^2(s) \right. \\
& \left. -\beta_0 \int_0^1 \left[W_2^1(s)\right]^2 ds \left[\frac{\mu_3}{2\beta_0} W_2^1(1) + \frac{\sqrt{\mu_4 - \mu_3^2 - 1}}{2\beta_0} W_2^2(1) \right] \right\} \\
=& \frac{1}{\sqrt{2}\sqrt{\int_0^1 \left[W_2^1(s)\right]^4 ds - \left[\int_0^1 \left[W_2^1(s)\right]^2 ds\right]^2}} \\
& \times \left\{ \mu_3 \int_0^1 \left[W_2^1(s)\right]^2 dW_2^1(s) + \sqrt{\mu_4 - \mu_3^2 - 1} \int_0^1 \left[W_2^1(s)\right]^2 dW_2^2(s) \right. \\
& \left. - \int_0^1 \left[W_2^1(s)\right]^2 ds \left[\mu_3 W_2^1(1) + \sqrt{\mu_4 - \mu_3^2 - 1} W_2^2(1) \right] \right\} \qquad (1.126)
\end{aligned}
$$

after some cancellations and the proof of Theorem 1 is completed. □

1.4 Size and Power Simulations

To evaluate the finite-sample performance of the test statistic in (1.15), we calculate critical values and run some size and power experiments. Starting values for $\bar{\alpha}$ are obtained by regressing observations x_t on x_{t-1} without intercept and taking the log of the least squares estimate. Then we regress the squared residuals from the first regression on an intercept and x_{t-1}^2 and save the respective least squares estimates, $\hat{\beta}$ and $\hat{\lambda}$, as starting values for β and λ. We substitute 10^{-6} for non-positive estimates of $e^{\bar{\alpha}}$, β, and λ. Values of β and λ are restricted to $(0, \infty)$ and $[0, \infty)$, respectively. That is, we simulate 100000 series of the null model of a random walk generated by the cumulative sum of independent $N(0,1)$ random variables for each sample size $n = 50, 100, 250, 500, 1000$ and estimate the critical values from the empirical distribution function. For comparison, we provide estimates from 100000 replications of the asymptotic quantiles derived from the limiting distribution in (1.50) to (1.52) where we assume $e_t \sim N(0,1)$, i.e. $\mu_3 = 0$ and $\mu_4 = 3$. Then we make use of the following discrete-time approximation

$$W_2^i(s) \approx \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} e_t^i \tag{1.127}$$

for $i = 1, 2$ and $n = 100000$ where the independent $N(0,1)$ random variables e_t^1 and e_t^2 are assumed to be mutually independent. Integrals over and with respect to Brownian motion are approximated by Riemann sums

$$\int_0^1 \left[W_2^1(s)\right]^j ds \approx \frac{1}{n} \sum_{s=0}^{1-1/n} \left[W_2^1(s)\right]^j \tag{1.128}$$

for $j = 2, 4$ and

$$\int_0^1 \left[W_2^1(s)\right]^2 dW_2^2(s) \approx \sum_{s=0}^{1-1/n} \left[W_2^1(s)\right]^2 \Delta_F W_2^2(s) \tag{1.129}$$

where Δ_F denotes forward differences. Table 1.1 presents the critical values corresponding to the 10%, 5%, and 1% level of significance for each sample size. The suggested procedure to receive starting values works quite well in case of a random walk. To ensure the global maximum of the log-pseudo-likelihood to be found in the size and power simulations below, we consider a more expensive approach where each sample is estimated five times. In a first run, the least squares estimates from above, $\hat{\beta}$ and $\hat{\lambda}$, serve as starting values. In the course of the next four runs, we draw initial values for β and

	Significance level		
n	10%	5%	1%
50	3.604	4.866	7.859
100	3.680	4.989	7.955
250	3.814	5.109	8.150
500	3.854	5.144	8.218
1000	3.939	5.243	8.379
∞	4.043	5.388	8.605

Table 1.1: Critical values.

λ from $LN(\log(\hat{\beta}), 100)$ and $LN(\log(\hat{\lambda}), 100)$, respectively, where the medians of the log-normal distributions correspond to the least squares estimates. The same initial value of β is used in estimation under the null and under the alternative. As $e^{\bar{\alpha}}$ is consistently estimated by least squares, the respective starting values are used in all five runs. The procedure is quite time-consuming but worthwhile, so we limit the number of replications to 1000. Again, β and λ are restricted to $(0, \infty)$ and $[0, \infty)$, respectively. As the critical values of the STUR test for sample sizes 50 to 1000 are very close to each other, we restrict our experiments to the sample size 250 and use the same critical values at the 5% level as Granger and Swanson (1997) for the DF/ADF and LMT/ALMT test, namely -1.95 and 0.168, respectively. For the STUR test, the 95% quantile is given in Table 1.1. We start with the size experiments where we compare the results of five different test statistics. Like Granger and Swanson (1997), we consider DF and ADF test of the null of a fixed unit root versus a stationary alternative using the t-statistic of ψ of the regression

$$\Delta x_t = \psi x_{t-1} + \sum_{i=1}^{p'} \varphi_i \Delta x_{t-i} + \epsilon_t \tag{1.130}$$

where the number of lagged differences $p' = \lfloor 12(n/100)^{0.25} \rfloor - 1$ is selected according to Schwert (1989) with $p' = 0$ in the DF test. We also apply LMT and ALMT test of the null of a fixed versus the alternative of a stochastic unit root given by the statistic

$$\hat{Z}_n = \frac{1}{n^{3/2}\hat{\sigma}_\epsilon^2\hat{\kappa}^2}\sum_{t=p''+3}^{n}\left(\sum_{j=p''+2}^{t-1}\hat{\epsilon}_j\right)^2\left(\hat{\epsilon}_j^2 - \hat{\sigma}_\epsilon^2\right) \tag{1.131}$$

obtained from

$$\Delta x_t = \gamma_0 + \gamma_1 t + \sum_{i=1}^{p''}\varphi_i \Delta x_{t-i} + \epsilon_t \tag{1.132}$$

where $\hat{\sigma}_\epsilon^2 = (1/n)\sum_{t=p''+2}^{n}\hat{\epsilon}_t^2$ and $\hat{\kappa}^2 = (1/n)\sum_{t=p''+2}^{n}(\hat{\epsilon}_t^2 - \hat{\sigma}_\epsilon^2)^2$ with $p'' = 0$ in the LMT test. As suggested by Leybourne et al. (1996), we include $p'' = 5$ lagged differences in the ALMT test. Then we simulate 1000 series of length 250 of an ARIMA(0,1,1) process

$$x_t = x_{t-1} + e_t + \theta e_{t-1} \tag{1.133}$$

with $\theta = -0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$ and $e_t \sim N(0,1)$. Rejection frequencies are summarized in Table 1.2. In case of strong positive correlation,

	θ								
Test	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
DF	0.951	0.618	0.306	0.137	0.046	0.012	0.011	0.005	0.008
ADF	0.046	0.033	0.051	0.043	0.049	0.038	0.053	0.045	0.039
LMT	0	0	0.003	0.031	0.06	0.083	0.148	0.163	0.168
ALMT	0.011	0.062	0.083	0.093	0.083	0.075	0.076	0.075	0.074
STUR	0.935	0.57	0.281	0.108	0.053	0.045	0.076	0.08	0.077

Table 1.2: Size in presence of serial correlation.

i.e. a negative MA coefficient of large modulus, DF and STUR strongly over-reject the null hypothesis whereas the LMT test performs weakly in presence of negative correlation. Using lagged differences in the ADF and ALMT test works to shift the empirical sizes close to their nominal values. For $\theta = 0$, rejection frequencies of all tests are almost equal to 5% except for ALMT where the nominal size is exceeded by about two percentage points. Although Leybourne et al. (1996) show in their Theorem 1 that the limiting distribution of $\hat{Z}_n$ is not affected by the number of lagged differences, severe differences in critical values may appear in finite samples. To illustrate this, we simulate critical values for 250 observations by generating 10000 replications, yielding 95% quantiles for LMT and ALMT with $p'' = 5$ of 0.178 and 0.204, respectively. Thus, the size of ALMT in Table 1.2 is overestimated with the critical value of 0.168 too low. Accounting for this effect, the nominal size of the ALMT test can be preserved. Next, we may study the absolute and relative power performance of the STUR test where we compare the results of three different test statistics. As the STUR test lacks the opportunity of augmenting, we skip this step here for reasons of comparison and report results for DF and LMT test only. First, we simulate 1000 series of length 350 of a stationary AR(1) process without intercept

$$x_t = \phi x_{t-1} + e_t \tag{1.134}$$

where $\phi = 0.8, 0.9, 0.95, 0.975, 0.99, 0.995$, $\sigma_t = 1$ for all t, and e_t independent $N(0,1)$ errors in equation (1.11) and drop the first 100. Results are shown in Table 1.3. The STUR test really competes with the DF test for $\phi \leq 0.9$,

	ϕ					
Test	0.8	0.9	0.95	0.975	0.99	0.995
DF	1	1	0.911	0.497	0.183	0.101
LMT	0	0.002	0.01	0.046	0.049	0.061
STUR	1	1	0.782	0.341	0.131	0.083

Table 1.3: Power against stationary alternatives.

showing moderate relative declines for coefficient values close to the unit root. Clearly, as the LMT test assumes any kind of unit root both under the null as under the alternative, we are not surprised by its weak performance. Next, we may study the power against STUR processes with a stochastic unit root, i.e. we consider the model class outlined in equations (1.1) to (1.3) with the mean of a_t equal to one. 1000 series of length 250 are simulated where we choose $\rho = 0.6$, $\sigma_\eta^2 = 0.00001, 0.0001, 0.001, 0.01, 0.1, 1$, and $\sigma_\varepsilon^2 = 100\sigma_\eta^2$ as in Granger and Swanson (1997), the intercept μ of the coefficient process α_t is set such that $E[a_t] = 1$. Obviously, the STUR test has power against stochastic

	σ_η^2					
Test	0.00001	0.0001	0.001	0.01	0.1	1
DF	0.042	0.048	0.042	0.114	0.967	0.991
LMT	0.051	0.06	0.161	0.644	0.928	0.862
STUR	0.045	0.055	0.308	0.971	1	1

Table 1.4: Power against STUR processes with $E[a_t] = 1$.

unit root processes. As $\rho = const.$, an increase in σ_η^2 implies an increase in σ_α^2 and thus in $Var[a_t] = (e^{\sigma_\alpha^2} - 1)E^2[a_t]$ where $E[a_t] = 1$. So more frequent rejections are related to stronger coefficient variation. The AR-ARCH test outperforms DF and LMT for all values of σ_η^2. Increasing the sample size to 500 strongly improves the power performance of LMT and STUR resulting in rejection frequencies for $\sigma_\eta^2 = 0.001$ of 0.331 and 0.692, respectively. STUR processes allow for both stationary and explosive regimes. Particularly, the probability of a unit root ($a_t = 1$ for a certain t) is equal to zero. This is true for STUR processes with $E[a_t] = 1$ (stochastic unit root) as well as for STUR processes with $E[a_t] < 1$. So we repeat the experiments summarized in Table 1.4 for the case where $E[a_t] = 0.99$. The results in Table 1.5 come quite close to those in Table 1.4 with some improvements in DF. It seems to be more illustrating to compare the results with Table 1.3, column 6 where

	σ_η^2					
Test	0.00001	0.0001	0.001	0.01	0.1	1
DF	0.167	0.134	0.155	0.246	0.972	0.998
LMT	0.064	0.037	0.113	0.561	0.916	0.863
STUR	0.112	0.088	0.217	0.948	1	1

Table 1.5: Power against STUR processes with $E[a_t] = 0.99$.

$\phi = 0.99$ but $\sigma_\eta^2 = 0$. In the STUR simulations $\sigma_\varepsilon^2 = 100\sigma_\eta^2$ is scaled whereas in the stationary series with $\sigma_\varepsilon^2 = 1$ it is not. That is, given power is non-sensitive to scaling in near unit-root cases sufficient coefficient variation may increase the rejection performance. Finally, we study the sensitivity of power results with respect to the parameter of coefficient correlation ρ. For this purpose, we fix $\sigma_\eta^2 = 0.001$ and assume $\rho = 0, 0.2, 0.4, 0.6, 0.8, 0.9, 0.95, 0.99$. Therefore, as ρ increases $\sigma_\alpha^2 = \sigma_\eta^2/(1-\rho^2)$ increases as well and $Var[a_t]$ goes up. Again, a stronger coefficient variation results in higher power as reported in Table 1.6. The variance of a_t is a function in σ_α^2 and thus in ρ^2, so the

	ρ							
Test	0	0.2	0.4	0.6	0.8	0.9	0.95	0.99
DF	0.054	0.059	0.063	0.05	0.064	0.063	0.096	0.492
LMT	0.113	0.114	0.131	0.159	0.253	0.395	0.529	0.849
STUR	0.218	0.226	0.267	0.315	0.512	0.73	0.904	1

Table 1.6: Power against STUR processes with different degrees of coefficient correlation.

rejection frequencies in Table 1.6 are symmetric around $\rho = 0$.

1.5 Application

We apply the new unit-root test to survey-based, seasonally-adjusted, monthly data on unemployment rates of 10 countries provided by the OECD: Australia, Brazil, Canada, Chile, Finland, Japan, Mexico, Sweden, UK, and US. We test for the null of a unit root in 250 demeaned observations ranging from May 1989 to February 2010 and report the results of DF, LMT, and STUR in Table 1.7. Autocorrelation and partial autocorrelation functions of the most unemployment series call for AR(1) models. To find the global maximum of the log-pseudo-likelihood function, we use starting values produced by the same procedure as in the size and power simulations above for different values of the variance parameters of the log-normal distributions. Critical values at the 1%, 5%, 10% level are equal to -2.58, -1.95, -1.62 for DF (Fuller, 1976, 373), 0.289, 0.168, 0.122 for LMT (Granger and Swanson, 1997), and 8.150, 5.109, 3.814 for STUR (see Table 1.1), respectively. Assuming now

Country	DF	LMT	STUR
Australia	-0.617	0.303^{***}	0.382
Brazil	-2.279^{**}	0.008	5.159^{**}
Canada	-1.023	0.904^{***}	6.472^{**}
Chile	-2.063^{**}	0.018	4.237^{*}
Finland	-2.355^{**}	6.032^{***}	10.577^{***}
Japan	-1.088	0.092	1.185
Mexico	-1.816^{*}	0.610^{***}	15.288^{***}
Sweden	-1.906^{*}	0.280^{**}	3.623
UK	-0.254	-0.040	0.065
US	1.589	0.113	8.809^{***}

*, **, *** denotes significance at the 10%, 5%, 1% level.

Table 1.7: Unit-root tests.

that the data are generated by a stochastic process considered in Section 1.4 except for some variance scaling, we make use of the following simulation results: Given that unemployment is well represented by an ARIMA(0,1,1) process with negative MA coefficient or an AR(1) process, a rejection by the DF test may correspond to a rejection by the STUR test and vice versa. By the same reasoning, if unemployment is generated by a STUR process and the LMT test rejects the null, the STUR test may reject as well. Thus, the results in Table 1.7 might provide some evidence in favor of the random walk hypothesis for Australia, Chile, Japan, Sweden, and the UK, however, for Canada, Finland, Mexico, and the US, STUR seems to be an appropriate model. Particularly for Finland, Mexico, and the US, the null is strongly rejected by the STUR test, in case of Finland and Mexico this is also supported by LMT. By estimating ARIMA(0,1,1), we find that this model also applies to the Canadian data with a positive MA coefficient whereas for Australia, Mexico, and Sweden obtaining negative MA coefficients, we may abstain from ARIMA(0,1,1). To model Brazilian unemployment, either ARIMA(0,1,1) with a negative MA coefficient or AR(1) can be possible alternatives. Clearly, the simulations in Section 1.4 are based on a great many of replications, so this interpretation of the test results in Table 1.7 must be handled with care. A more reliable analysis requires fitting the respective models to the data, generating a large number of replications based on these models and simulating the distributions of the individual test statistics.

1.6 Conclusions

This work considers a new unit-root test based on the pseudo-likelihood ratio statistic of an AR(1)-ARCH(1) model under the null of a random walk. Under the alternative, the autoregressive coefficient is restricted to positive values whereas the remaining parameters are completely flexible. The limiting distribution is derived and depends on the third and fourth moments of the i.i.d. errors. So critical values have to be simulated according to the moments estimated in advance. Klüppelberg et al. (2002) note that the high quantiles are strongly affected by the peaks of the series, as the resulting distribution is a heavy-tailed. Rather than to estimate critical values directly from the empirical distribution function, they suggest to use methods offered by extreme value theory and refer to Borkovec (2000). Due to the strong sensitivity to starting values, finding the maximum of the pseudo-likelihood function is quite time-consuming. Using consistent estimates obtained from an alternative procedure to initialize the algorithm may shorten the operating time substantially. That is, new methods in estimating AR(1)-ARCH(1) models are welcome.

Monte Carlo experiments provide evidence that the STUR test has power against stationary AR(1) processes and comes quite close to the performance of the DF test. However, as in the DF test, in presence of serial correlation its nominal size cannot be maintained. Future work may consider a test statistic allowing for more general STUR models concerning a higher lag order as well as dependent errors. The STUR test has strong power against STUR alternatives compared to the LMT test. It does not assume any kind of unit root to be present in the data-generating process which seems to be more realistic, particularly, as STUR models with coefficient mean equal to one do not differ qualitatively from those with coefficient mean less than one. In near-unit root cases, some coefficient variation may suffice to reject the random walk. Applying the ALMT test with five lagged differences, we report some overrejection and recommend to simulate separate critical values for each lag order. Otherwise, the null is rejected too often. In general, the power performance improves with the degree of coefficient variation, either driven directly or by coefficient correlation.

To sum up, we have constructed a new unit-root test which outperforms the

LMT test in presence of STUR alternatives, especially in case of small coefficient variation, and it also has power against stationary processes as the conventional unit-root tests. A final application of DF, LMT, and STUR test to unemployment rates provides some evidence that the majority of the series is generated by nonstationary processes. For some of them, the STUR process seems to be an attractive alternative to the random walk.

2 Fitting STUR Models to Unemployment – Evaluation of a new Unit-Root Test

2.1 Introduction

In the present study, a new unit-root test having power against so-called stochastic unit-root (STUR) models introduced in Holl (2013a) is evaluated. For this reason, STUR models are fitted to ten series of unemployment rates tested in Holl (2013a) to look whether the results of the estimations correspond to the decisions of the test. We estimate the STUR model presented in Granger and Swanson (1997)

$$u_t = a_t u_{t-1} + \varepsilon_t \tag{2.1}$$

where $a_t = e^{\alpha_t}$ and

$$\alpha_t = \mu + \rho\alpha_{t-1} + \eta_t \tag{2.2}$$

Not submitted.

with $|\rho| < 1$, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, $\eta_t \sim N(0, \sigma_\eta^2)$ and η_t independent of ε_t. That is, a_t follows a log-normal distribution with mean

$$E[a_t] = e^{m+\sigma_\alpha^2/2} \tag{2.3}$$

and variance

$$Var[a_t] = (e^{\sigma_\alpha^2} - 1)e^{2m+\sigma_\alpha^2} \tag{2.4}$$

where $m = \mu/(1-\rho)$ and $\sigma_\alpha^2 = \sigma_\eta^2/(1-\rho^2)$. To estimate the parameters in equations (2.1) and (2.2), Granger and Swanson (1997) consider two strategies: approximate maximum likelihood (AML) as discussed in Guyton et al. (1986) and approximate minimum sum squares (AMSS). Only AML performs successful in simulations under certain conditions. Effects resulting from a randomization parameter in use are not that clear. Estimates are fairly imprecise for smaller values of σ_η^2. For both AML and AMSS asymptotic behavior is still unsolved. In course of the present paper, STUR models are fitted to unemployment rates. Problems arise as the coefficient process is unobserved. Thus, we apply the Bayesian approach developed in Jones and Marriott (1999) to overcome that issue. In particular, Bayesian estimation allows for simulation of the distribution of the coefficient process. That is, on the one hand a STUR model can be fitted and on the other hand its 'distance' to a stationary AR(1) or random walk model can be evaluated by studying the resulting coefficient distribution in terms of e.g. mean, variance and symmetry.

The stochastic unit-root (STUR) model introduced by Granger and Swanson (1997) considers the autoregressive model of order one (AR(1)) where the coefficient is allowed to vary around a constant mean. That is, the resulting process may behave stationary for some periods and explosive for others. In case of a stochastic unit root, the respective coefficient mean is equal to one. Compared to a fixed unit-root process where the first differences are stationary, a STUR process having a stochastic unit root cannot be differenced to stationarity. However, conventional cointegration and error-correction techniques rely on the assumption of difference stationarity, calling for being careful in advance.

In Section 2.2, the relevance of STUR models in economics is motivated using the example of unemployment rates. Then, in Section 2.3, we have a look at the unit-root test results on unemployment rates reported in Holl (2013a). There, a new test of the null hypothesis of a unit root is introduced which i.a. has power against STUR models. To evaluate the results, we fit STUR models to the unemployment series by using Bayesian techniques discussed in Section 2.4. Diagnostics used to ensure convergence are introduced in Section 2.5. Estimation results are presented in Section 2.6. As the Bayesian procedure is quite time-consuming, we consider an alternative rough estimator in Section 2.7. Finally, we may draw some conclusions on whether STUR is adequate to model unemployment rates.

2.2 Motivation

Unfortunately, there is little economic theory resulting in fixed unit-root processes or STUR processes with a stochastic unit root. Granger and Swanson (1997) refer to Hall (1978) who is testing the life cycle-permanent income hypothesis by using the regression $c_t = \lambda_t c_{t-1} + \varepsilon_t$ where $\lambda_t = \lambda$ is a function of the rate of subjective time preference, the real rate of interest and the elasticity of marginal utility which are assumed constant over time. $\lambda = 1$ is associated with a fixed unit root corresponding to the life cycle-permanent income hypothesis. In contrast, by allowing the real rate of interest to change over time such that λ_t is varying around a mean equal to one, a stochastic unit root will appear.

To discriminate between the natural-rate and the hysteresis hypothesis in unemployment, Phelps and Zoega (1998) discuss the estimation of the regression

$$u_t = u_t^*(S_j) + \beta[u_{t-1} - u_t^*(S_j)] + \varepsilon_t(S_j) \tag{2.5}$$

where S_j denotes different states of the world corresponding to aggregate supply, $0 \leq \beta \leq 1$ serves as a measure of persistence and ε_t is independent and identically distributed (i.i.d.) with mean 0 and variance $\sigma_\varepsilon^2(S_j)$. For $\beta < 1$, unemployment is stationary around its respective means $u_t^*(S_j)$, for $\beta = 1$ unemployment follows a random walk, the intercept term vanishes. Explosive regimes where $\beta > 1$ are excluded a priori. We either have to opt for a stationary or a difference-stationary regime and nothing in between. However, given that the natural rate of unemployment depends on past unemployment rather than on different states of the world, such that

$$u_t^* = \phi_0 + \phi_1 u_{t-1} \tag{2.6}$$

we obtain

$$u_t = (1-\beta)\phi_0 + [(1-\beta)\phi_1 + \beta]u_{t-1} + \varepsilon_t \tag{2.7}$$

by substituting the right-hand side of equation (2.6) for $u_t^*(S_j)$ in equation (2.5) and therefore β, the measure of persistence, is not identified anymore.

Alternatively, we may depart from the accelerationist Phillips curve as supported by Friedman (1968)

$$\pi_t = \pi_{t-1} + \gamma(u_t - u^*). \tag{2.8}$$

where u^* is determined by supply-side factors. In contrast to Friedman, we allow the natural rate of unemployment to depend on past unemployment and thus on aggregate demand too which is transmitted by a certain hysteresis mechanism

$$u_t^* = s_t + a_t u_{t-1} \tag{2.9}$$

where changes due to aggregate supply and/or a more sophisticated demand structure are caught by s_t and may vary over time as well. Ball (2009), thoroughly studying the empirics of unemployment and its natural rate, notes that changes in u sometimes cause changes in u^* and sometimes do not. He describes the hysteresis mechanism as one depending on the past history of u^* and the length of time that u is pushed away from u^*. To somehow capture this behavior, we allow for a random coefficient a_t which may be correlated over time. By substituting u_t^* in equation (2.9) for u^* in equation (2.8), we obtain

$$u_t = a_t u_{t-1} + \varepsilon_t \tag{2.10}$$

assuming

$$\varepsilon_t = \frac{\pi_t - \pi_{t-1}}{\gamma} + s_t \tag{2.11}$$

to be $i.i.d.(0, \sigma_\varepsilon^2)$ such that the STUR model by Granger and Swanson (1997) can be applied. To gain some knowledge about the supply-side influences on the natural rate, the residuals of the STUR model can be studied afterwards. STUR models allow for stationary and explosive regimes (and the random walk in between) corresponding to the random autoregressive coefficient which may be correlated over time. That is, apart from the natural rate itself regimes are path-dependent.

2.3 Unit-Root Tests

We consider the results of the tests for the null of a unit root in 250 demeaned observations ranging from May 1989 to February 2010 presented in Holl (2013a). Test statistics from Dickey and Fuller (1979) (DF), Leybourne et al. (1996) (LMT), and Holl (2013a) (STUR) are calculated for survey-based, seasonally-adjusted, monthly data on unemployment rates for Australia, Brazil, Canada, Chile, Finland, Japan, Mexico, Sweden, UK, and US. DF, LMT, and STUR all share the same null hypothesis of a fixed unit root. DF and LMT test against a stationary alternative and a stochastic unit root, respectively. The STUR test, however, based on the model given in equations (2.1) and (2.2) allows for both under the alternative. Under the null hypothesis of a random walk, $\mu = 0$ and $\sigma_\eta^2 = 0$ and thus a_t is constant and equal to one. $\mu < 0$ and $\sigma_\eta^2 = 0$ represents a stationary AR(1) model. $m + \sigma_\alpha^2/2 = 0$ and $\sigma_\eta^2 > 0$ is associated with a stochastic unit root corresponding to $E[a_t] = 1$. Furthermore, the test has power against processes generated by parameter values $m + \sigma_\alpha^2/2 < 0$ and $\sigma_\eta^2 > 0$. Results of DF, LMT, and STUR are reported in Table 2.1. The STUR test rejects the null of a unit

Country	DF	LMT	STUR
Australia	-0.617	0.303^{***}	0.382
Brazil	-2.279^{**}	0.008	5.159^{**}
Canada	-1.023	0.904^{***}	6.472^{**}
Chile	-2.063^{**}	0.018	4.237^{*}
Finland	-2.355^{**}	6.032^{***}	10.577^{***}
Japan	-1.088	0.092	1.185
Mexico	-1.816^{*}	0.610^{***}	15.288^{***}
Sweden	-1.906^{*}	0.280^{**}	3.623
UK	-0.254	-0.040	0.065
US	1.589	0.113	8.809^{***}

*, **, *** denotes significance at the 10%, 5%, 1% level.

Table 2.1: Unit-root tests.

root for Brazil, Canada, Finland, Mexico, and the US at least at the 5% level of significance. Particularly for Canada, Finland, and Mexico, the LMT test agrees with the STUR decision. According to simulations based on a great many of replications in Holl (2013a), rejection by the STUR test in presence of an ARIMA(0,1,1) process with negative MA coefficient or an AR(1) process may correspond to rejection by DF and vice versa. Given that the true process is STUR and the LMT test rejects the null, the STUR test may reject as well. Thus, to evaluate the results in Table 2.1, we fit STUR models to the respective series to decide whether a certain test may have rejected the null or not. Stationary AR(1) and the random walk are nested in STUR. Estimated coefficient variances close to zero indicate so-called fixed-coefficient processes.

2.4 Bayesian Estimation

Parameters of the STUR model by Granger and Swanson (1997) as presented in equations (2.1) and (2.2) are now estimated using Markov chain Monte Carlo techniques. The vector of the parameters of interest is equal to

$$\theta = (m, \rho, h_\varepsilon, h_\eta)' \tag{2.12}$$

where $m = \mu/(1-\rho)$ is the mean of α_t, $h_\varepsilon = 1/\sigma_\varepsilon^2$ and $h_\eta = 1/\sigma_\eta^2$ represent the respective error precisions. The vector including α_t, $t = 1, \ldots, n$ is treated as a vector of parameters as well.

Now the posterior densities given the data u and conditional on u_1 and α_1 can be calculated where we mainly follow Jones and Marriott (1999) except for the prior specification. Corresponding to Yang and Leon-Gonzalez (2010), we assume the parameters in equation (2.12) to be independent of each other such that

$$p(\theta) = p(m)p(\rho)p(h_\varepsilon)p(h_\eta) \tag{2.13}$$

where $p(\cdot)$ denotes the probability density function of the respective random variable. In general, notation is due to Koop (2003). By using Bayes' rule, we obtain

$$p(\theta, \alpha|u) = \frac{p(m|\theta_{-m}, \alpha, u)p(\theta_{-m}, \alpha, u)}{p(u)} = \frac{p(u|\alpha, \theta)p(\alpha|\alpha_{-t}, \theta)p(\theta)}{p(u)} \tag{2.14}$$

with $u = (u_1, u_2, \ldots, u_n)'$, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)'$ and thus $\alpha_{-t} = (\alpha_1, \ldots, \alpha_{t-1}, \alpha_{t+1}, \ldots, \alpha_n)'$ for a certain t, $\theta_{-m} = (\rho, h_\varepsilon, h_\eta)'$. The posterior density of m results in

$$\begin{aligned} p(m|\theta_{-m}, \alpha, u) &= \frac{p(u|\alpha, \theta)p(\alpha|\alpha_{-t}, \theta)p(\theta)}{p(\theta_{-m}, \alpha, u)} \\ &= \frac{p(u|\alpha, \theta)p(\alpha|\alpha_{-t}, \theta)p(m)p(\theta_{-m})}{p(u|\alpha, \theta_{-m})p(\alpha, \theta_{-m})} \\ &\propto p(\alpha|\alpha_{-t}, \theta)p(m) \end{aligned} \tag{2.15}$$

as m is independent of θ_{-m}, $p(u|\alpha, \theta) = p(u|\alpha, \theta_{-m})$ and θ_{-m} and α are given. Analogous to equation (2.14),

$$p(\theta, \alpha|u) = \frac{p(\rho|\theta_{-\rho}, \alpha, u)p(\theta_{-\rho}, \alpha, u)}{p(u)} = \frac{p(u|\alpha, \theta)p(\alpha|\alpha_{-t}, \theta)p(\theta)}{p(u)} \tag{2.16}$$

where $\theta_{-\rho} = (m, h_\varepsilon, h_\eta)'$. Hence, for the posterior of ρ we obtain

$$\begin{aligned} p(\rho|\theta_{-\rho}, \alpha, u) &= \frac{p(u|\alpha, \theta)p(\alpha|\alpha_{-t}, \theta)p(\theta)}{p(\theta_{-\rho}, \alpha, u)} \\ &= \frac{p(u|\alpha, \theta)p(\alpha|\alpha_{-t}, \theta)p(\rho)p(\theta_{-\rho})}{p(u|\alpha, \theta_{-\rho})p(\alpha, \theta_{-\rho})} \\ &\propto p(\alpha|\alpha_{-t}, \theta)p(\rho) \end{aligned} \tag{2.17}$$

as ρ is independent of $\theta_{-\rho}$, $p(u|\alpha, \theta) = p(u|\alpha, \theta_{-\rho})$ and $\theta_{-\rho}$ and α are given. We turn now to the calculation of the posterior densities of the error precisions where

$$p(\theta, \alpha|u) = \frac{p(h_\varepsilon|\theta_{-h_\varepsilon}, \alpha, u)p(\theta_{-h_\varepsilon}, \alpha, u)}{p(u)} = \frac{p(u|\alpha, \theta)p(\alpha, \theta)}{p(u)} \tag{2.18}$$

with $\theta_{-h_\varepsilon} = (m, \rho, h_\eta)'$ and thus

$$\begin{aligned}
p(h_\varepsilon|\theta_{-h_\varepsilon}, \alpha, u) &= \frac{p(u|\alpha,\theta)p(\alpha,\theta)}{p(\theta_{-h_\varepsilon}, \alpha, u)} \\
&= \frac{p(u|\alpha,\theta)p(\alpha, h_\varepsilon, \theta_{-h_\varepsilon})}{p(\theta_{-h_\varepsilon}, \alpha, u)} \\
&= \frac{p(u|\alpha,\theta)p(\alpha, \theta_{-h_\varepsilon}|h_\varepsilon)p(h_\varepsilon)}{p(\theta_{-h_\varepsilon}, \alpha, u)} \\
&\propto p(u|\alpha,\theta)p(h_\varepsilon) \qquad (2.19)
\end{aligned}$$

as θ_{-h_ε}, α and u are given. The posterior of h_η derives from

$$p(\theta, \alpha|u) = \frac{p(h_\eta|\theta_{-h_\eta}, \alpha, u)p(\theta_{-h_\eta}, \alpha, u)}{p(u)} = \frac{p(u|\alpha,\theta)p(\alpha|\alpha_{-t}, \theta)p(\theta)}{p(u)} \qquad (2.20)$$

with $\theta_{-h_\eta} = (m, \rho, h_\varepsilon)'$ and is equal to

$$\begin{aligned}
p(h_\eta|\theta_{-h_\eta}, \alpha, u) &= \frac{p(u|\alpha,\theta)p(\alpha|\alpha_{-t}, \theta)p(\theta)}{p(\theta_{-h_\eta}, \alpha, u)} \\
&= \frac{p(u|\alpha,\theta)p(\alpha|\alpha_{-t}, \theta)p(h_\eta)p(\theta_{-h_\eta})}{p(u|\alpha, \theta_{-h_\eta})p(\alpha, \theta_{-h_\eta})} \\
&\propto p(\alpha|\alpha_{-t}, \theta)p(h_\eta) \qquad (2.21)
\end{aligned}$$

as h_η is independent of θ_{-h_η}, $p(u|\alpha,\theta) = p(u|\alpha, \theta_{-h_\eta})$ and θ_{-h_η} and α are given. And finally, the posterior of a certain α_t may be calculated from

$$p(\theta, \alpha|u) = \frac{p(\alpha_t|\theta, \alpha_{-t}, u)p(\theta, \alpha_{-t}, u)}{p(u)} = \frac{p(u|\alpha,\theta)p(\alpha,\theta)}{p(u)} \qquad (2.22)$$

and results in

$$
\begin{aligned}
p(\alpha_t|\theta, \alpha_{-t}, u) = & \quad \frac{p(u|\alpha, \theta)p(\alpha, \theta)}{p(\theta, \alpha_{-t}, u)} \\
= & \quad \frac{p(u|\alpha, \theta)p(\alpha|\alpha_{-t}, \theta)p(\theta)}{p(\theta, \alpha_{-t}, u)} \\
\propto & \quad p(u|\alpha, \theta)p(\alpha|\alpha_{-t}, \theta) \qquad (2.23)
\end{aligned}
$$

as θ, α_{-t} and u are given. To calculate the full conditional densities in equations (2.15), (2.17), (2.19), (2.21), and (2.23), expressions for $p(\alpha|\alpha_{-t}, \theta)$, $p(u|\alpha, \theta)$, and the prior densities of m, ρ, h_ε, and h_η are required. As $\eta_t \sim N(0, \sigma_\eta^2)$ and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, the conditional density of α_t is equal to

$$
\begin{aligned}
p(\alpha_t|\alpha_{t-1}, \theta) = & \quad \frac{1}{\sqrt{2\pi\sigma_\eta^2}} e^{-\frac{1}{2}\frac{[\alpha_t - (\mu + \rho\alpha_{t-1})]^2}{\sigma_\eta^2}} \\
= & \quad \frac{1}{\sqrt{2\pi\sigma_\eta^2}} e^{-\frac{1}{2}\frac{[(\alpha_t - m) - \rho(\alpha_{t-1} - m)]^2}{\sigma_\eta^2}} \qquad (2.24)
\end{aligned}
$$

as $\mu = m(1-\rho)$ and the conditional density of u_t results in

$$
p(u_t|\alpha_t, u_{t-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\frac{1}{2}\frac{(u_t - e^{\alpha_t}u_{t-1})^2}{\sigma_\varepsilon^2}}. \qquad (2.25)
$$

By using equations (2.24) and (2.25), $p(\alpha|\alpha_{-t}, \theta)$ and $p(u|\alpha, \theta)$ can be expressed as products of Gaussian densities such that

$$
\begin{aligned}
p(\alpha|\alpha_{-t}, \theta) = & \quad \prod_{t=2}^{n} p(\alpha_t|\alpha_{-t}, \theta) \\
= & \quad \frac{1}{\left(\sqrt{2\pi\sigma_\eta^2}\right)^{n-1}} e^{-\frac{1}{2\sigma_\eta^2}\sum_{t=2}^{n}[(\alpha_t - m) - \rho(\alpha_{t-1} - m)]^2} \qquad (2.26)
\end{aligned}
$$

and

$$\begin{aligned} p(u|\alpha,\theta) &= \prod_{t=2}^{n} p(u_t|\alpha_t, u_{t-1}, \theta) \\ &= \frac{1}{\left(\sqrt{2\pi\sigma_\varepsilon^2}\right)^{n-1}} e^{-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=2}^{n}(u_t - e^{\alpha_t}u_{t-1})^2}. \end{aligned} \tag{2.27}$$

Next, normal prior densities for m and ρ are specified, $N(\underline{\mu}_m, \underline{V}_m)$ and $N(\underline{\mu}_\rho, \underline{V}_\rho)$ over $|\rho| < 1$, respectively, thus

$$p(m) = \frac{1}{\sqrt{2\pi\underline{V}_m}} e^{-\frac{(m-\underline{\mu}_m)^2}{2\underline{V}_m}} \tag{2.28}$$

and

$$p(\rho) = \frac{1}{\underline{C}\sqrt{2\pi\underline{V}_\rho}} e^{-\frac{(\rho-\underline{\mu}_\rho)^2}{2\underline{V}_\rho}} \tag{2.29}$$

where $|\rho| < 1$, $\underline{C}$ denotes the normalizing constant over the restricted region. We choose Gamma prior densities for h_ε and h_η, $Gamma(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon)$ and $Gamma(\underline{\alpha}_\eta, \underline{\beta}_\eta)$, respectively, that is

$$p(h_\varepsilon) = \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon}\Gamma(\underline{\alpha}_\varepsilon)} h_\varepsilon^{\underline{\alpha}_\varepsilon - 1} e^{-\frac{h_\varepsilon}{\underline{\beta}_\varepsilon}} \tag{2.30}$$

and

$$p(h_\eta) = \frac{1}{\underline{\beta}_\eta^{\underline{\alpha}_\eta}\Gamma(\underline{\alpha}_\eta)} h_\eta^{\underline{\alpha}_\eta - 1} e^{-\frac{h_\eta}{\underline{\beta}_\eta}}. \tag{2.31}$$

Now the full conditional densities of m, ρ, h_ε, h_η, and α_t can be calculated.

From equations (2.15), (2.26), and (2.28) follows

$$
\begin{aligned}
p(m|\theta_{-m},\alpha,u) \propto \quad & \frac{1}{\left(\sqrt{2\pi\sigma_\eta^2}\right)^{n-1}} e^{-\frac{1}{2\sigma_\eta^2}\sum_{t=2}^{n}[(\alpha_t-m)-\rho(\alpha_{t-1}-m)]^2} \frac{1}{\sqrt{2\pi\underline{V}_m}} e^{-\frac{(m-\underline{\mu}_m)^2}{2\underline{V}_m}} \\
\propto \quad & e^{-\frac{1}{2\sigma_\eta^2}\sum_{t=2}^{n}[(\alpha_t-m)-\rho(\alpha_{t-1}-m)]^2} e^{-\frac{(m-\underline{\mu}_m)^2}{2\underline{V}_m}} \\
= \quad & e^{-\frac{1}{2\sigma_\eta^2}\sum_{t=2}^{n}[(\alpha_t-\rho\alpha_{t-1})-m(1-\rho)]^2} e^{-\frac{m^2-2m\underline{\mu}_m+\underline{\mu}_m^2}{2\underline{V}_m}\frac{h_\eta}{h_\eta}} \\
= \quad & e^{-\frac{h_\eta}{2}\sum_{t=2}^{n}[m(1-\rho)-(\alpha_t-\rho\alpha_{t-1})]^2} e^{-\frac{h_\eta}{2}\frac{m^2-2m\underline{\mu}_m+\underline{\mu}_m^2}{h_\eta\underline{V}_m}} \\
= \quad & e^{-\frac{h_\eta}{2}\sum_{t=2}^{n}[m^2(1-\rho)^2-2m(1-\rho)(\alpha_t-\rho\alpha_{t-1})+(\alpha_t-\rho\alpha_{t-1})^2]} \\
& \times e^{-\frac{h_\eta}{2}\left(\frac{m^2}{h_\eta\underline{V}_m}-\frac{2m\underline{\mu}_m}{h_\eta\underline{V}_m}+\frac{\underline{\mu}_m^2}{h_\eta\underline{V}_m}\right)} \\
= \quad & e^{-\frac{h_\eta}{2}\left[m^2(n-1)(1-\rho)^2-2m(1-\rho)\sum_{t=2}^{n}(\alpha_t-\rho\alpha_{t-1})+\sum_{t=2}^{n}(\alpha_t-\rho\alpha_{t-1})^2\right]} \\
& \times e^{-\frac{h_\eta}{2}\left(\frac{m^2}{h_\eta\underline{V}_m}-\frac{2m\underline{\mu}_m}{h_\eta\underline{V}_m}+\frac{\underline{\mu}_m^2}{h_\eta\underline{V}_m}\right)} \\
= \quad & e^{-\frac{h_\eta}{2}\left\{m^2\left[(n-1)(1-\rho)^2+\frac{1}{h_\eta\underline{V}_m}\right]-2m\left[(1-\rho)\sum_{t=2}^{n}(\alpha_t-\rho\alpha_{t-1})+\frac{\underline{\mu}_m}{h_\eta\underline{V}_m}\right]\right\}} \\
& \times e^{-\frac{h_\eta}{2}\left[\sum_{t=2}^{n}(\alpha_t-\rho\alpha_{t-1})^2+\frac{\underline{\mu}_m^2}{h_\eta\underline{V}_m}\right]} \\
\propto \quad & e^{-\frac{h_\eta}{2}\left\{m^2\left[(n-1)(1-\rho)^2+\frac{1}{h_\eta\underline{V}_m}\right]-2m\left[(1-\rho)\sum_{t=2}^{n}(\alpha_t-\rho\alpha_{t-1})+\frac{\underline{\mu}_m}{h_\eta\underline{V}_m}\right]\right\}} \\
= \quad & e^{-\frac{1}{2}\left(\frac{m^2}{\overline{V}_m}-2m\frac{\overline{\mu}_m}{\overline{V}_m}\right)}
\end{aligned}
\tag{2.32}
$$

where

$$\overline{V}_m = \frac{1}{h_\eta\left[(n-1)(1-\rho)^2+\frac{1}{h_\eta\underline{V}_m}\right]} \tag{2.33}$$

and

$$\overline{\mu}_m = \frac{(1-\rho)\sum_{t=2}^{n}(\alpha_t-\rho\alpha_{t-1})+\frac{\underline{\mu}_m}{h_\eta\underline{V}_m}}{(n-1)(1-\rho)^2+\frac{1}{h_\eta\underline{V}_m}} \tag{2.34}$$

are the variance and the mean of the normal full conditional distribution of m, respectively. From equations (2.17), (2.26), and (2.29) follows

$$\begin{aligned}
p(\rho|\theta_{-\rho,\alpha,u}) &\propto \frac{1}{\left(\sqrt{2\pi\sigma_\eta^2}\right)^{n-1}} e^{-\frac{1}{2\sigma_\eta^2}\sum_{t=2}^{n}[(\alpha_t-m)-\rho(\alpha_{t-1}-m)]^2} \\
&\quad \times \frac{1}{\underline{C}\sqrt{2\pi\underline{V}_\rho}} e^{-\frac{(\rho-\underline{\mu}_\rho)^2}{2\underline{V}_\rho}} \\
&\propto e^{-\frac{h_\eta}{2}\sum_{t=2}^{n}[(\alpha_t-m)-\rho(\alpha_{t-1}-m)]^2-\frac{1}{2\underline{V}_\rho}(\rho-\underline{mu}_\rho)^2} \\
&= e^{-\frac{h_\eta}{2}\left\{\sum_{t=2}^{n}[(\alpha_t-m)-\rho(\alpha_{t-1}-m)]^2+\frac{1}{h_\eta\underline{V}_\rho}(\rho-\underline{mu}_\rho)^2\right\}} \\
&= e^{-\frac{h_\eta}{2}\left\{\sum_{t=2}^{n}[\rho(\alpha_{t-1}-m)-(\alpha_t-m)]^2+\frac{1}{h_\eta\underline{V}_\rho}(\rho-\underline{mu}_\rho)^2\right\}} \\
&\propto e^{-\frac{h_\eta}{2}\left\{\rho^2\sum_{t=2}^{n}(\alpha_{t-1}-m)^2-2\rho\sum_{t=2}^{n}(\alpha_{t-1}-m)(\alpha_t-m)+\frac{1}{h_\eta\underline{V}_\rho}(\rho^2-2\rho\underline{mu}_\rho)\right\}} \\
&= e^{-\frac{h_\eta}{2}\left\{\rho^2\left[\sum_{t=2}^{n}(\alpha_{t-1}-m)^2+\frac{1}{h_\eta\underline{V}_\rho}\right]-2\rho\left[\sum_{t=2}^{n}(\alpha_{t-1}-m)(\alpha_t-m)+\frac{\underline{mu}_\rho}{h_\eta\underline{V}_\rho}\right]\right\}} \\
&= e^{-\frac{1}{2}\left(\frac{\rho^2}{\overline{V}_\rho}-2\rho\frac{\overline{mu}_\rho}{\overline{V}_\rho}\right)}
\end{aligned} \tag{2.35}$$

where

$$\overline{V}_\rho = \frac{1}{h_\eta\sum_{t=2}^{n}(\alpha_{t-1}-m)^2+\frac{1}{\underline{V}_\rho}} \tag{2.36}$$

and

$$\overline{\mu}_\rho = \frac{h_\eta\left[\sum_{t=2}^{n}(\alpha_{t-1}-m)(\alpha_t-m)+\frac{\underline{mu}_\rho}{h_\eta\underline{V}_\rho}\right]}{h_\eta\sum_{t=2}^{n}(\alpha_{t-1}-m)^2+\frac{1}{\underline{V}_\rho}} \tag{2.37}$$

are the variance and the mean of the normal full conditional distribution of ρ, respectively. Combining equations (2.17), (2.26), and (2.29), results in

$$\begin{aligned} p(h_\varepsilon|\theta_{-h_\varepsilon},\alpha,u) &\propto \frac{1}{\left(\sqrt{2\pi\sigma_\varepsilon^2}\right)^{n-1}} e^{-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=2}^{n}(u_t-e^{\alpha_t}u_{t-1})^2} \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon}\Gamma(\underline{\alpha}_\varepsilon)} h_\varepsilon^{\underline{\alpha}_\varepsilon-1} e^{-\frac{h_\varepsilon}{\underline{\beta}_\varepsilon}} \\ &\propto h_\varepsilon^{\frac{n-1}{2}} e^{-\frac{h_\varepsilon}{2}\sum_{t=2}^{n}(u_t-e^{\alpha_t}u_{t-1})^2} h_\varepsilon^{\underline{\alpha}_\varepsilon-1} e^{-\frac{h_\varepsilon}{\underline{\beta}_\varepsilon}} \\ &= h_\varepsilon^{\underline{\alpha}_\varepsilon+\frac{n-1}{2}-1} e^{-h_\varepsilon\left[\frac{1}{\underline{\beta}_\varepsilon}+\frac{1}{2}\sum_{t=2}^{n}(u_t-e^{\alpha_t}u_{t-1})^2\right]} \\ &= h_\varepsilon^{\overline{\alpha}_\varepsilon-1} e^{-\frac{h_\varepsilon}{\overline{\beta}_\varepsilon}} \end{aligned} \tag{2.38}$$

where

$$\overline{\alpha}_\varepsilon = \underline{\alpha}_\varepsilon + \frac{n-1}{2} \tag{2.39}$$

and

$$\overline{\beta}_\varepsilon = \frac{1}{\frac{1}{\underline{\beta}_\varepsilon} + \frac{1}{2}\sum_{t=2}^{n}(u_t - e^{\alpha_t}u_{t-1})^2} \tag{2.40}$$

are the shape and the scale parameter of the Gamma full conditional distribution of h_ε, respectively. Equations (2.21), (2.26), and (2.31) are used to derive

$$\begin{aligned} p(h_\eta|\theta_{-h_\eta},\alpha,u) &\propto \frac{1}{\left(\sqrt{2\pi\sigma_\eta^2}\right)^{n-1}} e^{-\frac{1}{2\sigma_\eta^2}\sum_{t=2}^{n}[(\alpha_t-m)-\rho(\alpha_{t-1}-m)]^2} \\ &\quad \times \frac{1}{\underline{\beta}_\eta^{\underline{\alpha}_\eta}\Gamma(\underline{\alpha}_\eta)} h_\eta^{\underline{\alpha}_\eta-1} e^{-\frac{h_\eta}{\underline{\beta}_\eta}} \\ &\propto h_\eta^{\frac{n-1}{2}} e^{-\frac{h_\eta}{2}\sum_{t=2}^{n}[(\alpha_t-m)-\rho(\alpha_{t-1}-m)]^2} h_\eta^{\underline{\alpha}_\eta-1} e^{-\frac{h_\eta}{\underline{\beta}_\eta}} \\ &= h_\eta^{\underline{\alpha}_\eta+\frac{n-1}{2}-1} e^{-h_\eta\left\{\frac{1}{\underline{\beta}_\eta}+\frac{1}{2}\sum_{t=2}^{n}[(\alpha_t-m)-\rho(\alpha_{t-1}-m)]^2\right\}} \\ &= h_\eta^{\overline{\alpha}_\eta-1} e^{-\frac{h_\eta}{\overline{\beta}_\eta}} \end{aligned} \tag{2.41}$$

where

$$\overline{\alpha}_\eta = \underline{\alpha}_\eta + \frac{n-1}{2} \tag{2.42}$$

and

$$\overline{\beta}_\eta = \frac{1}{\frac{1}{\underline{\beta}_\eta} + \frac{1}{2}\sum_{t=2}^{n}[(\alpha_t - m) - \rho(\alpha_{t-1} - m)]^2} \tag{2.43}$$

are the shape and the scale parameter of the Gamma full conditional distribution of h_η, respectively. And finally, from equations (2.23), (2.26), and (2.27), the following result can be calculated

$$\begin{aligned} p(\alpha_t|\theta, \alpha_{-t}, u) \propto & \quad \frac{1}{\left(\sqrt{2\pi\sigma_\varepsilon^2}\right)^{n-1}} e^{-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=2}^{n}(u_t - e^{\alpha_t}u_{t-1})^2} \\ & \times \frac{1}{\left(\sqrt{2\pi\sigma_\eta^2}\right)^{n-1}} e^{-\frac{1}{2\sigma_\eta^2}\sum_{t=2}^{n}[(\alpha_t - m) - \rho(\alpha_{t-1} - m)]^2} \\ \propto & \quad e^{-\frac{h_\varepsilon}{2}\sum_{t=2}^{n}(u_t - e^{\alpha_t}u_{t-1})^2} e^{-\frac{h_\eta}{2}\sum_{t=2}^{n}[(\alpha_t - m) - \rho(\alpha_{t-1} - m)]^2} \\ \propto & \quad e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t} - \frac{u_t}{u_{t-1}}\right)^2} e^{-\frac{h_\eta}{2}\sum_{t=2}^{n}[(\alpha_t - m) - \rho(\alpha_{t-1} - m)]^2} \end{aligned} \tag{2.44}$$

where for $t = 2, \ldots, n-1$

$$
\begin{aligned}
p(\alpha_t|\theta, \alpha_{-t}, u) \propto\ & e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} \\
& \times e^{-\frac{h_\eta}{2}\{[(\alpha_t-m)-\rho(\alpha_{t-1}-m)]^2+[(\alpha_{t+1}-m)-\rho(\alpha_t-m)]^2\}} \\
=\ & e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} \\
& \times e^{-\frac{h_\eta}{2}[(\alpha_t-m)^2-2(\alpha_t-m)\rho(\alpha_{t-1}-m)+\rho^2(\alpha_{t-1}-m)^2]} \\
& \times e^{-\frac{h_\eta}{2}[(\alpha_{t+1}-m)^2-2(\alpha_{t+1}-m)\rho(\alpha_t-m)+\rho^2(\alpha_t-m)^2]} \\
\propto\ & e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} \\
& \times e^{-\frac{h_\eta}{2}\{[(\alpha_t-m)^2(1+\rho^2)-2\rho(\alpha_t-m)[(\alpha_{t-1}-m)+(\alpha_{t+1}-m)]\}} \\
\propto\ & e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} \\
& \times e^{-\frac{h_\eta}{2}\{[(\alpha_t^2-2\alpha_t m)(1+\rho^2)-2\rho\alpha_t[(\alpha_{t-1}-m)+(\alpha_{t+1}-m)]\}} \\
=\ & e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} \\
& \times e^{-\frac{h_\eta}{2}\{\alpha_t^2(1+\rho^2)-2\alpha_t[m(1+\rho^2)+\rho(\alpha_{t-1}+\alpha_{t+1}-2m)]\}} \\
=\ & e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} \\
& \times e^{-\frac{h_\eta}{2}\{\alpha_t^2(1+\rho^2)-2\alpha_t[m(1+\rho^2-2\rho)+\rho(\alpha_{t-1}+\alpha_{t+1})]\}} \\
=\ & e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} \\
& \times e^{-\frac{h_\eta}{2}\{\alpha_t^2(1+\rho^2)-2\alpha_t[m(1-\rho)^2+\rho(\alpha_{t-1}+\alpha_{t+1})]\}} \qquad (2.45)
\end{aligned}
$$

and for $t = n$

$$\begin{aligned}
p(\alpha_t|\theta, \alpha_{-t}, u) \propto & \; e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} e^{-\frac{h_\eta}{2}[(\alpha_t-m)-\rho(\alpha_{t-1}-m)]^2} \\
= & \; e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} \\
& \times e^{-\frac{h_\eta}{2}[(\alpha_t-m)^2-2(\alpha_t-m)\rho(\alpha_{t-1}-m)+\rho^2(\alpha_{t-1}-m)^2]} \\
\propto & \; e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} e^{-\frac{h_\eta}{2}[\alpha_t^2-2\alpha_t m+m^2-2\alpha_t\rho(\alpha_{t-1}-m)]} \\
= & \; e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} e^{-\frac{h_\eta}{2}\{\alpha_t^2-2\alpha_t[m+\rho(\alpha_{t-1}-m)]\}}. \qquad (2.46)
\end{aligned}$$

Results from equations 2.45 and 2.46 are equal except for the terms in the exponents of the second exponential functions and can be expressed by one equation

$$p(\alpha_t|\theta, \alpha_{-t}, u) \propto e^{-\frac{h_\varepsilon u_{t-1}^2}{2}\left(e^{\alpha_t}-\frac{u_t}{u_{t-1}}\right)^2} e^{-\frac{h_\eta}{2}[\alpha_t^2\vartheta-2\alpha_t\tau]} \qquad (2.47)$$

with $\vartheta = 1+\rho^2$ and $\tau = m(1-\rho)^2+\rho(\alpha_{t-1}+\alpha_{t+1})$ for $t = 2, \ldots, n-1$ and $\vartheta = 1$ and $\tau = m(1-\rho)+\rho\alpha_{t-1}$ for $t = n$. Now the Gibbs sampler can be applied. The variables m and ρ as well as h_ε and h_η may be easily drawn from normal and Gamma distributions, respectively. In case of α_t, there is no standard density function available. As suggested in Yang and Leon-Gonzalez (2010), we use Independent Chain Metropolis-Hastings steps to sample α_t where we take candidate draws from the t-distribution with one degree of freedom. We assume the same prior parameter values as Yang and Leon-Gonzalez (2010), summarized in Table 2.2.

2.5 Diagnostics

To test for the separate null hypotheses of $m = 0$, $\rho = 0$, $\sigma_\eta^2 = 0$, and $\sigma_\varepsilon^2 = 0$, we make use of a central limit theorem addressing a function g of the parameters of interest θ given the data u implying

$$\frac{1}{NSE_1}\{\hat{g}_{S_1} - E[g(\theta)|u]\} \to N(0,1) \tag{2.48}$$

as S_1 goes to infinity where $\hat{g}_{S_1}$ estimates $g(\theta|u)$ by averaging over S_1 observations of the Markov chain. NSE_1 denotes the numerical standard error of $\hat{g}_{S_1}$ and is equal to $\sqrt{f_1(0)}/\sqrt{S_1}$ where $f_1(0)$ is the spectrum at frequency zero (see e.g. Koop, 2003, 65).

After running 100000 cycles, we drop the first $S_0 = 10000$ observations to ensure starting values have no effect anymore. The remaining $S_1 = 90000$ are used for estimation. Convergence is monitored by the convergence diagnostic (CD) introduced in Geweke (1992). That is, we test for the null hypothesis of equal means in the first 10% and the last 50% of the Markov chain by using the following test statistic

$$CD = \frac{\hat{g}_{S_A} - \hat{g}_{S_C}}{NSE_A + NSE_C} \to N(0,1) \tag{2.49}$$

as S_1 goes to infinity where g is a function of the parameters of interest θ. $\hat{g}_{S_A}$ and $\hat{g}_{S_C}$ are estimates of g corresponding to the first S_A and the last S_C observations, respectively, with $S_A = 9000$ and $S_C = 45000$ in the present case. Estimates are calculated as unweighted averages from the respective draws. $NSE_A = \sqrt{f_A(0)}/\sqrt{S_A}$ and $NSE_C = \sqrt{f_C(0)}/\sqrt{S_C}$ denote the numerical standard errors of $\hat{g}_{S_A}$ and $\hat{g}_{S_C}$, respectively, where the variances are estimated by $f_A(0)$ and $f_C(0)$, that is by the spectra at frequency zero to take account of serial correlation.

Parameter	Selected Prior Values	
m	$\underline{\mu}_m = \ln(0.9)$	$\underline{V}_m = 0.01$
ρ	$\underline{\mu}_\rho = 1$	$\underline{V}_\rho = 0.1$
h_η	$\underline{\alpha}_\eta = 1.5$	$\underline{\beta}_\eta = 2.5$
h_ε	$\underline{\alpha}_\varepsilon = 1.1$	$\underline{\beta}_\varepsilon = 0.2$

Table 2.2: Selected prior parameter values.

2.6 Results

Results for m, ρ, σ_η^2, and σ_ε^2 are reported in Table 2.3. The parameters of interest are estimated by the sample means of the respective Markov chains after discarding the first $S_0 = 10000$ cycles. To account for serial correlation, numerical standard errors given in parenthesis are calculated using the spectrum at frequency zero $f_1(0)$. Estimates for all countries and all parameters based on $S_1 = 90000$ observations are significantly different from zero at any reasonable level. As α_t in equation (2.2) is normally distributed with mean m and variance $\sigma_\alpha^2 = \sigma_\eta^2/(1-\rho^2)$, a_t in equation (2.1) follows the log-normal distribution with mean

$$E[a_t] = e^{m+\sigma_\alpha^2/2} \tag{2.50}$$

and variance

$$Var[a_t] = \left(e^{\sigma_\alpha^2} - 1\right) e^{2m+\sigma_\alpha^2}. \tag{2.51}$$

By using the estimates for m, ρ, and σ_η^2, estimates of the mean and the variance of the autoregressive coefficient can be derived, reported in columns 6 and 7 of Table 2.3. Obviously, the countries can be divided into two groups corresponding to the size of $Var[a_t]$. Australia, Canada, Finland, Japan, and UK showing moderate dispersion and a coefficient mean close to one are natural candidates either for a random walk or a STUR model with a stochastic unit root. For Swedish unemployment, a small coefficient variance is estimated as well with a coefficient mean probably smaller than one. Thus, AR(1) or STUR with a coefficient mean smaller than one are the relevant models. Brazil, Chile, and Mexico seem to be well represented by STUR models, again with a coefficient mean smaller than one. For the US, a STUR model with a stochastic unit root might be an adequate decision. Alternatively, we may

Country	m	ρ	σ_η^2	σ_ε^2	$E[a_t]$	$Var[a_t]$
Australia	-0.01607^{***}	0.21384^{***}	0.01407^{***}	0.08496^{***}	0.99134	0.01460
	(0.00033)	(0.00328)	(0.00004)	(0.00010)		
Brazil	-0.06626^{***}	0.38013^{***}	0.02665^{***}	0.24202^{***}	0.95058	0.02860
	(0.00086)	(0.00535)	(0.00015)	(0.00029)		
Canada	-0.02377^{***}	0.23436^{***}	0.01663^{***}	0.08555^{***}	0.98514	0.01723
	(0.00035)	(0.00358)	(0.00005)	(0.00008)		
Chile	-0.06097^{***}	0.55116^{***}	0.02997^{***}	0.17982^{***}	0.96132	0.04065
	(0.00091)	(0.00248)	(0.00015)	(0.00022)		
Finland	-0.02168^{***}	0.60667^{***}	0.00938^{***}	0.07759^{***}	0.98584	0.01453
	(0.00050)	(0.00292)	(0.00002)	(0.00014)		
Japan	-0.02764^{***}	0.23024^{***}	0.01705^{***}	0.06241^{***}	0.98153	0.01750
	(0.00053)	(0.00309)	(0.00006)	(0.00006)		
Mexico	-0.07218^{***}	0.32661^{***}	0.03459^{***}	0.10623^{***}	0.94855	0.03552
	(0.00106)	(0.00584)	(0.00020)	(0.00006)		
Sweden	-0.04316^{***}	0.17146^{***}	0.02050^{***}	0.18763^{***}	0.96792	0.02000
	(0.00050)	(0.00501)	(0.00008)	(0.00015)		
UK	-0.01265^{***}	0.27300^{***}	0.01175^{***}	0.06552^{***}	0.99372	0.01262
	(0.00038)	(0.00334)	(0.00004)	(0.00009)		
US	-0.01682^{***}	0.36532^{***}	0.01902^{***}	0.06886^{***}	0.99418	0.02193
	(0.00069)	(0.00429)	(0.00008)	(0.00007)		

*, **, *** denotes significance at the 10%, 5%, 1% level.

Table 2.3: Bayesian estimates of the STUR model.

sample 10000 values of both m, ρ and σ_η^2 from the simulated distributions and estimate the density of $E[a_t]$. Density plots related to the degree of dispersion are displayed in Figures 2.1 and 2.2, respectively. Clearly, the variance of a_t transmits to the variance of the mean of a_t. To control for convergence, CD statistics for all countries and parameters are calculated. The null hypothesis of sufficient convergence is not rejected at the 5% level of significance for any case. In three cases, the null is rejected at the 10% level. Finally, we

Country	m	ρ	σ_η^2	σ_ε^2
Australia	-0.16471	1.65892*	-0.32284	-0.14614
Brazil	-1.08425	0.37573	-0.43591	0.73211
Canada	1.65123*	-0.53225	0.89991	0.47659
Chile	-0.67347	-0.89615	-1.08460	0.50517
Finland	0.02720	0.17438	0.29621	0.83479
Japan	-0.50476	0.30368	0.73794	0.11676
Mexico	-0.59307	0.81786	-0.32016	-0.40073
Sweden	-1.41864	1.74546*	0.07555	-0.65490
UK	-0.62134	0.19063	-0.14414	0.58342
US	1.55418	-0.42295	1.13313	1.44847

*, **, *** denotes significance at the 10%, 5%, 1% level.

Table 2.4: CD statistics.

have a look at the residuals associated with some realizations of the coefficient process α_t. The Jarque-Bera test of the null hypothesis of normally distributed random variables rejects in many cases. Positive serial correlation is regularly present. Consider the situation where the errors follow an AR(1) process without intercept, that is

$$\varepsilon_t = \phi\varepsilon_{t-1} + \nu_t \tag{2.52}$$

with $\phi > 0$ and $\nu_t \sim N(0, \sigma_\nu^2)$. Using equation (2.52) and substituting $u_{t-1} - a_{t-1}u_{t-2}$ for ε_{t-1}, it follows from equation (2.1) that

$$u_t = (\phi + a_t)u_{t-1} - \phi a_{t-1}u_{t-2} + \nu_t \tag{2.53}$$

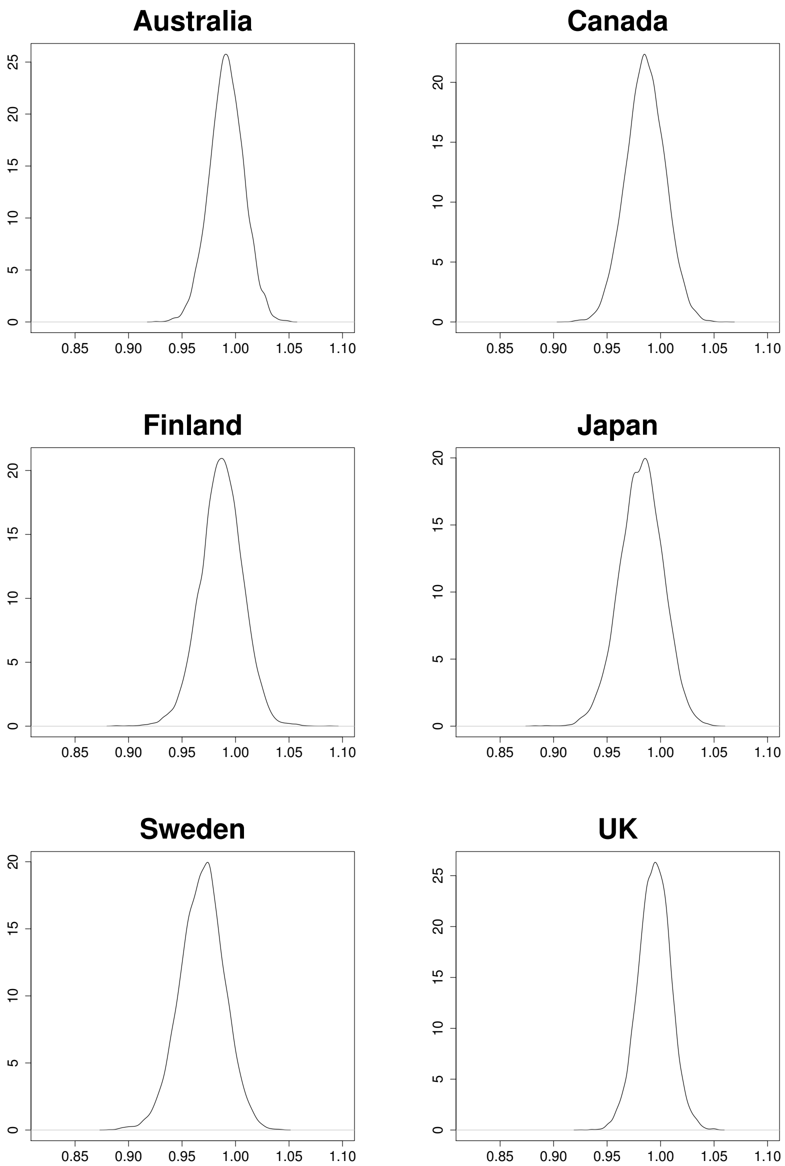

Figure 2.1: Density estimates of $E[a_t]$ with small variances.

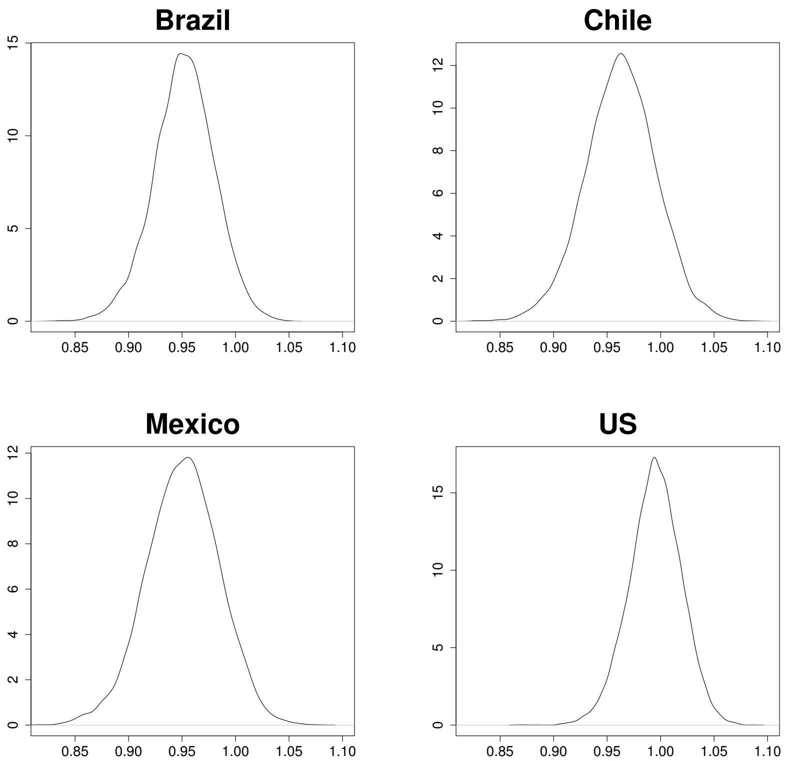

Figure 2.2: Density estimates of $E[a_t]$ with large variances.

and thus the mean of a_t in the STUR model is overestimated by the correlation of the errors over time. But note that another hysteresis effect enters one lag behind which is strictly negative. Particularly, it does not suffice to enter fixed-coefficient autoregressive terms of higher order to tackle serial correlation. Again we have to allow for random coefficients which makes the Metropolis-Hastings procedure grow enormously.

2.7 Alternative Estimation

To summarize, the Bayesian procedure elaborated in Section 2.4 is a very time-consuming one, thus to allow for extensive simulation experiments, faster alternative methods are in demand. The STUR test introduced in Holl (2013a) is based on an AR(1) model with autoregressive conditional heteroskedastic errors of order one (ARCH(1)) which is equivalent in first and second conditional moment to the STUR model in equations (2.1) and (2.2) given that the coefficient process a_t is approximated by an infinite-order Taylor series around its mean $E[a_t]$. That is, to obtain rough estimators for $E[a_t]$, $Var[a_t]$ and σ_ε^2, as suggested in Klüppelberg et al. (2002), it seems conceivable to maximize the log-pseudo-likelihood function (conditional on x_0) of the AR(1)-ARCH(1) model

$$x_t = \varphi x_{t-1} + \sigma_t e_t \tag{2.54}$$

for $t = 1, \dots, n$ where e_t follows a standardized normal distribution and denotes the value of $\hat{e}_t = (x_t - e^{m+\sigma_\alpha^2/2} x_{t-1})/\sigma_t$ when the parameters take their true values, $\varphi = E[a_t]$, $\sigma_t = \sqrt{\beta + \lambda x_{t-1}^2}$ with $\beta = \sigma_\varepsilon^2$ and $\lambda = Var[a_t]$. The ARCH part is driven by previous observations as introduced in Weiss (1984) rather than by previous innovations. See Holl (2013a) for details, particularly on the specific form of the log-pseudo-likelihood function.

As in Granger and Swanson (1997), we suggest to use an approximate method and check its accuracy by simulations. The procedure described above is restricted to provide estimates for $E[a_t]$, $Var[a_t]$ and σ_ε^2 which suffices, for instance, to calculate one-step forecasts by conditional means. Especially, in case that a_t is not correlated over time, i.e. $\rho = 0$, one can derive any h-step forecast by conditional means with these estimates. For every country 1000 replications of a STUR process are generated according to the estimates in Table 2.3. Then, for each replication the log-pseudo-likelihood of the corresponding AR(1)-ARCH(1) model is maximized to obtain the parameter estimates. Average parameter estimates and standard errors in parentheses are shown in Table 2.5 and can be compared to the Bayesian estimates in Table 2.3. For a great many of replications, this new estimator provides really satisfying results. Thus it seems to be worth studying its theoretical characteristics which we may leave to a future project.

2.8 Conclusions

In the present paper, hysteresis in unemployment is revisited. Ball (2009) defines hysteresis as the behavior of unemployment depending on the past history of its natural rate and the length of time that it is pushed away from it. This is to some extent reflected in the STUR model by Granger and Swanson (1997). We make use of a new unit-root test introduced in Holl (2013a) which has power against STUR processes. Moreover, a process generated by a STUR model cannot be differenced to stationarity and therefore the results may affect

Country	$E[a_t]$	$Var[a_t]$	σ_ε^2
Australia	0.98743	0.01388	0.08668
	(0.01765)	(0.00497)	(0.01399)
Brazil	0.95597	0.02731	0.24480
	(0.02596)	(0.01044)	(0.03264)
Canada	0.98089	0.01647	0.08654
	(0.02046)	(0.00617)	(0.01219)
Chile	0.98069	0.04119	0.18138
	(0.02829)	(0.01254)	(0.02429)
Finland	0.98985	0.01385	0.08416
	(0.01786)	(0.00370)	(0.09733)
Japan	0.97779	0.01649	0.06360
	(0.02016)	(0.00634)	(0.00951)
Mexico	0.95371	0.03416	0.10734
	(0.02894)	(0.01303)	(0.01448)
Sweden	0.96504	0.01862	0.18942
	(0.02099)	(0.00681)	(0.02422)
UK	0.99132	0.01210	0.06653
	(0.01673)	(0.00439)	(0.00996)
US	0.99477	0.02113	0.06979
	(0.01955)	(0.00570)	(0.01117)

Table 2.5: Alternative estimates of the STUR model.

methods beyond the univariate analysis, especially, the cointegration analysis. To evaluate the results of the test, we estimate STUR models by Bayesian techniques. As the stationary AR(1) and the random walk model are nested in the STUR model, we gain some broad insight into the test results and why the decisions have come up by simulating the mean of the random autoregressive coefficient.

To sum up, all STUR estimates are significantly different from zero. For Brazil, Canada, Chile, Finland, Mexico, and the US the null hypothesis of a unit root in unemployment is rejected at least at the 10% level of significance. In case of Brazil, Chile, Mexico, and the US, this result seems to be due to the large variance of the autoregressive coefficient. For the US, we may actually assume a stochastic unit root. Brazil, Chile, and Mexico are more frequently represented by stationary regimes associated with smaller coefficient means. Particularly, for the UK and Japan a random walk is suggested, as no test in Table 2.1 rejects the null. Furthermore, the Bayesian procedure results in small coefficient variances around the unit mean. The test results for Australia, Canada, Finland, and Sweden cannot be interpreted that easily.

We recommend to generate a large number of replications from competing models fitted to the data. Simulating the distributions of the test statistics may help to draw final conclusions. Evaluating predictive accuracy of the STUR model when compared to competing models can shed light on that issue as well. To do so, one can use an alternative estimation method based on maximizing the log-pseudo-likelihood using conditional mean and conditional variance of the STUR model which is much less time-consuming than the Bayesian technique. Simulation results are quite promising.

3 Forecasting Unemployment Using STUR Models – Evaluation of a new Unit-Root Test

3.1 Introduction

To evaluate the results of the tests of the null hypothesis of a random walk in OECD unemployment rates for ten countries in Holl (2013a), we generate a great many of replications of artificial data and derive the empirical distribution functions of the test statistics in use. Particularly, we calculate rejection frequencies for models considered in Holl (2013a) for each test and each country.

In Holl (2013a), test statistics from Dickey and Fuller (1979) (DF), Leybourne et al. (1996) (LMT), and Holl (2013a) (STUR) are calculated where LMT and STUR have power against stochastic unit-root (STUR) processes as defined in Granger and Swanson (1997). DF and STUR have power against AR(1), however, both tend to overreject in presence of a unit root with positive serial

Not submitted.

correlation. In contrast, LMT overrejects in presence of negative serial correlation. Thus, we consider four competing models which may have generated the real data, namely, the random walk, the AR(1), the ARIMA(0,1,1), and the STUR model. In simulations, we also make use of the augmented version of DF (ADF) which adresses the issue of positive serial correlation.

By evaluating the results of all tests, we decide which model is most likely to have generated the real data for each country individually. Then we confront the results with different forecast situations and look whether the preferred models of the unit-root tests correspond to good forecasts measured by mean squared forecast error. Plotting cumulative forecast errors for different countries and horizons illustrates potential systematic distortions resulting from a STUR model compared to its competitors. In an extended variety of combined forecasts, we allow different models to contribute different parts to a performant forecast.

We summarize the results of the unit-root tests in Holl (2013a) in Section 3.2. In Section 3.3, the respective models are introduced and the simulation procedure is explained. Size and power estimates resulting from the simulations are discussed. Forecasts by conditional means of the models presented in Section 3.3 are calculated in Section 3.4 for different forecast horizons. Forecast combinations resulting from different weighting methods are considered in Section 3.5. Whether STUR models are good 1-step forecast models at all, is checked by simulations elaborated in Section 3.7. And finally, in Section 3.8 we may draw some conclusions.

3.2 Unit-Root Tests

In Holl (2013a), results of the tests for the null of a unit root in 250 demeaned observations ranging from May 1989 to February 2010 are presented. DF, LMT, and STUR statistics are calculated for survey-based, seasonally-adjusted, monthly data on unemployment rates for the following countries: Australia, Brazil, Canada, Chile, Finland, Japan, Mexico, Sweden, UK, and US. There, the following test statistics are used:

1. *Dickey-Fuller and Augmented Dickey-Fuller test:*

   The null of a fixed unit root versus a stationary alternative is tested using the t-statistic of ψ of the regression

   $$\Delta u_t = \psi u_{t-1} + \sum_{i=1}^{p'} \varphi_i \Delta u_{t-i} + \epsilon_t. \tag{3.1}$$

   with $p' = 0$ in Holl (2013a). For size and power simulations, we also calculate the augmented version of DF (ADF) where the number of lagged differences is chosen according to Akaike information criterion (AIC) with the maximum number $p'_{max} = \lfloor 12(n/100)^{0.25} \rfloor - 1$ as suggested in Schwert (1989). Critical values at the 1%, 5%, 10% level from Fuller (1976, 373) are equal to -2.58, -1.95, -1.62.

2. *Leybourne-McCabe-Tremayne test:*

   Further, the LMT test of the null of a fixed versus the alternative of a stochastic unit root given by the statistic

   $$\hat{Z}_n = \frac{1}{n^{3/2}\hat{\sigma}_\epsilon^2 \hat{\kappa}^2} \sum_{t=3}^{n} \left(\sum_{j=2}^{t-1} \hat{\epsilon}_j \right)^2 \left(\hat{\epsilon}_j^2 - \hat{\sigma}_\epsilon^2 \right) \tag{3.2}$$

obtained from

$$\Delta u_t = \gamma_0 + \gamma_1 t + \epsilon_t \tag{3.3}$$

where $\hat{\sigma}_\epsilon^2 = (1/n)\sum_{t=2}^n \hat{\epsilon}_t^2$ and $\hat{\kappa}^2 = (1/n)\sum_{t=2}^n (\hat{\epsilon}_t^2 - \hat{\sigma}_\epsilon^2)^2$ is applied. Critical Values at the 1%, 5%, 10% level are equal to 0.289, 0.168, 0.122 (Granger and Swanson, 1997).

3. *STUR test:*

   Finally, the pseudo-likelihood ratio test introduced in Holl (2013a) for the null of a random walk versus alternatives nested in a STUR model is calculated using the deviance statistic

$$d_n = -2\left[\mathcal{L}_n(\hat{\bar{\alpha}}_{n,0}, \hat{\beta}_{n,0}, \hat{\lambda}_{n,0}) - \mathcal{L}_n(\hat{\bar{\alpha}}_{n,1}, \hat{\beta}_{n,1}, \hat{\lambda}_{n,1})\right] \tag{3.4}$$

   where $\mathcal{L}_n(\hat{\bar{\alpha}}_{n,0}, \hat{\beta}_{n,0}, \hat{\lambda}_{n,0})$ and $\mathcal{L}_n(\hat{\bar{\alpha}}_{n,1}, \hat{\beta}_{n,1}, \hat{\lambda}_{n,1})$ denote the maximum values of the log-pseudo-likelihood functions conditional on u_0 using conditional means and conditional variances of u_t under the null and under the alternative, respectively, where u_t is given by

$$u_t = e^{\bar{\alpha}} u_{t-1} + \sqrt{\beta + \lambda u_{t-1}^2}\epsilon_t. \tag{3.5}$$

   Under the null hypothesis, we have $\bar{\alpha} = 0$, $\beta > 0$, and $\lambda = 0$ whereas under the alternative hypothesis, $\bar{\alpha} \in \mathbb{R}$, $\beta > 0$, and $\lambda \geq 0$. Critical values at the 1%, 5%, 10% level from Holl (2013a) are equal to 8.150, 5.109, 3.814.

Country	DF	ADF	p'	LMT	STUR
Australia	-0.617	-0.900	13	0.303^{***}	0.382
Brazil	-2.279^{**}	-2.341^{**}	7	0.008	5.159^{**}
Canada	-1.023	-1.766^{*}	5	0.904^{***}	6.472^{**}
Chile	-2.063^{**}	-2.337^{**}	14	0.018	4.237^{*}
Finland	-2.355^{**}	-2.539^{**}	14	6.032^{***}	10.577^{***}
Japan	-1.088	-1.388	5	0.092	1.185
Mexico	-1.816^{*}	-2.728^{***}	6	0.610^{***}	15.288^{***}
Sweden	-1.906^{*}	-2.087^{**}	5	0.280^{**}	3.623
UK	-0.254	-1.517	6	-0.040	0.065
US	1.589	-3.130^{***}	7	0.113	8.809^{***}

*, **, ***: significant at the 10%, 5%, 1% level.

Table 3.1: Unit-root tests.

From Table 3.1 we see that the STUR test rejects the null of a unit root for Brazil, Canada, Finland, Mexico and the US at least at the 5% level of significance. For Canada, Finland, and Mexico, the LMT test corresponds to the STUR decision. ADF results and numbers of lagged differences p' are now added for simulations.

3.3 Size and Power Simulations

In this section, the parameters of the models considered in Holl (2013a) are estimated. By using estimates and, where necessary, starting values of the real series, artificial series of unemployment rates are generated for each model and each country.

In particular, the following models are estimated:

1. *Random walk or ARIMA(0,1,0) model:*

$$u_t = u_{t-1} + \varepsilon_t \tag{3.6}$$

where ε_t is white noise. To calculate size estimates, random walks have to be simulated as the random walk forms the null hypothesis of all four tests. $Var[\varepsilon_t] = \sigma_\varepsilon^2$ is estimated by a Kalman filter.

2. *ARIMA(0,1,1) model:*

$$u_t = u_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \tag{3.7}$$

with ε_t white noise. $\theta > 0$ and $\theta < 0$ correspond to negative and positive serial correlation of the errors, respectively. This process is taken account of as all tests have an issue with reaching the nominal size in case the true process is a random walk with serially correlated errors. θ is estimated by maximum likelihood via a state-space representation; $Var[\varepsilon_t] = \sigma_\varepsilon^2$ is again estimated by a Kalman filter.

3. *AR(1) model:*

$$u_t = au_{t-1} + \varepsilon_t \tag{3.8}$$

with a a constant over time and ε_t white noise. This is one of the two main competing processes under the alternatives of DF/ADF and STUR test. LMT does not have power against AR(1) even from a theoretical point of view (Holl, 2013a). Parameters a and $Var[\varepsilon_t] = \sigma_\varepsilon^2$ are estimated by ordinary least squares.

4. *STUR model:*

$$u_t = a_t u_{t-1} + \varepsilon_t \tag{3.9}$$

where $a_t = e^{\alpha_t}$ and

$$\alpha_t = \mu + \rho\alpha_{t-1} + \eta_t \tag{3.10}$$

with $|\rho| < 1$, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, $\eta_t \sim N(0, \sigma_\eta^2)$ and η_t independent of ε_t. As a consequence, α_t is normally distributed with mean $m = \mu/(1-\rho)$ and variance $\sigma_\alpha^2 = \sigma_\eta^2/(1-\rho^2)$. Thus, $a_t = e^{\alpha_t}$ is lognormally distributed with mean

$$E[a_t] = e^{m+\sigma_\alpha^2/2}. \tag{3.11}$$

DF/ADF, LMT as well as STUR test allow for a STUR process under the alternative. Parameters μ, ρ, σ_η^2 and σ_ε^2 are estimated in Holl (2013b) by means of a Bayesian procedure suggested in Jones and Marriott (1999).

Parameter estimates for all four models using 250 observations are shown in Table 3.2 with standard errors in parentheses (numerical standard errors for the Bayesian estimates). For STUR, all estimates are significant at the 1% level. Please note, that these results are based on a Bayesian procedure and cannot be compared to the results of frequentist's methods which have been used for random walk, ARIMA(0,1,1) and AR(1). The autoregressive

Country	RW	ARIMA(0,1,1)		AR(1)		STUR			
	σ_ε^2	θ	σ_ε^2	a	σ_ε^2	m	ρ	σ_η^2	σ_ε^2
Australia	0.03538	−0.07502	0.03510	0.99611***	0.03533	−0.01607***	0.21384***	0.01407***	0.08496***
		(0.05230)		(0.00629)		(0.00033)	(0.00328)	(0.00004)	(0.00010)
Brazil	0.21972	−0.09908*	0.21739	0.96666***	0.21521	−0.06626***	0.38013***	0.02665***	0.24202***
		(0.05982)		(0.01460)		(0.00086)	(0.00535)	(0.00015)	(0.00029)
Canada	0.04181	0.05493	0.04168	0.99183***	0.04163	−0.02377***	0.23436***	0.01663***	0.08555***
		(0.06476)		(0.00797)		(0.00035)	(0.00358)	(0.00005)	(0.00008)
Chile	0.17719	0.41134***	0.13306	0.96591***	0.17420	−0.06097***	0.55116***	0.02997***	0.17982***
		(0.04093)		(0.01649)		(0.00091)	(0.00248)	(0.00015)	(0.00022)
Finland	0.03843	0.70336***	0.01828	0.99202***	0.03759	−0.02168***	0.60667***	0.00938***	0.07759***
		(0.03654)		(0.00338)		(0.00050)	(0.00292)	(0.00002)	(0.00014)
Japan	0.01285	−0.04271	0.01283	0.99287***	0.01279	−0.02764***	0.23024***	0.01705***	0.06241***
		(0.06946)		(0.00654)		(0.00053)	(0.00309)	(0.00006)	(0.00006)
Mexico	0.07506	−0.18729***	0.07198	0.97015***	0.07408	−0.07218***	0.32661***	0.03459***	0.10623***
		(0.05534)		(0.01641)		(0.00106)	(0.00584)	(0.00020)	(0.00006)
Sweden	0.14414	−0.20035***	0.13777	0.97773***	0.14205	−0.04316***	0.17146***	0.02050***	0.18763***
		(0.05627)		(0.01166)		(0.00050)	(0.00501)	(0.00008)	(0.00015)
UK	0.01100	0.31585***	0.00937	0.99905***	0.01100	−0.01265***	0.27300***	0.01175***	0.06552***
		(0.04732)		(0.00372)		(0.00038)	(0.00334)	(0.00004)	(0.00009)
US	0.02438	0.09838*	0.02402	1.01233***	0.02413	−0.01682***	0.36532***	0.01902***	0.06886***
		(0.05165)		(0.00774)		(0.00069)	(0.00429)	(0.00008)	(0.00007)

*, **, *** denotes significance at the 10%, 5%, 1% level.

Table 3.2: Parameter estimates.

coefficients as well are significant at the 1% level for all countries. Clearly, the US coefficient estimate larger than one corresponds to an explosive process which is not plausible at least in the long run. However, as the STUR model allows for explosive regimes as well, the explosive AR(1) model is also used in simulations for the US data. Estimating an ARIMA(0,1,1) model reveals very nonhomogeneous patterns of serial correlation where we have moving-average coefficients significant at the 1% level for Chile, Finland, Mexico, Sweden and UK. Among these countries, only for Mexico and Sweden, there is a negative moving-average coefficient associated with positive serial correlation.

By running DF, ADF, LMT and STUR test on series generated according to the models fitted to the real data, size and power performance are evaluated. The optimum number of lagged differences for ADF, p', is selected for the real series of each country and kept the same for every replication (see Table 3.1). For each model, 250 observations are generated. 1000 series of a random walk are simulated using the starting value of the real series for each country to obtain size estimates. Errors are drawn from a normal distribution with mean zero and variance σ_ε^2 as shown in the second column of Table 3.2. To assess the extent of deviation from nominal size in case of a violated error assumption, 1000 series of an ARIMA(0,1,1) process are generated, again using the starting value of the real series, drawing errors ε_t from $N(0, \sigma_\varepsilon^2)$ where σ_ε^2 can be found in column 4 of Table 3.2. Power estimates are calculated for AR(1) and STUR process. For AR(1) 1000 series are generated where the first 100 observations of each series are dropped to obtain stationary series. This works for all countries except for the US where the autoregressive coefficient is larger than one. Errors follow an $N(0, \sigma_\varepsilon^2)$ distribution. Estimates are retrieved from columns 5 and 6 in Table 3.2. And finally, for STUR 1000 series are generated using the starting values of the real series of unemployment rates. Necessary parameter estimates are presented in the last four columns of Table 3.2.

The results of the size and power simulations are shown in Table 3.3. We may discuss the results for each country individually. For Australia, only LMT rejects the null hypothesis of a unit root, however at the 1% level of significance. On the one hand, if the true process were ARIMA(0,1,0), ARIMA(0,1,1) or AR(1), LMT should not reject. But on the other hand, if it were STUR, the STUR test, should reject as well. Thus, we may conclude that the LMT test rejects for different reasons. In case of Brazilian unemployment rates, DF, ADF and STUR reject at the 5% level of significance. Therefore, AR(1) and STUR seem to be feasible processes. For Canada, LMT and STUR come to a clear decision, i.e. the STUR model. We may exclude AR(1), as the DF test does not reject the null at any reasonable level. Chilean unemployment rates might be adequately represented either by an AR(1) or a STUR model. In case of Finland, all tests reject the null, especially LMT and STUR test at the 1% level of significance, therefore, we may expect STUR to be the adequate model. No test rejects, for the first time in this study, when applied to Japanese data. As a consequence, random walk and ARIMA(0,1,1) are the potential models. The negative moving-average coefficient associated with positive serial correlation in the third column of Table 3.2 is not significant at any reasonable level. For Mexico, the STUR model seems to be reasonable, particularly supported by LMT and STUR test. Sweden displays a similar-ambiguous pattern as Australia. Given that the underlying process were ARIMA(0,1,0) or ARIMA(0,1,1), DF and LMT should not reject the null. As LMT actually has no power against AR(1), we dismiss this process. If the true process were STUR, the STUR test should reject the null hypothesis.

Country	RW				ARIMA(0,1,1)				AR(1)				STUR			
	DF	ADF	LMT	STUR	DF	ADF	LMT	STUR	DF	ADF	LMT	STUR	DF	ADF	LMT	STUR
Australia	0.053	0.053	0.063	0.054	0.069	0.050	0.038	0.060	0.083	0.080	0.046	0.068	0.176	0.175	0.692	0.965
Brazil	0.060	0.054	0.052	0.051	0.067	0.066	0.046	0.053	0.664	0.547	0.024	0.484	0.753	0.815	0.585	1.000
Canada	0.062	0.057	0.041	0.057	0.044	0.053	0.076	0.053	0.144	0.135	0.054	0.103	0.258	0.310	0.679	0.970
Chile	0.039	0.034	0.050	0.046	0.005	0.035	0.146	0.073	0.661	0.436	0.025	0.477	0.360	0.549	0.887	1.000
Finland	0.053	0.059	0.041	0.063	0.089	0.051	0.167	0.164	0.143	0.123	0.061	0.106	0.304	0.354	0.768	1.000
Japan	0.051	0.058	0.065	0.058	0.056	0.057	0.048	0.060	0.136	0.120	0.065	0.083	0.358	0.414	0.651	0.984
Mexico	0.043	0.039	0.052	0.050	0.125	0.045	0.016	0.102	0.576	0.489	0.026	0.410	0.756	0.806	0.704	0.998
Sweden	0.048	0.045	0.051	0.053	0.068	0.046	0.016	0.065	0.418	0.364	0.038	0.286	0.742	0.703	0.473	0.998
UK	0.053	0.050	0.059	0.049	0.015	0.047	0.119	0.070	0.060	0.062	0.053	0.046	0.133	0.157	0.688	0.943
US	0.051	0.045	0.050	0.046	0.028	0.038	0.080	0.044	0.001	0.000	0.079	0.955	0.132	0.251	0.821	0.996

Table 3.3: Size and power estimates.

For the UK data, results are quite similar to the Japanese one's, random walk and ARIMA(0,1,1) are the preferred models. In contrast, the moving-average coefficient which is strongly significant has a positive sign for this country and is thus related to negative serial correlation. Finally, for the US the STUR test rejects the null at the 1% level of significance, hence AR(1) and STUR model come into consideration. As the autoregressive coefficient of AR(1) is larger than one, the fitted model is an explosive one where DF has no power against it.

Interpretations of the size and power estimates are summarized in Table 3.4. Apparently, for six out of ten countries STUR may be an adequate model, namely for Brazil, Canada, Chile, Finland, Mexico and the US. The stationary AR(1) model only works for Brazil and Chile, the difference-stationary models do so for Japan and the UK. To check whether those decisions hold in the context of forecasting, out-of-sample forecasts resulting from the respective models are calculated as part of the next section.

Country	RW	ARIMA(0,1,1)	AR(1)	STUR
Australia	☐	☐	☐	☐
Brazil	☐	☐	☑	☑
Canada	☐	☐	☐	☑
Chile	☐	☐	☑	☑
Finland	☐	☐	☐	☑
Japan	☑	☑	☐	☐
Mexico	☐	☐	☐	☑
Sweden	☐	☐	☐	☐
UK	☑	☑	☐	☐
US	☐	☐	☑	☑

Table 3.4: Potential models.

3.4 Forecasts by Conditional Means

Unit-root tests in Section 3.2 and estimations in Section 3.3 are based on 250 demeaned observations ranging from May 1989 to February 2010. Conditional means of the fitted models are now used to determine h-step forecasts, $h = \{1, 3, 12\}$, for the periods March 2010 to November 2011, May 2010 to November 2011 and February 2011 to November 2011, respectively, corresponding to numbers of predicted observations of 21, 19 and 10. That is, we consider one-month-ahead, one-quarter-ahead and one-year-ahead forecasts conditional on the information known in period $t-1$, denoted $\hat{u}_{t-1}(1) := \hat{E}[u_t|\mathcal{I}_{t-1}]$, $\hat{u}_{t-1}(3) := \hat{E}[u_{t+2}|\mathcal{I}_{t-1}]$ and $\hat{u}_{t-1}(12) := \hat{E}[u_{t+11}|\mathcal{I}_{t-1}]$, respectively, where $\mathcal{I}_{t-1}$ contains the information cumulated up to period $t-1$; the hat above the expectations operator indicates that estimates are substituted for unknown parameters and variables. As the Bayesian procedure to estimate STUR models applied in Holl (2013b) is a very time-consuming one, in a first scenario

STUR models are not estimated for each and every information set. Thus, we assume parameters to be fixed and equal to the parameters associated with the first 250 observations estimated from 100000 observations of a Markov chain where the first 10000 are dropped as in Holl (2013b). To ensure a fair forecast competition, the same holds for parameter estimates of random walk, ARIMA(0,1,1) and AR(1). Let's go through the individual forecasting models:

1. *Random walk or ARIMA(0,1,0) model:*

$$u_{t+h-1} = u_{t+h-2} + \varepsilon_{t+h-1} \tag{3.12}$$

   is to be forecast where we obtain by repeated substitution

$$\hat{u}_{t-1}(h) = E[u_{t+h-1}|\mathcal{I}_{t-1}] = u_{t-1}. \tag{3.13}$$

2. *ARIMA(0,1,1) model:*

$$u_{t+h-1} = u_{t+h-2} + \varepsilon_{t+h-1} + \theta\varepsilon_{t+h-2} \tag{3.14}$$

   is required which is equal to equation (3.12) except for the correlated error term. In period $t-1$, ε_{t-1} already exists but cannot be observed and thus has to be estimated. Again, by repeated substitution we find

$$E[u_{t+h-1}|\mathcal{I}_{t-1}] = u_{t-1} + \theta\varepsilon_{t-1} \tag{3.15}$$

where θ and ε_{t-1} are unobserved and therefore have to be replaced by estimates. Particularly, ε_{t-1} is estimated by a Kalman filter. And we obtain

$$\hat{u}_{t-1}(h) = u_{t-1} + \hat{\theta}\hat{\varepsilon}_{t-1}. \tag{3.16}$$

As we want to keep the estimated parts constant, $\hat{\theta}\hat{\varepsilon}_{t-1}$ from $\hat{u}_{t-1}(h)$ where $t-1 = 250$ is stored to be added up to any u_{t-1} related to an information set $\mathcal{I}_{t-1}$ with $t-1 \in \{251, 252, \ldots, 270\}$. Particularly, forecasts from ARIMA(0,1,1) correspond to model-free forecasts from single exponential smoothing where the smoothing parameter is equal to $1+\theta$. However, the smoothing parameter is restricted to the interval $(0,1)$ and thus this holds only for $-1 > \theta < 0$ which is associated with positive serial correlation.

3. *AR(1) model:*

$$u_{t+h-1} = au_{t+h-2} + \varepsilon_{t+h-1} \tag{3.17}$$

is the variable to be estimated. Substituting repeatedly and taking conditional means results in

$$E[u_{t+h-1}|\mathcal{I}_{t-1}] = a^h u_{t-1} \tag{3.18}$$

where the estimate $\hat{a}$ is substituted for the unknown parameter a to obtain

$$\hat{u}_{t-1}(h) = \hat{a}^h u_{t-1}. \tag{3.19}$$

4. *STUR model:*

   Last but not least, the h-step forecast from a STUR model is derived. Quite similar to the AR(1) model u_{t+h-1} amounts to

$$u_{t+h-1} = a_{t+h-1}u_{t+h-2} + \varepsilon_{t+h-1} \tag{3.20}$$

   where now the autoregressive coefficient is not constant over time any more. By repeated substitution, we obtain

$$\begin{aligned} u_{t+h-1} = \quad & a_{t+h-1}a_{t+h-2}\dots a_t u_{t-1} \\ & +\varepsilon_{t+h-1} \\ & +a_{t+h-1}\varepsilon_{t+h-2} \\ & +a_{t+h-1}a_{t+h-2}\varepsilon_{t+h-3} \\ & \vdots \\ & +a_{t+h-1}a_{t+h-2}\dots a_{t+1}\varepsilon_t \end{aligned} \tag{3.21}$$

where we may take conditional expectations of equation (3.21) to obtain

$$\begin{aligned}\hat{u}_{t-1}(h) = E[u_{t+h-1}|\mathcal{I}_{t-1}] = \quad & E[a_{t+h-1}a_{t+h-2}\dots a_t u_{t-1}|\mathcal{I}_{t-1}] \\ & +E[\varepsilon_{t+h-1}|\mathcal{I}_{t-1}] \\ & +E[a_{t+h-1}\varepsilon_{t+h-2}|\mathcal{I}_{t-1}] \\ & +E[a_{t+h-1}a_{t+h-2}\varepsilon_{t+h-3}|\mathcal{I}_{t-1}] \\ & \vdots \\ & +E[a_{t+h-1}a_{t+h-2}\dots a_{t+1}\varepsilon_t|\mathcal{I}_{t-1}] \end{aligned} \tag{3.22}$$

α_t given in equation 3.10 follows a stationary AR(1) process and thus has an infinite-order moving-average representation

$$\alpha_t = \frac{\mu}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \eta_{t-j} \tag{3.23}$$

and therefore, $a_t = e^{\alpha_t}$ is a function of η_{t-j} for $j = 0, 1, 2, \dots$ Moreover, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, η_t is assumed to be independent of ε_t and u_{t-1} is known and thus non-random. Hence, equation (3.22) reduces to

$$E[u_{t+h-1}|\mathcal{I}_{t-1}] = \quad E[a_{t+h-1}a_{t+h-2}\dots a_t|\mathcal{I}_{t-1}]u_{t-1} \tag{3.24}$$

where information up to period $t-1$ does not simplify calculating the mean of the term $a_{t+h-1}a_{t+h-2}\dots a_t$, so we may look for its unconditional mean. As $a_t = e^{\alpha_t}$, $a_{t+h-1}a_{t+h-2}\dots a_t$ can be written as

$$a_{t+h-1}a_{t+h-2}\dots a_t = e^{S_{\alpha_t}(h)} \tag{3.25}$$

where $S_{\alpha_t}(h) = \sum_{i=0}^{h-1} \alpha_{t+i}$ with $S_{\alpha_t}(0) = 0$ by definition. α_t is normally distributed with mean m and variance σ_α^2. As $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_t)'$ may be represented by a linear transformation of a vector of independent standardized normal variables corresponding to Wold decomposition, it is multivariate normally distributed. $S_{\alpha_t}(h)$ as a sum of h multivariate normally distributed random variables is normally distributed as well. For its mean it holds that

$$E[S_{\alpha_t}(h)] = hm \tag{3.26}$$

and for its variance, we have to take account of the correlation between the random variables. The variance of a sum of correlated random variables is equal to the sum of covariances between every two individual random variables

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov[X_i, X_j] \tag{3.27}$$

where $Var[X_i] = Cov[X_i, X_j]$ for $i = j$ and $Cov[X_i, X_j] = Cov[X_j, X_i]$. Considering the h random variables $\alpha_t, \alpha_{t+1}, \dots, \alpha_{t+h-1}$, there are $h-1$ pairs having a distance of one period, $h-2$ pairs with a distance of two, and so on, with finally one pair having a distance of $h-1$. Furthermore, we may substitute $Corr[X_i, X_j]\sqrt{Var[X_i]}\sqrt{Var[X_j]}$ for $Cov[X_i, X_j]$ to obtain the variance of $S_{\alpha_t}(h)$

$$Var[S_{\alpha_t}(h)] = h\sigma_\alpha^2 + 2\sigma_\alpha^2 \sum_{r=1}^{h-1}(h-r)\rho_\alpha(r) \tag{3.28}$$

as $Var[a_{t+h-1}] = Var[a_{t+h-2}] = \dots = Var[a_t] = \sigma_\alpha^2$ and where $\rho_\alpha(r)$ denotes the correlation between random variables facing a distance in time equal to r. For the autoregressive model of order one in equation (3.10), $\rho_\alpha(r) = \rho^r$. That is, $S_{\alpha_t}(h)$ is normally distributed with mean $E[S_{\alpha_t}(h)]$

and variance $Var[S_{\alpha_t}(h)]$ given in equations (3.26) and (3.28). Thus, $e^{S_{\alpha_t}(h)}$ in equation (3.25) is log-normally distributed with mean

$$E\left[e^{S_{\alpha_t}(h)}\right] = e^{E\left[S_{\alpha_t}(h)\right]+\frac{1}{2}Var\left[S_{\alpha_t}(h)\right]} \tag{3.29}$$

and therefore the h-step forecast in equation (3.24) results in

$$E[u_{t+h-1}|\mathcal{I}_{t-1}] = e^{hm+\frac{\sigma_\alpha^2}{2}\left[h+2\sum_{r=1}^{h-1}(h-r)\rho^r\right]}u_{t-1} \tag{3.30}$$

where $\sigma_\alpha^2 = \sigma_\eta^2/(1-\rho)$; estimates for m, σ_η^2 and ρ are provided in Table 3.2.

We are now ready to evaluate predictive accuracy of the different forecasting models. Please note, that the unit-root tests in Section 3.2 as well as the estimations in Section 3.3 are applied to demeaned data. That is, the mean of the first 250 observations is subtracted from every single observation 1 to 250. Demeaning the data using the mean of all 271 observations would imply adding future information to the respective information set $\mathcal{I}_{t-1}$. Anyway, the last 21 observations have to be demeaned as well and so we continue subtracting the mean related to observations 1 to 250 from observations 251 to 271. Particularly, this procedure corresponds to our practice of using fixed parameters for periods 251 to 271.

Finally, before we switch over to the results of forecast evaluation, the measure of predictive accuracy is introduced. Mean squared forecast errors defined by

$$MSFE = \frac{1}{n_f}\sum_{t=251}^{271-h+1}[u_{t+h-1} - \hat{u}_{t-1}(h)]^2 \tag{3.31}$$

are used where $n_f = 20 - h + 2$ denotes the number of forecast observations to be evaluated which, in this study, is equal to the total number of forecast observations, i.e. $n_f = 21, 19, 10$ for 1-step, 3-step, 12-step forecasts, respectively.

Country	RW		ARIMA(0,1,1)		AR(1)		STUR	
Australia	0.01667	2	0.01664	1	0.01672	3	0.01692	4
Brazil	0.03286	1	0.03960	2	0.06693	3	0.09421	4
Canada	0.01381	2	0.01349	1	0.01394	3	0.01409	4
Chile	0.07000	3	0.08130	4	0.06962	1	0.06969	2
Finland	0.01048	1	0.01268	3	0.01231	2	0.01399	4
Japan	0.03048	3	0.03050	4	0.03007	2	0.02966	1
Mexico	0.05619	1	0.05882	3	0.05712	2	0.06105	4
Sweden	0.09381	4	0.09293	3	0.08666	2	0.08492	1
UK	0.01048	1	0.01167	4	0.01052	2	0.01079	3
US	0.02952	2	0.02981	3	0.03608	4	0.02785	1
Rank Sum		20		28		24		28

Table 3.5: MSFE for 1-step forecasts.

Country	RW		ARIMA(0,1,1)		AR(1)		STUR	
Australia	0.02421	2	0.02400	1	0.02521	3	0.02631	4
Brazil	0.07316	1	0.08699	2	0.32017	3	0.40163	4
Canada	0.04842	2	0.04714	1	0.05047	3	0.05150	4
Chile	0.25789	3	0.27995	4	0.25045	1	0.25195	2
Finland	0.04526	1	0.05050	2	0.05920	4	0.05623	3
Japan	0.04842	3	0.04854	4	0.04402	2	0.04029	1
Mexico	0.12263	1	0.12904	3	0.12501	2	0.13846	4
Sweden	0.10105	4	0.09825	3	0.05030	2	0.04263	1
UK	0.04684	1	0.05005	4	0.04723	2	0.04844	3
US	0.09053	1	0.09139	2	0.15356	4	0.09224	3
Rank Sum		19		26		26		29

Table 3.6: MSFE for 3-step forecasts.

Country	RW		ARIMA(0,1,1)		AR(1)		STUR	
Australia	0.07200	2	0.07042	1	0.09843	3	0.10889	4
Brazil	0.64500	1	0.69769	2	3.50238	3	3.52547	4
Canada	0.35400	2	0.34821	1	0.37675	3	0.38211	4
Chile	1.08200	2	1.17281	4	1.12248	3	1.08079	1
Finland	0.32600	2	0.34288	3	0.49175	4	0.25617	1
Japan	0.34000	3	0.34062	4	0.23413	2	0.16052	1
Mexico	0.08400	1	0.09484	2	0.27031	3	0.35188	4
Sweden	0.97200	4	0.96021	3	0.18347	2	0.10353	1
UK	0.13000	1	0.13776	3	0.13404	2	0.13917	4
US	0.47900	1	0.48280	2	1.69838	4	0.87347	3
Rank Sum		19		25		29		27

Table 3.7: MSFE for 12-step forecasts.

Let's start with mean squared forecast errors calculated for one-month-ahead forecasts. Results are presented in Table 3.5. Ranks for each country are given next to mean squared forecast errors where 1 corresponds to the smallest MSFE. For Japan, Sweden and the US, the STUR model performs best. Particularly, in case of Japan, we are somewhat surprised given the results from unit-root testing where a fixed-coefficient difference-stationary model is suggested (see Table 3.4). There, for the US AR(1) and STUR are recommended where now forecasting strongly prefers STUR. Moreover, the ambiguous situations for Sweden and Australia are now decided in favor of STUR and ARIMA(0,1,1), respectively. The best forecasts for Brazil, Finland, Mexico and the UK are calculated from a random walk model which was only expected for the UK. For Canada, ARIMA(0,1,1) wins the contest, again somewhat surprising. Especially for Chile, expecting AR(1) or STUR model to be adequate, forecasting reveals a consistent result.

In one-quarter-ahead forecasts, summarized in Table 3.6, STUR performs best for Japan and Sweden as in one-month-ahead forecasts. However, now the US data are better forecast by random walk as well as the Brazilian, Finnish, Mexican and UK data. Again, ARIMA(0,1,1) seems to be adequate for Australia and Canada and AR(1) still works for Chile.

Finally, mean squared forecast errors corresponding to one-year-ahead forecasts are presented in Table 3.7. The STUR model generates the best forecasts for Chile, Finland, Japan and Sweden showing up with a great performance especially for the latter three countries. Quite similarly, the random walk model performs best for Brazil, Mexico, UK and US. ARIMA(0,1,1) is still limited to Australian and Canadian unemployment rates and AR(1) does not succeed in providing best forecasts for any country any more.

Preparing these forecasts, all 271 observations have been demeaned by the mean of the first 250 observations implying the assumption of a constant mean over the whole sample. Assume, for instance, that the mean of observations 251 to 271 is underestimated by the mean of the first 250 observations. That is, the mean after observation 250 could increase in time or shift to a higher level. Then, forecasts generated from the AR(1) model based on observations 1 to 250 are going to underestimate observations 251 to 271. In case of trending or shifting means the nonstationary models are likely to be favored over the stationary one. To overcome this difficulty, we may demean the observations of a certain information set $\mathcal{I}_{t-1}$ using the mean of all observations up to $t-1$. Summing up, $u_1, u_2, \ldots, u_{t-1}$ are demeaned by

$$\bar{u}_{t-1} = \frac{1}{t-1} \sum_{i=1}^{t-1} u_i \tag{3.32}$$

for $t = 251, \ldots, 271$. By this procedure, no future information is used in forecasting and, in particular, no present information appearing in realized series of unemployment rates is destroyed. Mean squared forecast errors from forecasts using data demeaned in such a way are presented in Tables 3.8, 3.9 and 3.10.

What are now the consequences of different demeaning? For 1-step forecasts using ARIMA(0,1,1), the rank of Australia changes from 1 to 4, but note that the mean squared forecast errors of all models are very close to each other. Using AR(1), the rank of Australia changes from 3 to 1. If 1-step forecasts are generated by the STUR model and applied to the data demeaned by changing means, its relative performance improves associated with rank 3 rather than rank 4. In case of 3-step forecasts, the random walk model now seems to be more appropriate for Australia, the rank is 1 instead of 2. However, forecast performance of Mexico becomes worse with rank equal to 2 being 1 before. The same holds for Mexico using ARIMA(0,1,1) as well as for Australia with ranks changing from 3 to 4 and 1 to 2, respectively. Mexico improves from rank 2 to 1 when using the AR(1) model and from rank 4 to 3 when using the STUR model. Doing 12-step forecasts, ranks are not assigned in a different way from Table 3.7. As we would like to address the forecast performance of the AR(1) model, two results have to be emphasized again. For a forecast horizon equal to one, performance for Australia improves, for a horizon equal to 3, performance for Mexico improves. In both cases, the STUR model benefits from new demeaning as well. Maybe this points to the need for a separate handling of level shifts when using STUR models.

Country	RW		ARIMA(0,1,1)		AR(1)		STUR	
Australia	0.01658	2	0.01663	4	0.01653	1	0.01661	3
Brazil	0.03071	1	0.03650	2	0.05965	3	0.08365	4
Canada	0.01366	2	0.01336	1	0.01377	3	0.01389	4
Chile	0.06961	3	0.08065	4	0.06900	1	0.06903	2
Finland	0.00968	1	0.01169	3	0.01126	2	0.01274	4
Japan	0.03049	3	0.03052	4	0.03005	2	0.02956	1
Mexico	0.05601	1	0.05928	3	0.05618	2	0.05933	4
Sweden	0.09474	4	0.09377	3	0.08695	2	0.08483	1
UK	0.01028	1	0.01130	4	0.01030	2	0.01051	3
US	0.03081	2	0.03117	3	0.03818	4	0.02865	1
Rank Sum		20		31		22		27

Table 3.8: MSFE for 1-step forecasts allowing for changing means.

Country	RW		ARIMA(0,1,1)		AR(1)		STUR	
Australia	0.02360	1	0.02363	2	0.02378	3	0.02433	4
Brazil	0.05978	1	0.07075	2	0.26679	3	0.33797	4
Canada	0.04690	2	0.04568	1	0.04871	3	0.04963	4
Chile	0.25555	3	0.27677	4	0.24648	1	0.24876	2
Finland	0.03916	1	0.04380	2	0.05106	4	0.04850	3
Japan	0.05037	3	0.05050	4	0.04567	2	0.04153	1
Mexico	0.12368	2	0.13204	4	0.11992	1	0.13013	3
Sweden	0.11220	4	0.10912	3	0.05577	2	0.04579	1
UK	0.04493	1	0.04765	4	0.04522	2	0.04615	3
US	0.10379	1	0.10489	2	0.17496	4	0.10586	3
Rank Sum		19		28		25		28

Table 3.9: MSFE for 3-step forecasts allowing for changing means.

Country	RW		ARIMA(0,1,1)		AR(1)		STUR	
Australia	0.05529	2	0.05468	1	0.06840	3	0.07476	4
Brazil	0.42192	1	0.46301	2	2.84286	3	2.86325	4
Canada	0.32609	2	0.32057	1	0.34712	3	0.35209	4
Chile	1.02669	2	1.11399	4	1.05757	3	1.02665	1
Finland	0.24401	2	0.25853	3	0.38542	4	0.18602	1
Japan	0.39008	3	0.39076	4	0.27710	2	0.19675	1
Mexico	0.10121	1	0.11982	2	0.20165	3	0.26769	4
Sweden	1.14071	4	1.12783	3	0.25685	2	0.15179	1
UK	0.11462	1	0.12050	3	0.11746	2	0.12115	4
US	0.71076	1	0.71549	2	2.09820	4	1.17589	3
Rank Sum		19		25		29		27

Table 3.10: MSFE for 12-step forecasts allowing for changing means.

Finally, parameter estimates are calculated for each information set with STUR estimates obtained from 10000 rather than from 100000 cycles of the Markov chain Monte Carlo procedure. Again, the first 10% of the observations are dropped. For random walk, ARIMA(0,1,1) and AR(1) this corresponds to a more efficient use of information which holds for STUR too. STUR however is now accompanied by an additional estimation error resulting from possibly insufficient convergence in Bayesian estimation. The standard-normally distributed convergence diagnostic (CD) introduced in Geweke (1992) which is used to ensure convergence ranges from -7.665 to 5.317. Means corresponding to the information sets at hand are subtracted from the observations before estimating. Thus, results can be compared to the changing-means scenario above. Mean sqared forecast errors are presented in Tables 3.11 to 3.13.

The forecast performance of ARIMA(0,1,1) strongly improves and is responsible for most of the rank adjustments occuring to the competing models.

Calculating separate forecasts for every information set affects the medium-run and long-run performance of STUR. One-quarter-ahead forecasts from the STUR model are now superior for the US. For Brazil and UK, one-year-ahead forecasts improve, however starting from a weak initial position. In the long run, STUR seems to outperform AR(1) which is indicated by a rank sum equal to 26 for STUR compared to 33 for AR(1).

Country	RW		ARIMA(0,1,1)		AR(1)		STUR	
Australia	0.01658	2	0.01616	1	0.01659	3	0.01689	4
Brazil	0.03071	1	0.03180	2	0.05338	3	0.07488	4
Canada	0.01366	2	0.01352	1	0.01377	3	0.01383	4
Chile	0.06961	3	0.07623	4	0.06906	1	0.06943	2
Finland	0.00968	1	0.01168	3	0.01119	2	0.01224	4
Japan	0.03049	4	0.03048	3	0.03007	2	0.02968	1
Mexico	0.05601	2	0.05424	1	0.05654	3	0.05836	4
Sweden	0.09474	4	0.08073	1	0.08714	3	0.08470	2
UK	0.01028	1	0.01079	4	0.01031	2	0.01047	3
US	0.03081	3	0.03010	2	0.03467	4	0.02873	1
Rank Sum		23		22		26		29

Table 3.11: MSFE for 1-step forecasts with adjusted estimates.

Country	RW		ARIMA(0,1,1)		AR(1)		STUR	
Australia	0.02360	2	0.02304	1	0.02400	3	0.02419	4
Brazil	0.05978	1	0.06186	2	0.22630	3	0.28848	4
Canada	0.04690	1	0.04704	2	0.04870	3	0.04920	4
Chile	0.25555	3	0.26048	4	0.24691	1	0.24751	2
Finland	0.03916	2	0.03722	1	0.05057	4	0.04547	3
Japan	0.05037	4	0.05006	3	0.04579	2	0.04148	1
Mexico	0.12368	4	0.11572	1	0.12183	2	0.12233	3
Sweden	0.11220	3	0.11575	4	0.05448	2	0.04423	1
UK	0.04493	2	0.04149	1	0.04535	3	0.04588	4
US	0.10379	2	0.10465	3	0.14723	4	0.09218	1
Rank Sum		24		22		27		27

Table 3.12: MSFE for 3-step forecasts with adjusted estimates.

Country	RW		ARIMA(0,1,1)		AR(1)		STUR	
Australia	0.05529	2	0.05329	1	0.06947	3	0.07640	4
Brazil	0.42192	1	0.43794	2	2.61985	4	2.35464	3
Canada	0.32609	2	0.32397	1	0.34705	3	0.34810	4
Chile	1.02669	3	0.96364	1	1.05940	4	1.02239	2
Finland	0.24401	3	0.23042	2	0.38142	4	0.15998	1
Japan	0.39008	4	0.38934	3	0.28649	2	0.18445	1
Mexico	0.10121	2	0.09113	1	0.18543	3	0.21055	4
Sweden	1.14071	3	1.16120	4	0.24345	2	0.10908	1
UK	0.11462	2	0.11369	1	0.11871	4	0.11673	3
US	0.71076	1	0.71612	2	1.74998	4	0.74617	3
Rank Sum		23		18		33		26

Table 3.13: MSFE for 12-step forecasts with adjusted estimates.

3.5 Forecast Combinations

We evaluate the results from combined forecasts, again starting with the scenario where the estimates for observations 1 to 250 are kept constant over the forecast period. As constant estimates bring multicollinearity into the design matrix consisting of the forecast values, combination strategies based on regression techniques cannot be applied. Thus, models are arranged in three different groups, namely a group made up of all models, a constant-coefficient group containing random walk, ARIMA(0,1,1) and AR(1) and a nonstationary group consisting of random walk, ARIMA(0,1,1) and STUR. For AR(1) applied to US data, an explosive model is suggested which is treated as part of the constant-coefficient group like the stationary models. Combined forecasts with weights resulting from inverse mean sqared forecast errors are considered. Mean squared forecast errors correspond to the evaluation periods predicting observations 230 to 250 for 1-step, 232 to 250 for 3-step and 241 to 250 for 12-step forecasts. Parameters are estimated for observations 1 to 229 and again kept constant over the whole evaluation period. Group forecasts using inverse-MSFE weights are calculated for data demeaned by updated means with results presented in Tables 3.14 to 3.16 where ranks are given next to mean squared forecast errors and overall ranks, i.e. compared to all single and combined forecasts, in parentheses.

Using an weighted average of forecasts generated by each individual model results in one-month-ahead combined forecasts for Australia, Chile, Japan, Mexico, Sweden and the UK outperforming the competing forecast combinations. However, when this is compared to the individual forecasts it is only superior for Australia and Mexico. The group consisting of nonstationary models performs quite similarly. In this case, the rank sum is a bit misleading given that some mean squared forecast errors of both groups are very close to each other. Compared to individual forecasts, this combination is not superior for any country. Particularly, disregarding STUR seems to affect the performance of the constant-coefficient combination negatively whereas ignoring AR(1) for the nonstationary combination has not the same effect.

In one-quarter-ahead forecasts, rank sums approximate to each other. For Australia and Mexico the constant-coefficient combination and the combination including all models, respectively, are able to outperform the individual forecasts.

Considering one-year-ahead forecasts, for the first time the all-together combination is showing up with some deficits. Either ignoring STUR or ignoring AR(1) improves performance, suggesting that the difference-stationary models are more adequate to forecast at this horizon. For Mexico the combination made up of nonstationary models succeeds even when compared to individual forecasts.

Finally, combined forecasts are calculated using regularly updated parameter estimates for every information set. Weights are obtained from three different procedures. First, following Granger and Ramanathan (1984) we regress realized values on values of different forecasts without an intercept. Parameters are estimated by ordinary least squares. Fitted values serve as combined forecasts. Second, we make us of the forecast encompassing test introduced in Harvey and Newbold (2000) where the forecast error

$$\hat{e}_t(h) = u_t - \hat{u}_{t-h}(h) \tag{3.33}$$

for a certain model is regressed on its differences to the forecast errors resulting from the competing models. That is, following regressions are estimated

$$\hat{e}_t^{(1)} = \beta_1(\hat{e}_t^{(1)} - \hat{e}_t^{(2)}) + \beta_2(\hat{e}_t^{(1)} - \hat{e}_t^{(3)}) + \beta_3(\hat{e}_t^{(1)} - \hat{e}_t^{(4)}) + \epsilon_t^{(1)} \tag{3.34}$$

$$\hat{e}_t^{(2)} = \beta_1(\hat{e}_t^{(2)} - \hat{e}_t^{(1)}) + \beta_2(\hat{e}_t^{(2)} - \hat{e}_t^{(3)}) + \beta_3(\hat{e}_t^{(2)} - \hat{e}_t^{(4)}) + \epsilon_t^{(2)} \tag{3.35}$$

$$\hat{e}_t^{(3)} = \beta_1(\hat{e}_t^{(3)} - \hat{e}_t^{(1)}) + \beta_2(\hat{e}_t^{(3)} - \hat{e}_t^{(2)}) + \beta_3(\hat{e}_t^{(3)} - \hat{e}_t^{(4)}) + \epsilon_t^{(3)} \tag{3.36}$$

$$\hat{e}_t^{(4)} = \beta_1(\hat{e}_t^{(4)} - \hat{e}_t^{(1)}) + \beta_2(\hat{e}_t^{(4)} - \hat{e}_t^{(2)}) + \beta_3(\hat{e}_t^{(4)} - \hat{e}_t^{(3)}) + \epsilon_t^{(4)} \tag{3.37}$$

where the symbol indicating *function of* h is omitted for convenience, superscripts (1), (2), (3) and (4) denote random walk, ARIMA(0,1,1), AR(1) and STUR model, respectively. A certain model m is said to forecast-encompass its rivals if the F-statistic of the regression with dependent variable $\hat{e}_t^{(m)}$ is not significant at a specific level. The combined forecast results from an unweighted average of forecasts from all forecast-encompassing models. In case that there are no such forecast-encompassing models, an unweighted average over all models is calculated. And third, combined forecasts are obtained with weights corresponding to inverse mean squared forecast errors. Results are shown in Tables 3.17, 3.18 and 3.19 where rank averages are associated with ties.

Country	all together			const. coeff.			nonstationary		
Australia	0.01653	1	(1)	0.01655	3	(4)	0.01653	2	(2)
Brazil	0.04709	3	(5)	0.03960	1	(3)	0.04359	2	(4)
Canada	0.01364	3	(4)	0.01358	1	(2)	0.01361	2	(3)
Chile	0.07163	1	(4)	0.07256	3	(6)	0.07255	2	(5)
Finland	0.01137	2	(4)	0.01097	1	(2)	0.01140	3	(5)
Japan	0.03009	1	(3)	0.03034	3	(5)	0.03011	2	(4)
Mexico	0.05507	1	(1)	0.05550	3	(3)	0.05512	2	(2)
Sweden	0.08938	1	(3)	0.09142	3	(5)	0.09031	2	(4)
UK	0.01058	1	(4)	0.01060	2	(5)	0.01068	3	(6)
US	0.03181	2	(5)	0.03293	3	(6)	0.03021	1	(2)
Rank Sum		16	(34)		23	(41)		21	(37)

Table 3.14: MSFE for 1-step combined forecasts allowing for changing means.

Country	all together			const. coeff.			nonstationary		
Australia	0.02359	3	(3)	0.02354	1	(1)	0.02356	2	(2)
Brazil	0.13363	3	(5)	0.09995	1	(3)	0.10651	2	(4)
Canada	0.04767	3	(5)	0.04706	1	(3)	0.04734	2	(4)
Chile	0.25561	1	(4)	0.25796	2	(5)	0.25962	3	(6)
Finland	0.04556	3	(5)	0.04468	2	(4)	0.04372	1	(2)
Japan	0.04666	1	(3)	0.04873	3	(5)	0.04700	2	(4)
Mexico	0.11604	1	(1)	0.11891	3	(3)	0.11774	2	(2)
Sweden	0.07827	1	(3)	0.09083	3	(5)	0.08680	2	(4)
UK	0.04597	2	(4)	0.04592	1	(3)	0.04623	3	(6)
US	0.11893	2	(5)	0.12283	3	(6)	0.10476	1	(2)
Rank Sum		20	(38)		20	(38)		20	(36)

Table 3.15: MSFE for 3-step combined forecasts allowing for changing means.

Country	all together			const. coeff.			nonstationary		
Australia	0.06142	3	(5)	0.05803	1	(3)	0.05961	2	(4)
Brazil	0.92067	3	(5)	0.68463	1	(3)	0.71151	2	(4)
Canada	0.33681	3	(5)	0.33146	1	(3)	0.33330	2	(4)
Chile	1.05164	1	(3)	1.06104	3	(6)	1.05654	2	(4)
Finland	0.27569	2	(5)	0.29892	3	(6)	0.23269	1	(2)
Japan	0.31004	1	(3)	0.35118	3	(5)	0.32138	2	(4)
Mexico	0.07890	3	(3)	0.07550	2	(2)	0.07464	1	(1)
Sweden	0.59934	1	(3)	0.79547	3	(5)	0.73756	2	(4)
UK	0.11837	2	(4)	0.11750	1	(3)	0.11867	3	(5)
US	1.09793	3	(5)	1.07479	2	(4)	0.84319	1	(3)
Rank Sum		22	(41)		20	(40)		18	(35)

Table 3.16: MSFE for 12-step combined forecasts allowing for changing means.

Country	by regression			by encompassing			by MSFE		
Australia	0.03413	3	(7)	0.01650	1	(2)	0.01650	2	(3)
Brazil	0.89896	3	(7)	0.04336	2	(4)	0.04239	1	(3)
Canada	0.13034	3	(7)	0.01367	1	(3)	0.01367	2	(4)
Chile	0.08804	3	(7)	0.06877	1	(1)	0.06879	2	(2)
Finland	0.02162	3	(7)	0.00999	2	(3)	0.00981	1	(2)
Japan	0.07005	3	(7)	0.03011	2	(4)	0.03010	1	(3)
Mexico	0.10699	3	(7)	0.05493	2	(3)	0.05487	1	(2)
Sweden	0.29765	3	(7)	0.08504	2	(4)	0.08487	1	(3)
UK	0.01673	3	(7)	0.01079	2	(5.5)	0.01026	1	(1)
US	0.36190	3	(7)	0.03071	1	(3)	0.03079	2	(4)
Rank Sum		30	(70)		16	(32.5)		14	(27)

Table 3.17: MSFE for 1-step combined forecasts with adjusted estimates.

Country	by regression			by encompassing			by MSFE		
Australia	0.02879	3	(7)	0.02419	2	(5.5)	0.02349	1	(2)
Brazil	0.42992	3	(7)	0.13240	2	(4)	0.11761	1	(3)
Canada	0.57642	3	(7)	0.04788	1	(3)	0.04792	2	(4)
Chile	0.47212	3	(7)	0.24908	2	(4)	0.24904	1	(3)
Finland	0.15689	3	(7)	0.04159	2	(4)	0.04147	1	(3)
Japan	0.36412	3	(7)	0.04652	1	(3)	0.04653	2	(4)
Mexico	0.13637	3	(7)	0.11930	2	(3)	0.11436	1	(1)
Sweden	0.62576	3	(7)	0.11307	2	(5)	0.07697	1	(3)
UK	0.09040	3	(7)	0.04420	2	(3)	0.04408	1	(2)
US	3.67489	3	(7)	0.11022	1	(4)	0.11041	2	(5)
Rank Sum		30	(70)		17	(38.5)		13	(30)

Table 3.18: MSFE for 3-step combined forecasts with adjusted estimates.

Country	by regression			by encompassing			by MSFE		
Australia	0.07243	3	(6)	0.06131	1	(3)	0.06237	2	(4)
Brazil	69.16982	3	(7)	0.42985	1	(2)	0.83081	2	(4)
Canada	0.68371	3	(7)	0.33604	1	(3)	0.33694	2	(4)
Chile	12.08243	3	(7)	1.00945	1	(2)	1.01081	2	(3)
Finland	2.46186	3	(7)	0.24487	1	(4)	0.25616	2	(5)
Japan	3.76799	3	(7)	0.30521	1	(3)	0.30786	2	(4)
Mexico	67.33295	3	(7)	0.07933	2	(2)	0.07483	1	(1)
Sweden	208.68746	3	(7)	0.52939	1	(3)	0.56822	2	(4)
UK	2.89242	3	(7)	0.11564	2	(4)	0.11562	1	(3)
US	23.27379	3	(7)	0.74617	1	(3)	0.93055	2	(5)
Rank Sum		30	(69)		12	(29.5)		18	(37)

Table 3.19: MSFE for 12-step combined forecasts with adjusted estimates.

Obviously, obtaining weights by an unconstrained, homogeneous regression equation is not an adequate technique at all. Note that the estimates used to calculate the combined forecasts are in fact not weights, as they are not restricted to positive values summing up to one. Using combinations made up of forecast-encompassing models works, particularly, at the long horizon. There, for the US STUR is encompassing all rival models; for Brazil, the difference-stationary models are encompassing. At the medium horizon, the STUR model is forecast-encompassing for Australia; for Mexico and Sweden, ARIMA(0,1,0) and ARIMA(0,1,1) are used exclusively in combinations. Finally, at the short horizon, for the UK ARIMA(0,1,1) is the only model used in forecasting corresponding to the results of the forecast-encompassing test. For the remaining situations, there are no encompassing models. Deriving weights from inverse mean squared forecast errors performs quite similarly with some more strength at the short and medium horizon. In many cases, combining does not succeed in generating a superior forecast when compared to certain individual forecasts. Particularly, combined forecasts are affected by additional estimation errors resulting from weights identification. Given that there are strong similarities among the rival models, forecast combinations cannot merge complementary characteristics for what they are actually used for. Similarities among the rival models are studied by means of cumulative forecast errors as part of the next section.

3.6 Cumulative Forecast Errors

Alternatively, to compare forecasts from rival models and to check whether certain models systematically under- or overestimate the realized values in finite samples, cumulative forecast errors can be evaluated. We consider the cumulative sum of forecast errors as defined by

$$CSFE_i = \sum_{t=251}^{i} [u_{t+h-1} - \hat{u}_{t-1}(h)] \tag{3.38}$$

for $i = 251, 252, \ldots, 271 - h + 1$ where the number of forecast observations to be evaluated is equal to $n_f = 21, 19, 10$ for 1-step, 3-step, 12-step forecasts, respectively, as already used for the mean squared forecast error. Cumulative forecast errors of the four competitors are plotted for each country over the whole forecast period. Results are presented in Figures 3.1 to 3.6 where random walk, ARIMA(0,1,1), AR(1) and STUR are marked with circles, triangles, crosses and squares, respectively.

A good forecast is associated with cumulative forecast errors around zero, with small variation and no trends in deviating. In case of 1-step forecasts, except for Mexico and the UK, there is some tendency in forecast values to be too low when compared to realized values. For Australia, Brazil, Canada and Mexico, random walk and ARIMA(0,1,1) work best; for Japan, Sweden and the US, STUR is the winner whereas for Chile and Finland, ARIMA(0,1,1) has definitely the lowest sum of cumulative forecast errors in absolute value. For the UK, all models are very close to each other. In many cases this corresponds to the results of mean squared forecast errors in Tables 3.5, 3.8 and 3.11. Conflicting situations are related to ARIMA(0,1,1) for Chile and Finland.

Cumulative sums of forecast errors of 3-step forecasts show up with tendencies to under- or overforecast quite similar to the situation with 1-step forecasts. Except for Mexico and the UK, underestimation is present. Superior models are distributed as before. This is in line with mean squared forecast errors in Tables 3.6, 3.9 and 3.12, except for some cases with ARIMA(0,1,1). Cumulative forecast errors coincide with the results from mean squared forecast errors for the US STUR model in Table 3.12 where estimates are updated for every information set.

For 12-step forecasts of US unemployment rates now random walk and ARIMA(0,1,1) perform best; for Finland, the STUR model is superior. Overall patterns look very similar to before, also when compared to mean squared forecast errors in Tables 3.7, 3.10 and 3.13. Obviously, the stationary AR(1) model does not play any role from the perspective of cumulative forecast errors. However, measured by mean squared forecast errors AR(1) has its benefits, especially at shorter forecast horizons which may be due to less larger forecast errors resulting from dampening AR(1) forecasts applied to constrained data like unemployment rates.

3.7 Forecasts from Simulated Data

For some countries, the forecast results from Section 3.4 suggest that the STUR model is not a great forecast model at all. To check this impression, we may generate 1000 replications of a STUR process with 271 observations corresponding to the estimates presented in Table 3.2 for each country. 1-step forecasts are calculated for the last 21 observations with new estimates for every information set. As the Bayesian procedure applied to obtain the estimates in Table 3.2 is a very expensive one, we make use of the alternative estimators suggested in Klüppelberg et al. (2002) which are based on maximizing the log-pseudo-likelihood of realizations of independent normal random variables having mean and variance equal to conditional mean and conditional variance, respectively, of the STUR model given in equations (3.9) and (3.10). Simulations in Holl (2013b) where this method is used are quite promising. Forecast models are ranked according to mean squared forecast errors. Rank distributions and rank sums (RS) are shown in Table 3.20.

To sum up, the STUR model is more adequate than random walk or ARIMA(0,1,1) to forecast realizations generated by a STUR model. However, AR(1) strongly outperforms STUR at this forecast horizon for all countries which corresponds to the results in Tables 3.5, 3.8 and 3.11.

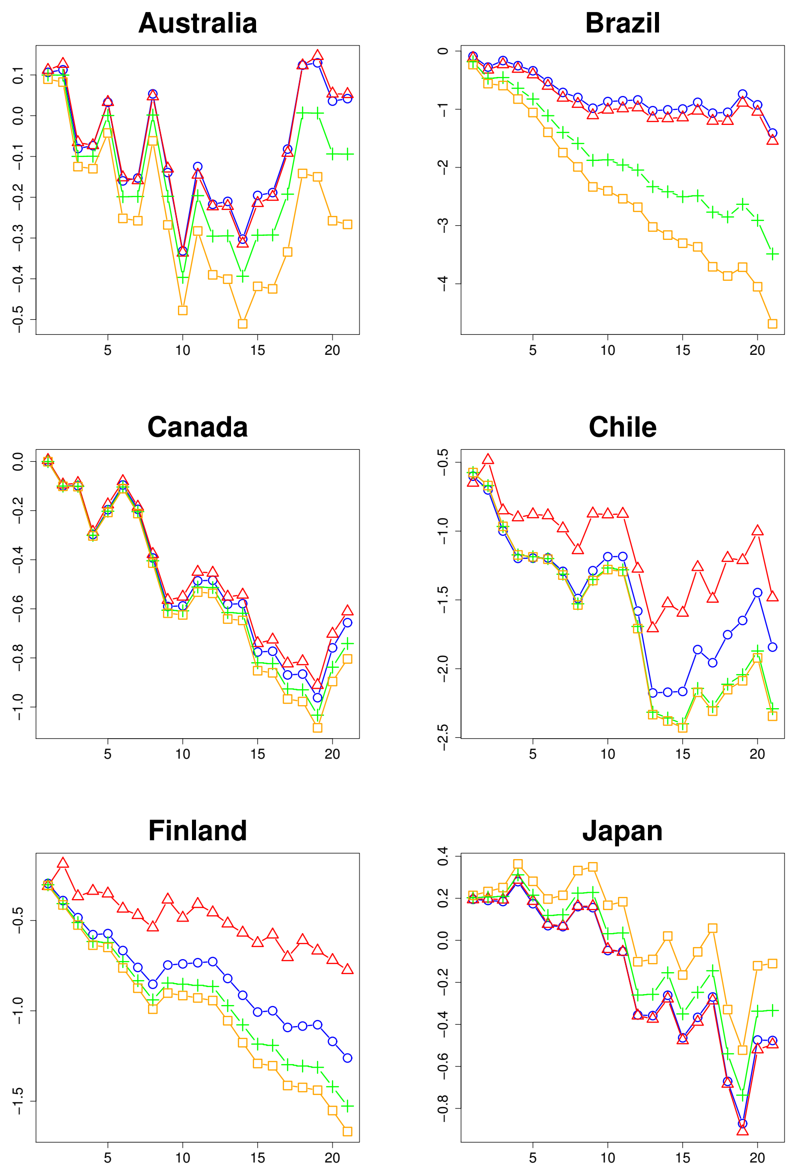

Figure 3.1: Cumulative forecast errors for 1-step forecasts (1/2).

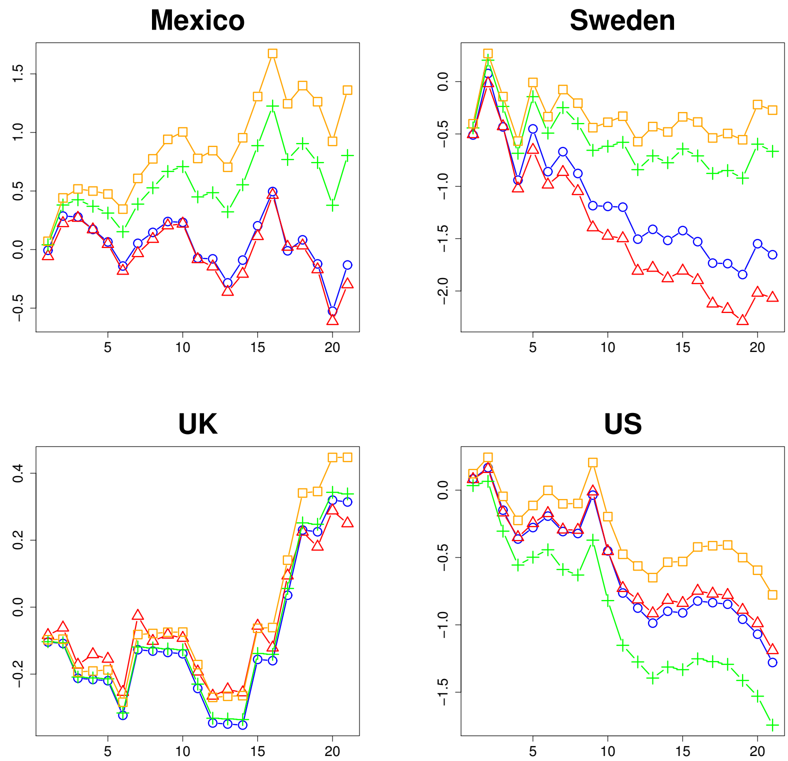

Figure 3.2: Cumulative forecast errors for 1-step forecasts (2/2).

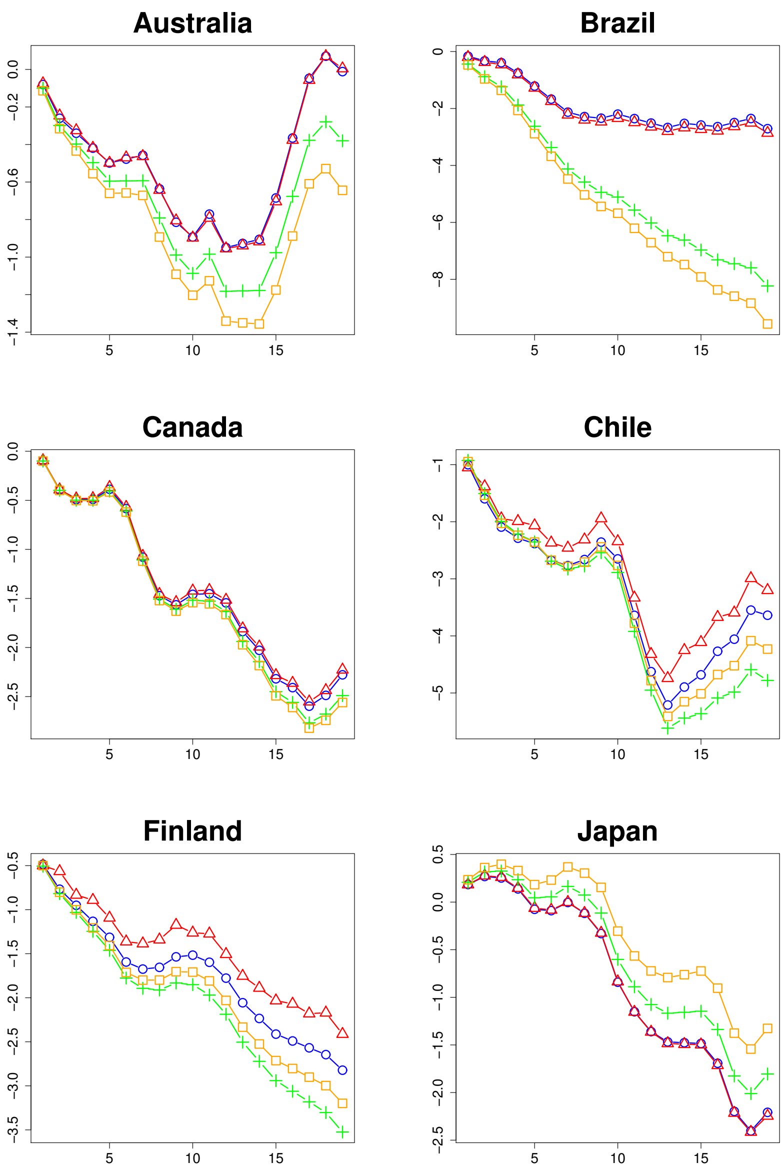

Figure 3.3: Cumulative forecast errors for 3-step forecasts (1/2).

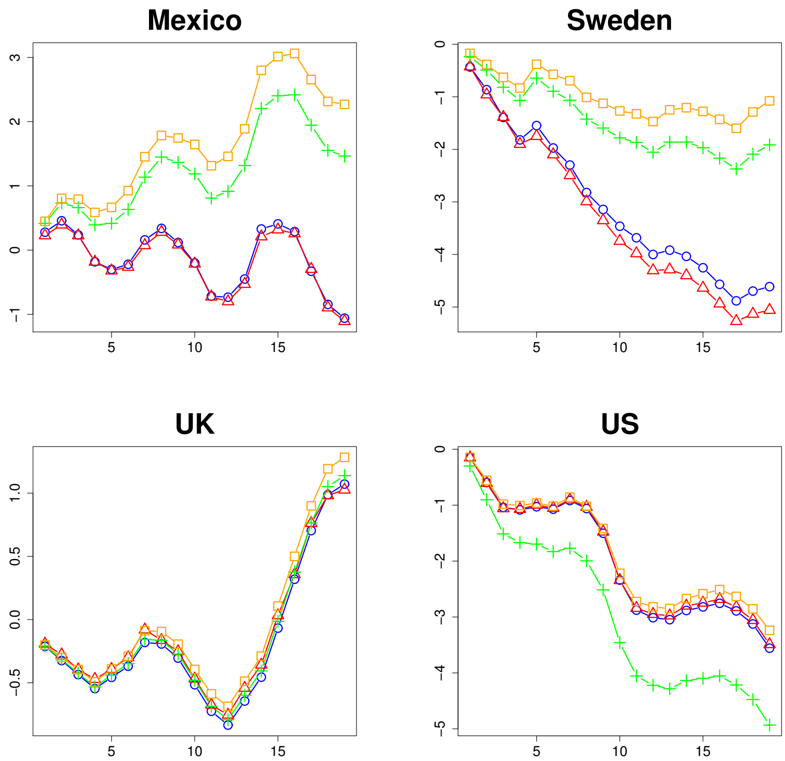

Figure 3.4: Cumulative forecast errors for 3-step forecasts (2/2).

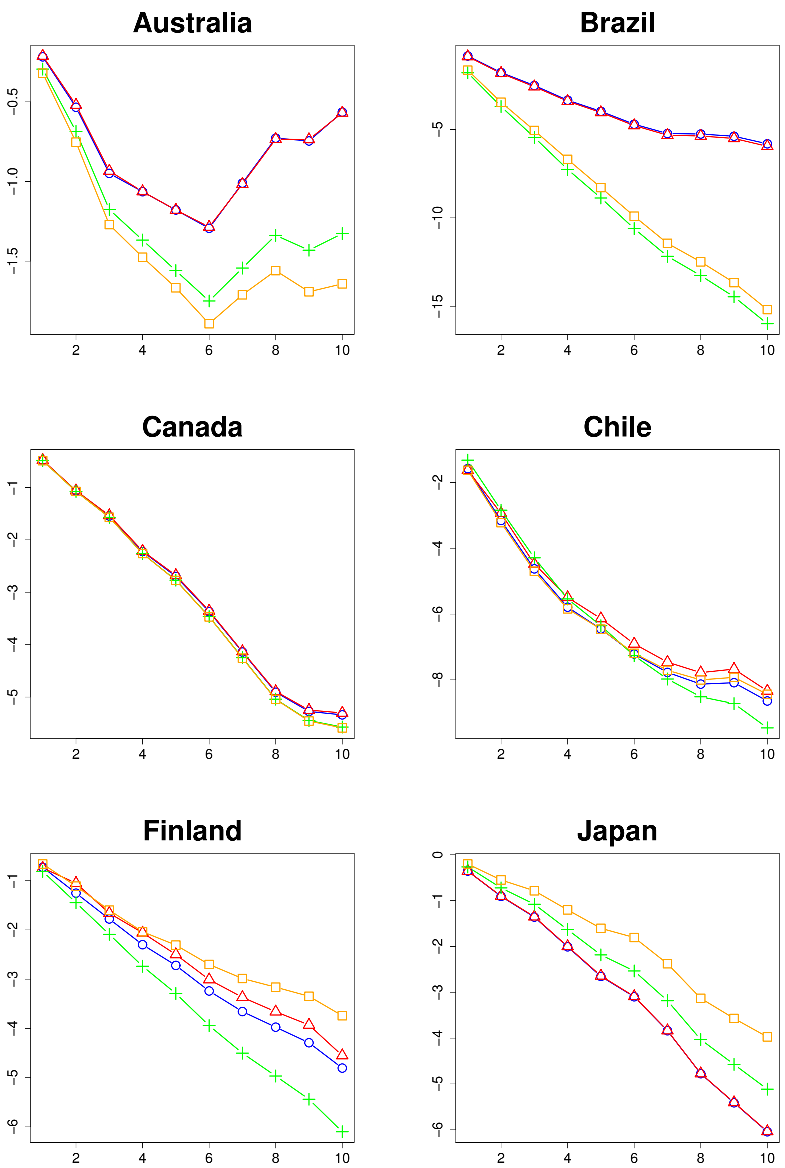

Figure 3.5: Cumulative forecast errors for 12-step forecasts (1/2).

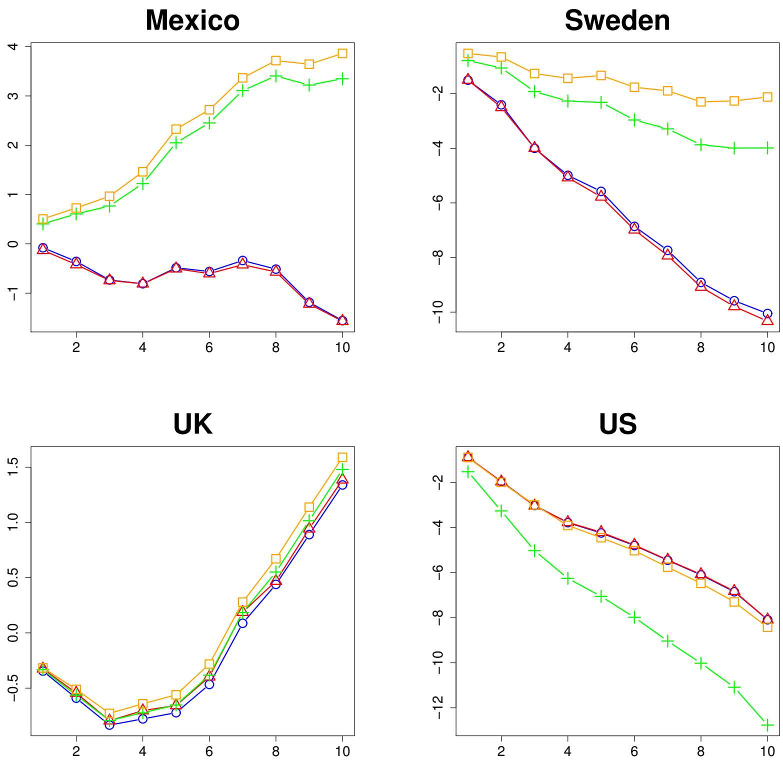

Figure 3.6: Cumulative forecast errors for 12-step forecasts (2/2).

	RW					ARIMA(0,1,1)					AR(1)					STUR				
Country	1	2	3	4	RS	1	2	3	4	RS	1	2	3	4	RS	1	2	3	4	RS
Australia	146	237	398	219	2690	238	146	196	420	2798	515	146	111	228	2052	101	471	295	133	2460
Brazil	103	132	482	283	2945	149	90	240	521	3133	484	295	95	126	1863	264	483	183	70	2059
Canada	138	217	397	248	2755	219	132	208	441	2871	522	166	101	211	2001	121	485	294	100	2373
Chile	110	250	479	161	2691	179	78	144	599	3163	503	254	110	133	1873	208	418	267	107	2273
Finland	179	242	446	133	2533	231	71	94	604	3071	406	272	163	159	2075	184	415	297	104	2321
Japan	123	212	421	244	2786	216	101	214	469	2936	523	158	97	222	2018	138	529	268	65	2260
Mexico	90	108	517	285	2997	114	92	254	540	3220	546	255	72	127	1780	250	545	157	48	2003
Sweden	118	130	435	317	2951	157	98	276	469	3057	513	232	97	158	1900	212	540	192	56	2092
UK	149	269	378	204	2637	249	115	185	451	2838	481	182	130	207	2063	121	434	307	138	2462
US	131	381	361	127	2484	241	108	148	503	2913	507	143	138	212	2055	121	368	353	158	2548
Total Sum	1287	2178	4314	2221	27469	1993	1031	1959	5017	30000	5000	2103	1114	1783	19680	1720	4688	2613	979	22851

Table 3.20: Ranks for 1-step forecasts from simulated data.

3.8 Conclusions

The present study tackles the question whether STUR models can be used to forecast unemployment rates adequately. In a broader sense, this addresses the question whether unemployment rates can be modelled appropriately by STUR. Unfortunately, this cannot be answered unambiguously. The object of research is discussed from three different perspectives. Unit-root test results presented in Holl (2013a) are confronted with estimation results concerning the STUR model as done in Holl (2013b) and concerning some standard time series models as done in course of the present paper. Parameter estimates are used to generate a great many of replications to obtain size and power estimates. The models involved are used to calculate out-of-sample forecasts by conditional means.

Unit-root tests prepared to have power against STUR processes reject the null hypothesis of a unit root for five out of ten countries at least at the 5% level of significance. Particularly, only for Japan and the UK, there is no test rejecting the null at this level. As a consequence, we could expect processes different from a random walk to be appropriate for the majority of countries.

Fitting AR(1) and STUR model to each series reveals strongly significant estimates in line with the test results. Simulations support the case for a difference-stationary model merely for Japan and the UK. There, the STUR model is suggested for six out of ten countries. After extracting in-sample information, the rejection of the null hypothesis in unit-root testing seems quite reasonable, however the reason why the null is rejected is left unclear. Thus we consult the forecast performance of each model which may shed some light on that issue.

On average, i.e. over all countries, the random walk model performs best, measured by rank sums. In particular, this holds independent of the forecast horizon selected (see Tables 3.5 to 3.7, 3.8 to 3.10 and 3.11 to 3.13). STUR shows up with a rather weak performance on average associated with a strong dispersion. That is, there are series where the STUR model performs great and series where it performs poorly. The poor results seem to be more stable over different forecast horizons. STUR tends to have some slight advantages in the long run. However, the AR(1) model performs bad as well, particularly at a longer forecast horizon. Forecasts generated from the stationary AR(1) model are not that excellent for hardly any country. Adjusting the demeaning procedure for a changing mean somewhat positively affects the results of AR(1) and STUR at the shorter forecast horizons. Consistent recommendations resulting both from testing and forecasting are obtained for the UK with a random walk and for Chile where STUR or maybe AR(1) seem to be appropriate models.

By arranging combined forecasts a further phenomenon becomes evident. There are only a few cases where a combined forecast outperforms the best individual one. Ignoring either AR(1) or STUR at the short horizon is costly whereas at the long horizon it makes sense – on average (see Tables 3.14 to 3.16). Obtaining weights by forecast encompassing tests has its benefits at the medium and long horizon whereas using inverse mean squared forecast errors in weighting performs better at the short and medium horizon (see Tables 3.17 to 3.19).

Plotting cumulative forecast errors provides evidence concerning somewhat strong similarities among the rival models. From that perspective, AR(1) is not superior at any forecast horizon which is not in line with the results from mean squared forecast errors. Conflicts between cumulative forecast errors and mean squared forecast errors also arise associated with ARIMA(0,1,1).

Finally, the question is raised whether STUR models are good forecast models at all. That is, given that the data are generated by a STUR model, does it outperform the rival models in forecasting. Results for 1-step forecasts applied to simulated data are rather discouraging. STUR can only outrival random walk and ARIMA(0,1,1) (see Table 3.20).

Clearly, forecaster's perspective focuses on predicting unemployment rates of individual countries. The present study says that there is no single model adequate to represent unemployment rates of all countries. Furthermore, testing considers in-sample behavior which need not and in most cases does not coincide with out-of-sample behavior. As the future of unemployment rates is difficult to forecast per se, the respective forecast horizon matters. We may expect different things for the next month than for the next year, not only from a quantitative point of view. In the context of forecasting, the STUR model extends our opportunities to form expectations concerning the future. Sometimes it succeeds.

Bibliography

Ball, L. M., 2009, Hysteresis in unemployment: old and new evidence. NBER Working Paper 14818.

Billingsley, P., 1968, Convergence of probability measures. Wiley, New York.

Borkovec, M., 2000, Extremal behavior of the autoregressive process with ARCH(1) errors. Stochastic Processes and their Applications 85/2, 189–207.

Dickey, D. A. and W. A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431.

Durrett, R., 1991, Probability: theory and examples. Wadsworth & Brooks/Cole, Pacific Grove, California.

Friedman, M., 1968, The role of monetary policy. American Economic Review 58/1, 1–17

Fuller, W. A., 1976, Introduction to statistical time series. Wiley, New York.

Geweke, J., 1992, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Bayesian Statistics 4, 169–193.

Granger, C. W. J. and R. Ramanathan, 1984, Improved methods of combining forecasts. Journal of Forecasting 3/2, 197–204.

Granger, C. W. J. and N. R. Swanson, 1997, An introduction to stochastic unit-root processes. Journal of Econometrics 80, 35–62.

Guyton, D. A., N.-F. Zhang and R. V. Foutz, 1986, A random parameter process for modeling and forecasting time series. Journal of Time Series Analysis 7/2, 105–115.

Hall, R. E., 1978, Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. Journal of Political Economy 86/6, 971–987.

Harvey, D. and P. Newbold, 2000, Tests for multiple forecast encompassing. Journal of Applied Econometrics 15/5, 471–482.

Holl, J., 2013a, Testing for unit roots using the AR-ARCH structure of STUR models. Available on request.

Holl, J., 2013b, Fitting STUR Models to Unemployment – Evaluation of a new Unit-Root Test. Available on request.

Jones, C. R. and J. M. Marriott, A Bayesian analysis of stochastic unit root models. Bayesian Statistics 6, 785–794.

Klüppelberg, C., R. A. Maller, M. Van De Vyver and D. Wee, 2002, Testing for reduction to random walk in autoregressive conditional heteroskedasticity models. Econometrics Journal 5, 387–416.

Koop, G., 2003, Bayesian econometrics. Wiley, Chichester.

Kurtz, T. G. and P. Protter, 1991, Weak limit theorems for stochastic integrals and stochastic differential equations. Annals of Probability 19/3, 1035–1070.

Leybourne, S. J., B. P. M. McCabe and A. R. Tremayne, 1996, Can economic time series be differenced to stationarity? Journal of Business & Economic Statistics 14/4, 435–446.

Phelps, E. S. and G. Zoega, 1998, Natural-rate theory and OECD unemployment. Economic Journal, 108/448, 782–801.

Phillips, P. C. B., 1987, Time series regression with a unit root. Econometrica 55/2, 277–301.

Protter, P. E., 2005, Stochastic integration and differential equations. Springer, Berlin.

Said, S. E. and D. A. Dickey, 1984, Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71/3, 599–607.

Schwert, G. W., 1989, Tests for unit roots: A Monte Carlo investigation. Journal of Business & Economic Statistics 7/2, 147–159.

Tsay, R. S., 1987, Conditional heteroscedastic time series models. Journal of the American Statistical Association 82/398, 590–604.

Weiss, A. A., 1984, ARMA models with ARCH errors. Journal of Time Series Analysis 5, 129–143.

Yang, F. and R. Leon-Gonzalez, 2010, Bayesian estimation and model selection in the generalized stochastic unit root model. Studies in Nonlinear Dynamics & Econometrics, 14/4, Article 5

Abstract

The dissertation at hand consists of three essays focussing on so-called stochastic unit-root (STUR) models in economics. A test of the null hypothesis of a unit root in an autoregressive model of order one (AR(1)) against alternatives nested in a STUR model is introduced. STUR models allow for changes between stationary and explosive regimes. The pseudo-likelihood ratio test is based on the equivalence in first and second conditional moment of STUR models and AR(1) models with errors from an autoregressive conditional heteroskedastic model of order one. Asymptotics are derived and critical values simulated. Monte Carlo experiments show that the test really has power against stationary and STUR alternatives. An application to unemployment rates of ten countries is provided. To evaluate the decisions of unit-root tests having power against STUR processes, STUR models are fitted to the unemployment rates by using Bayesian techniques. The Bayesian procedure in use allows to assess the deviation of STUR models from standard fixed-coefficient time series models. Parameter estimates show up strongly significant for all countries. As the Bayesian method is a very time-consuming one, an alternative procedure is studied in simulations. The application of STUR models to unemployment rates is motivated from an economic point of view. Forecasts of the unemployment rates generated from STUR models are evaluated. Unit-root test results on original and simulated data are confronted with 1-step, 3-step and 12-step forecast results where STUR is compared to three standard time series models. Additionally, combined forecasts are calculated. The question whether STUR is a good forecast model at all is addressed. Testing and forecasting do not coincide for every country. However, the discussion suggests that STUR is relevant for certain countries and should be considered a real alternative.

Zusammenfassung

Die vorliegende Dissertation besteht aus drei Essays, die von sogenannten stochastischen Einheitswurzel-(STUR-)Modellen in der Volkswirtschaftslehre handeln. Ein Test der Nullhypothese einer Einheitswurzel in einem autoregressiven Modell erster Ordnung (AR(1)) gegen Alternativen, die in einem STUR-Modell eingebettet sind, wird vorgestellt. STUR-Modelle ermöglichen den Wechsel zwischen stationären und explosiven Regimen. Der Pseudo-Likelihood-Quotienten-Test basiert auf der Äquivalenz der ersten und zweiten bedingten Momente von STUR-Modellen und AR(1)-Modellen mit Störtermen eines autoregressiven bedingt heteroskedastischen Modells erster Ordnung. Die asymptotische Verteilung wird hergeleitet und kritische Werte werden simuliert. Monte-Carlo-Experimente zeigen, dass der Test tatsächlich Teststärke gegen stationäre und STUR-Alternativen besitzt. Der Test wird auf die Arbeitslosenquoten von zehn Staaten angewandt. Unter Verwendung Bayesianischer Methoden werden STUR-Modelle für diese Arbeitslosenquoten gefittet, um die Entscheidungen von Einheitswurzel-Tests, die Teststärke gegen STUR-Prozesse besitzen, zu evaluieren. Das Bayesianische Verfahren ermöglicht die Einschätzung der Abweichung der STUR-Modelle von gewöhnlichen Zeitreihenmodellen mit konstanten Koeffizienten. Die geschätzten Werte für die Parameter sind hoch signifikant für alle Staaten. Da das Bayesianische Verfahren sehr zeitaufwendig ist, wird ein alternatives Verfahren in Simulationen getestet. Die Anwendung der STUR-Modelle auf Arbeitslosenquoten wird aus einer ökonomischen Perspektive motiviert. Prognosen der Arbeitslosenquoten, die sich aus STUR-Modellen errechnen, werden evaluiert. Resultate der Einheitswurzel-Tests angewandt auf reale und simulierte Daten werden den Resultaten von 1-Schritt-, 3-Schritt und 12-Schritt-Prognosen gegenübergestellt; STUR wird dabei mit drei gewöhnlichen Zeitreihenmodellen verglichen. Darüberhinaus werden kombinierte Prognosen berechnet. Die Fra-

ge, ob STUR überhaupt ein gutes Prognosemodell ist, wird erörtert. Test und Prognose führen nicht für jeden Staat zum selben Ergebnis, jedoch für bestimmte Staaten ist STUR ein relevantes Modell und sollte als echte Alternative betrachtet werden.

Curriculum Vitae

Personal Data

Name:	Jürgen Holl
Date of Birth:	May 19, 1982
Place of Birth:	Wels, Upper Austria
Citizenship:	Austria
Marital Status:	married

Education

Oct. 2007 until now:	PhD studies in economics at the University of Vienna Dissertation: "Stochastic Unit-Root Models in Economics – Essays on Testing, Estimating and Forecasting", supervised by Prof. Robert M. Kunst and Prof. Erhard Reschenhofer
Sept. 2007:	Master in economics (Mag.rer.soc.oec.) Thesis: "Detecting Hysteresis in Unemployment. A Nonlinear Challenge", supervised by Prof. Robert M. Kunst
Oct. 2003 – Sept. 2007:	Diploma studies in economics and philosophy at the University of Vienna
Oct. 2001 – Jan. 2002:	Diploma studies in economics and sociology at the Johannes Kepler University Linz
June 2001	Reifeprüfung and Diploma Certificate
1996 – 2001:	Secondary College for Mechanical Engineering in Wels
1992 – 1996:	Grammar school ("Realgymnasium") in Wels
1988 – 1992:	Elementary school in Stadt Haag, Lower Austria

Work Experience

Apr. 2012 until now:	Statistical Modeler, Synthesis Forschung
Sept. 2011 – Feb. 2012:	Statistician, Macroeconomic Statistics Directorate, Statistics Austria
Mar. 2011 – Aug. 2011:	Research assistant, Department of Economics and Philosophy of Science, University of Applied Sciences Wiener Neustadt
Oct. 2007 – Dec. 2010:	Research assistant, Initiative Group "Issues in the Global Economy: Dynamics, Governance, and Information", Department of Economics, University of Vienna

Publications

"Unit Root in Unemployment – New Evidence from Nonparametric Tests"
(with Robert M. Kunst), Applied Economics Letters 18, 509-512 (2011)