An Introduction to Stochastic Unit Root Processes

Clive W.J. Granger and Norman R. Swanson*

University of California, San Diego, La Jolla, California and
Pennsylvania State University, University Park, PA, respectively

February 1994
this revision: April 1996

A class of nonlinear processes which have a root that is not constant, but is stochastic, and varying around unity is introduced. The process can be stationary for some periods, and mildly explosive for others. *Stochastic* unit roots are seen to arise naturally in economic theory, as well as in everyday macroeconomic applications. It is shown that standard tests, such as the augmented Dickey-Fuller test, cannot easily distinguish between *exact* unit roots and *stochastic* unit roots. An alternative test which has difference stationarity as the null suggests that exact unit roots models are often rejected in favor of more general nonlinear stochastic unit root (STUR) models. Estimation is discussed, and, a forecast comparison of linear random walk and AR(p) models, time varying parameter models, and STUR models suggests that this new class of processes is potentially useful, particularly when the objective is multi-step ahead forecasting.

JEL CLASSIFICATION C22, C51, C53

KEYWORDS doubly stochastic, stochastic unit root, cointegration, approximate maximum likelihood

---

Correspondence to: C.W.J. Granger, Department of Economics 0508, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508.

* Many thanks to three anonymous referees, the editor, Valentina Corradi, Marjorie Flavin, Scott Gilbert, Stephane Gregoir, Eilev Jansen, Valerie Ramey and Hal White for helpful suggestions. Research was supported by NSF Grants SES 9023037, 9308295 and by Social Sciences and Humanities Research Council of Canada Award No. 753-91-0416.

1. Introduction

It is well established that many economic series contain dominant, smooth components, even after removal of simple deterministic trends. This is particularly true for series from finance and macroeconomics. A traditional way to model these components is as random walks, leading to I(1) and ARIMA (p,l,q) models in the Box and Jenkins' (1970) terminology. The success of such models with economic data and relationships between groups of such series, including the idea of cointegration, has resulted in a considerable literature and special statistical techniques to handle the problems of inference with data having this property. A small part of this literature is included in the book of readings by Engle and Granger (1991).

If a series is such that its first difference is stationary (and has positive spectrum at zero frequency) then the series has an exact (or pure) unit root. It is quite possible that such roots occur in practice. However, there is little economic theory that produces them, and, in practice, there are many types of processes that have near unit roots and are very difficult to distinguish from perfect unit root processes, given a finite sample. In this paper, a particular class of such processes are introduced, having a root that is not constant, but is stochastic, and varying around unity. In this way, the process is stationary for some periods, and mildly explosive for others. However, on average, the series may seem to be I(1), according to standard tests.

The class of processes here is a stochastic version of the time-varying and seasonal unit root processes suggested by Osborn (1988), and considered by Franses (1994), where, for instance

$$x_t = \phi_s \, x_{t-1} + \varepsilon_t$$

and x_t is observed quarterly. The parameters, ϕ_s, are allowed to change each quarter, such that $\prod_{s=1}^{4} \phi_s = 1$. Alternately, the stochastic unit root (STUR) process can be thought of as a special case of the time-varying parameter processes discussed in Andel (1976), with the special feature that the variable is nonstationary. In this sense, STUR processes combine unit root as well as time varying parameter characteristics which are of some interest in economics.

Interestingly, many economic series appear to be modeled well as STUR processes, based on a forecasting analysis which compares four types of models: (i) random walk (with drift) processes; (ii) fixed parameter autoregressive processes; (iii) time varying parameter models (using a Kalman filter for estimation); and (iv) STUR processes. In particular, STUR models perform well at multi-step ahead forecast horizons.

The rest of the paper is organized as follows. STUR processes are introduced in Section 2, and various properties are derived. Standard I(1) tests are discussed in Section 3. Section 4 presents a STUR test, and Section 5 examines estimation strategies. Section 6 compares the forecasting ability of the STUR model with standard time varying parameter models, linear autoregressive models, and random walk models. The final section briefly discusses bivariate extensions, and concludes.

2. Stochastic Unit Root Processes

Consider a series x_t generated by

$$x_t = a_t\, x_{t-1} \;+\; \varepsilon_t \;, \tag{2.1}$$

where ε_t is zero mean, *i.i.d.* with variance σ_ε^2, and where

$$a_t \;=\; \exp(\alpha_t) \tag{2.2}$$

with α_t a Gaussian stationary series having mean m, variance σ_α^2 and power spectrum $g_\alpha(\omega)$. This is an example of a doubly stochastic process as considered by Tjøstheim (1986), and has been specifically analyzed by Brandt (1986), and Pourahmadi (1986, 1988). However, the exponential form (2.2) is a convenient case for consideration of the long-run properties of the series. A particular example that will be used to illustrate the results is the case where α_t is given by an AR(1) process

$$\alpha_t \;=\; \mu + \rho\, \alpha_{t-1} + \eta_t \tag{2.3}$$

where $|\rho| < 1$, and η_t is *i.i.d.* normally distributed $N(0,\sigma_\eta^2)$ and is independent of the series ε_t. It will be assumed throughout this section that α_t is generated exogenously from x_t, so that

$$g\,(\,\alpha_{t+1} \mid \alpha_{t-j}\,,\, x_{t-j}\,,\, j \geq 0\,) \;=\; g\,(\,\alpha_{t+1} \mid \alpha_{t-j}\,,\, j \geq 0\,)$$

where $g(\alpha \mid I)$ is a conditional distribution. For the example (2.3), $m = \mu/(1-\rho)$ and $\sigma_\alpha^2 = \sigma_\eta^2/(1-\rho^2)$. (2.1) can be solved as

$$x_t \;=\; \varepsilon_t + a_t\, \varepsilon_{t-1} + a_t a_{t-1}\, \varepsilon_{t-2} + \ldots + a_t a_{t-1} \cdot \ldots \cdot a_{t-k+2}\, \varepsilon_{t-k+1} + a_t a_{t-1} \cdot \ldots \cdot a_{t-k+1} x_{t-k}$$

or,

$$x_t = \varepsilon_t + \pi_{t,1}\,\varepsilon_{t-1} + \pi_{t,2}\,\varepsilon_{t-2} + \ldots + \pi_{t,k-1}\,\varepsilon_{t-k+1} + \pi_{t,k}x_{t-k} \tag{2.4}$$

for any integer k, $0 \le k \le t$, where,

$$\pi_{t,j} = \exp(S_{\alpha t}(j)) ,$$

and,

$$S_{\alpha t}(j) = \sum_{i=0}^{j-1} \alpha_{t-i}$$

with the notation $S_{\alpha t}(0) = 1$. It should be noted that for j>0

$$E_j \equiv E[S_{\alpha t}(j)] = jm \ , \tag{2.5}$$

and,

$$V_j = var[S_{\alpha t}(j)] = \sigma_\alpha^2[j + \sum_{r=1}^{j-1}(j-r)\rho_\alpha(r)] \tag{2.6}$$

where $\rho_\alpha(r) = corr(\alpha_t , \alpha_{t-r})$. For j large (2.6) can be approximated as

$$V_j \approx jf_\alpha(0) \tag{2.7}$$

as shown in Koopmans (1974, page 171) where $f_\alpha(w) = \frac{1}{2\pi} g_\alpha(w)$.

A standard result that is useful in this analysis is that if a random variable X is distributed $N(m , \sigma^2)$ then

$$E[\exp(kX)] = \exp(km + \tfrac{1}{2}k^2\sigma^2). \tag{2.8}$$

It follows that for k large,

$$E[\pi_{t,k}] = \exp(E_k + \tfrac{1}{2}V_k) \approx \exp(k\theta) \ , \tag{2.9}$$

where,

$$\theta = m + \tfrac{1}{2}f_\alpha(0). \tag{2.10}$$

Assuming for convenience that x_0 is deterministic, and taking expectations of (2.4) with $k = t$ gives

$$E[x_t] = x_0\exp(t\theta)$$

and, as the ε_t are *i.i.d.*,

$$var(x_t) = \sigma_\varepsilon^2 \sum_{j=0}^{t-1} E[\pi_{t,j}^2] = \sigma_\varepsilon^2 \sum_{j=0}^{t-1} \exp(2jm + 2V_j) \approx \sigma_\varepsilon^2 \sum_{j=0}^{t-1} \exp(2j\phi) \ , \tag{2.11}$$

where, $\phi = m + f_\alpha(0)$.[1] Thus,

$$\left.\begin{aligned} var(x_t) &= \sigma_\varepsilon^2 \left[\frac{1 - \exp(t2\phi)}{1 - \exp(2\phi)}\right] && \text{if } \phi \neq 0 \\ &= t\,\sigma_\varepsilon^2 && \text{if } \phi = 0 \end{aligned}\right\} \tag{2.12}$$

---

[1] This is an approximation of the true variance of x_t, since $E[\pi_{t,j}^2] \neq \exp(2j\phi)$ for small j.

Turning to consideration of (linear) regression coefficients $\beta_{t,k}$ as in:

$$x_t = \beta_{t,k} x_{t-k} + \text{error}$$

it can be noted from (2.4) that

$$E_L(x_t \mid x_{t-j}\,,\, j \geq k) = x_{t-k} E_L[\pi_{t,k} \mid x_{t-j}\,,\, j > k] = x_{t-k} E[\pi_{t,k}] \tag{2.13}$$

where $E_L(\cdot)$ is an expectation using only the components of the conditioning sets in a linear form and $E(\cdot)$ is an unconditional expectation. Thus, from (2.9),

$$\beta_{t,k} = E[\pi_{t,k}] = \exp(k\theta)\,, \text{ for large } k \tag{2.14}$$

Using the identity $corr(x_t\,,\, x_{t-k}) = \beta_{t,k} \sqrt{\frac{var(x_{t-k})}{var(x_t)}}$ thus gives the following approximate expressions for this autocorrelation when t and k are both large:

	$\theta < 0$	$\theta = 0$	$\theta > 0$
$\phi > 0$	$\exp(-kf)$	$\exp(-kf)$	$\exp(-kf)$
$\phi = 0$	$\exp(k\theta)(1 - k/t)^{1/2}$	1	-
$\phi < 0$	$\exp(k\theta)$	-	-

where $\theta = m + \tfrac{1}{2}f$, $\phi = \theta + \tfrac{1}{2}f$,and $f \equiv f_\alpha(0) = \frac{1}{2\pi} g_\alpha(0)$ where $g_\alpha(0)$ is the spectrum of α_t at zero frequency. Empty cells are cases that cannot occur, since by definition, $\phi \geq \theta$.

The autocorrelations decline exponentially (approximately) in all cases except when $\theta = \phi = 0$, so that $f_\alpha(0) = 0$ and $m = 0$. This special case occurs, for example, if $\alpha_t = 0$, all t, since it then follows that $E[e^{\alpha_t}] = 1$, and x_t is then a perfect unit root process.

Two alternative characterizations of the STUR process given by (2.1) and (2.2) are

(i) STURA: α_t is such that $E[e^{\alpha_t}] = 1$ so that

$$m + \tfrac{1}{2}\sigma_\alpha^2 = 0 \tag{2.15}$$

(ii) STURB: α_t is such that $E[\pi_{t,k}] = 1$, k large, where $\pi_{t,k}$ is given in (2.4), so that

$$\theta = 0 \quad \text{or} \quad m + \tfrac{1}{2} f_\alpha(0) = 0 \tag{2.16}$$

The persistence which arises when the restrictions given by (2.15) or (2.16) are met is similar to the persistence exhibited by perfect unit root processes. For instance, as in the case a random walk, shocks from the distant past continue to have an impact on current values of the process.

More precisely, the properties of STURA depend on the relative size of σ_α^2 and $f_\alpha(0)$. Suppose that $f_\alpha(0) > \sigma_\alpha^2$, so that, essentially, α_t is smoother than white noise. It follows that θ and ϕ are both positive,

so that from (2.12), the variance of x_t is increasing explosively, although probably mildly so, and $E[\pi_{t,k}]$ is increasing with k. Thus, distant shocks have greater impact on the current value of the series than do recent shocks. This property also holds for linear I(d), d>1, processes, for example, and might be called "super-persistence".

A STURB process will have $E[e^{\alpha_t}] < 1$ if $f_\alpha(0) > \sigma_\alpha^2$, so that a regression of

$$x_t = \beta_{t,k} x_{t-k} + error$$

will yield $\beta_{t,k}$ values which are nearer to unity for large values of k than for small values of k. The persistence property is now particularly interesting.

Recall that

$$x_t = \sum_{j=0}^{k-1} \pi_{t,j} \varepsilon_{t-j} + \pi_{t,k} x_0$$

and

$$\pi_{t,j} = \exp[S_{\alpha t}(j)], \quad j \geq 1 \quad \text{and} \quad \pi_{t,0} = 1$$

where, $S_{\alpha t}(j)$ has mean jm, variance $jf_\alpha(0)$, for j large. If α_t is normally distributed, then $\pi_{t,j}$ will have a lognormal distribution. With the constraint that $E[\pi_{t,j}] = 1$, which implies that m is negative, standard results then give

$$variance(\pi_{t,j}) = \exp(jf_\alpha(0)) - 1$$

$$mode(\pi_{t,j}) = \exp(-3/2\ jf_\alpha(0))$$

$$median(\pi_{t,j}) = \exp(-1/2\ jf_\alpha(0))$$

Also, a 95% confidence interval (about the median) is

$$\exp[-\tfrac{1}{2} jf_\alpha(0) \pm 2(jf_\alpha(0))^{1/2}]$$

Thus, as j increases, $\pi_{t,j}$ has a constant mean, an increasing variance, a mode and median which tend towards zero, and a $100(1-\delta)\%$ confidence interval that tends towards very small values. This implies that the form of persistence which is displayed is very fragile.

For a theoretical perspective, stochastic unit roots may be seen to arise quite naturally in economics. As an example, it is well known that many basic formulations of the permanent income hypothesis involve the term $\frac{1+\delta}{1+R_t}$ (see Hall (1978), for example). Using the standard approach, δ is the marginal

rate of time preference, and R_t is the real rate of interest. A frequently made assumption is that this term is identically unity. However, when that assumption is relaxed, and R_t is allowed to vary stochastically, then the term may well vary around unity, and probably has a mean very close to unity, as expected in the STUR case.

To summarize, the above results indicate that the properties of STUR processes are often markedly different from comparable properties of perfect unit root processes. Another characteristic of stochastic unit roots is that they are quite difficult to distinguish from perfect unit roots. This is not surprising given that evidence presented below indicates that variances of stochastic unit roots are often quite small. In this sense the usual power failures associated with unit root tests should apply. Nevertheless, before discussing an appropriate testing strategy for the presence of STUR, we provide an example of a STUR process which is readily mistaken for a pure unit root process based on a standard Dickey-Fuller test.

3. An Augmented Dickey-Fuller Test with a STUR Alternative

It is interesting to ask whether the oft applied augmented Dickey-Fuller (ADF) test has power against STUR alternatives. ADF tests have as the null hypothesis, that the series in question is distributed as I(1). However, with the amount of data available in macroeconomics, these tests are unlikely to have much power against a variety of processes which are, in some sense, near I(1). Processes against which unit root tests have little power may be called "generic unit root" processes (Granger (1993)).

In order to shed light on the power properties of the ADF test in the presence of STUR, two experiments were performed. In both experiments, a process was generated by (2.1), (2.2), (2.3). In the first experiment, $\eta_t \sim i.i.d.$ N(0, 0.0001), $\rho = 0.60$ and $\mu = -0.00003125$. The second experiment used $\eta_t \sim i.i.d.$ N(0, 0.000001), $\rho = 0.60$ and $\mu = -0.0000003125$. Thus, both experiments are STURA processes with $E(a_t) = 1$.

Augmented Dickey-Fuller tests using the regression

$$\Delta x_t = \beta_0 + \beta_1 t + \beta_2 x_{t-1} + \sum_{j=1}^{p} c_j \Delta x_{t-j} + \nu_t \tag{3.1}$$

where then run. The relevant test statistic is the standard t-value for β_2, the coefficient of x_{t-1}. The test can be run in various forms: without constraints (case 3), without trend (case 2, $\beta_1 = 0$), without trend

and intercept (case 1, $\beta_0 = \beta_1 = 0$), and for various amounts of augmentation (value of p). Monte Carlo results for various p values are given in Table 1. Although not reported, the results of the experiments varied very little for $p = 2,3,...,10$. The table lists the percentage of estimated t-values that lie above the theoretical 95% and 5% values, taken from the tables provided in Fuller (1986), for a sample size of 250. If the augmentation is estimated, being defined here as that which makes the estimated autocorrelations of the residuals of the augmented equation up to lag 20 all be in the region $\pm 2/\sqrt{250}$, the value of $p = 1$ is chosen in all cases. Thus, the experimental results for the ADF(1) tests are the most interesting. The other test results in Table 1 are for comparison.

To reiterate, in Table 1 results for $\sigma_\eta^2 = 0.0001$ and 0.000001 are reported. Clearly, in practice, the variance of the stochastic unit root depends on the particular characteristics of the data. For instance, estimation results which are presented in Section 6 suggest that σ_η^2 may be as large as 0.001, and as small as 0.0000001. In support of the values used in Table 1, note that the variance of the STUR term discussed above, $\frac{1+\delta}{1+R_t}$, was calculated to be 0.00026 using quarterly U.S. data from 1959:1-1993:2. (Defining $(1+R_t) = (1-mtr)i_t - [p_{t+1}/p_t - 1]$, where mtr=0.4 is the marginal tax rate, i_t is a quarterly average of the 3 month t-bill rate and p_t is the gross domestic income price deflator).

From both experiments, it is clear that the ADF(1) test cannot distinguish between the given STUR model and a standard pure unit root process. In particular, using the 95% critical value, the null of a pure unit root (I(1)) fails to reject about a 95% of the time. Thus, the power of the test is very close to the size. However, it should be noted that the results vary slightly depending on whether an intercept or trend is added to the regression, and depending on the number of lags, p. The above results represent only a brief glimpse at the properties of the ADF test for STUR processes. Various other experiments were also run, but are not reported here. Perhaps not surprisingly, it turns out that the pure unit root null fails to reject around 95% of the time for σ_η^2 values both larger (i.e. $\sigma_\eta^2 = 0.01$) and smaller than the values reported on. In the next section, a test is discussed which has some power even when $\sigma_\eta^2 = 0.0001$. Overall, these results seem to support the proposition that STUR processes are generic unit root processes and are often not rejected by an ADF unit root test having I(1) as the null.

4. Testing For Stochastic Unit Roots

That ADF tests have low power against some STUR alternatives is not particularly surprising, since the test is well known to exhibit limited power against a variety of "near" I(1) processes. In the STUR case it would be ideal if one were to construct a test which has STUR as the null hypothesis. This is difficult, though, given the rather complex nature of STUR processes, and is not considered in this paper. The approach adopted here has difference stationarity as the null hypothesis, and a random coefficient AR(p+1) model as the alternative. Leybourne, McCabe, and Tremayne (1994) develop the LM type test which we will discuss. Interestingly, the test is also applicable to STUR processes, and is robust to various departures under the null such as ARCH and GARCH. Of note is that the test starts by assuming that a series is non-stationary. In this way, the test has the null of a fixed unit root (or a linear I(1) process) and the alternative of some more general non-stationary data generating mechanism.

The model treated by Leybourne et al. has

$$x_t = \sum_{i=1}^{q+1} \delta_{i,t} x_{t-i} + \varepsilon_t \tag{4.1}$$

$$\delta_{1,t} = (a_t + \pi_1)\,,\ \delta_{i,t} = (\pi_i - a_t\pi_{i-1})\,, i=2,...,q\,,\ \delta_{q+1,t} = -a_t\pi_q\ ,$$

where $a_t \sim$ i.i.d. $(1, \omega^2)$ and $\varepsilon_t \sim$ i.i.d. $(1, \sigma_\varepsilon^2)$, and q is a non-negative integer. Trends may also be incorporated into (4.1), and some dependence is allowed between the a_t and ε_t. Also, note that under the null hypothesis that $\omega^2 = 0$, x_t is seen to be an AR(q+1) process with a unit root, since all of the coefficients sum to unity. The version of (4.1) which we will consider has q=1 and π_1=1:

$$x_t = a_t x_{t-1} + \varepsilon_t\,, \tag{4.2}$$

and is examined by Leybourne, McCabe and Tremayne (1994) and McCabe and Tremayne (1995). In this framework, consider: H_0: $\omega^2 = 0$ against H_1: $\omega^2 > 0$. Before writing the statistic, three features of the test should be mentioned. First, Leybourne et al. show that the test statistic below remains unchanged when the natural log of a_t in (4.2) is allowed to follow a stationary AR(1) process, as in (2.1), (2.2) and (2.3). Second, randomized coefficient models such as (4.1) and (4.2) can be written as models of conditional variation. However, the lagged level of the process itself drives the conditional heteroskedasticity, while in many G(ARCH) formulations it is the innovations that drive the conditional variation. Leybourne et al. have shown that assuming the ε_t are stationary is sufficient to ensure that the test has no

power against G(ARCH). They have also shown that the empirical size of the test in the G(ARCH) case is less than its' nominal value so that the test is conservative. [2] Unfortunately, this result doesn't hold when the form of heteroskedasticity is IGARCH. These and other related issues will be left to future work. Third, it should also be noted that the test is augmented in the same way as ADF tests to allow for various forms of serial correlation in the differences of x_t.

Assuming normality, using the standard conditioning argument, and substituting $\omega^2 = 0$ in the first order condition, yields

$$\partial L(\cdot)/\partial\omega^2 \mid_{\omega^2=0} = \sigma_\varepsilon^{-4} \sum_{t=q+2}^{T} (\sum_{j=1}^{t-1} \varepsilon_j)^2(\varepsilon_t^2 - \sigma_\varepsilon^2) + o_p(1)$$

The statistic is

$$\hat{Z}_T = T^{\frac{-3}{2}} \hat{\sigma}_\varepsilon^{-2} \hat{\kappa}^{-1} \sum_{t=q+3}^{T} (\sum_{j=q+2}^{t-1} \hat{\varepsilon}_j)^2(\hat{\varepsilon}_t^2 - \hat{\sigma}_\varepsilon^2) \ ,$$

where, $\hat{\varepsilon}_t$ is the residual from a regression of Δx_t on a constant, a trend (and lags of Δx_t), with the number of lags, q chosen accordingly (i.e. so that the ε_t are not autocorrelated), $\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum\hat{\varepsilon}_t^2$ and $\hat{\kappa}^2 = \frac{1}{T}\sum(\hat{\varepsilon}_t^2 - \hat{\sigma}_\varepsilon^2)^2$. Leybourne et al. show that the size and power of the statistic is not significantly affected by overfitting. Thus, they suggest that to err on the side of too many lags is better than under-specifying the number of lags. $\hat{Z}_T$ converges weakly to a sum of functionals of generalized Brownian bridges and Brownian motions, given by

$$\int_0^1 G_1(r)^2 dG_2(r) - \int_0^1 G_1(s)^2 ds G_2(1) \ ,$$

where $G_1(r) = W_1(r) - rW_1(1) + 6r(1-r)[W_1(1)/2 - \int_0^1 W_1(s)ds]$, $G_2(r) = \psi W_1 + (1-\psi^2)^{1/2} W_2(r)$,

the limit processes W_1 and W_2 are independent Brownian motions, G_1 is a generalized Brownian bridge, and the nuisance parameter ψ is the covariance between ε_t and ε_t^2. The critical values reported in the footnote of Table 2 are simulated under the assumption that ε_t is symmetrically distributed so that ψ=0. However, the distribution could be tabulated for different values of ψ, and an estimated value used to

---

[2] These finding were pointed out to the authors in correspondence with Steve Leybourne.

calibrate the critical value.

Table 2 contains size and power Monte Carlo results when the alternative is the STUR model specified by (2.1), (2.2) and (2.3). That is, the alternative is (4.1), but with $a_t = exp(\alpha_t)$ and

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

where $|\rho| < 1$, $\eta_t \sim i.i.d.\ N(0,\sigma_\eta^2)$ and is independent of the series ε_t, and $E(a_t) = 1$. Since the test has difference stationarity as the null, it may have power against STUR, as STUR processes are clearly not difference stationary, in theory. As might be expected $\hat{Z}_T$ has very good power against the STUR alternative for larger values of σ_η^2 and for larger sample sizes (i.e. σ_η^2= 0.1 - 0.001, T=500, and 1000). As σ_η^2 falls, the power of the test tails off significantly. The size of the test is also given in Table 1, and corresponds to the case where $\sigma_\eta^2 = 0$. For sample size T=250, the sizes reported are exactly 0.01, 0.05, and 0.10 because critical values were not reported for this sample size in Leybourne et al., and are calculated here using 10000 simulations. Based on the limited Monte Carlo evidence presented in Tables 1 and 2, as well as evidence from a number of other ADF test simulations (see discussion above), it seems that the difference-stationarity test has more power against STUR alternatives than standard ADF tests. For example, given that the ADF and $\hat{Z}_T$ tests are invariant to the scale of the data (σ_ε^2), one may compare the results from Panel (A) in Table 1 with those from Table 2, column 3. For a sample size of 250 observations, the power of the $\hat{Z}_T$ test is fairly low (0.152 for nominal size of 5%), but is still around three times as high as the power of the ADF test. In order to examine the relationship between ADF and $\hat{Z}_T$ tests further, the experiments reported in Table 1 where repeated for T=500. We found that the relative performance of the two tests remains approximately the same as for the T=250 case. Some test results for actual macroeconomic series are given in Section 6. Before discussing a series of forecasting experiments (Section 6), we briefly outline some strategies for estimating STUR processes.

5. Estimation Strategies

Two methods for estimating the parameters of STUR processes are considered. The first is the approximate maximum likelihood method discussed in Guyton, Zhang, and Foutz (1986). In contrast to Guyton et al., though, conditions are not found for the existence of a *stationary* generalized autoregressive process. Rather, the parameters of a nonstationary STUR process are estimated. Assume that

$$x_t = a_t x_{t-1} + \varepsilon_t \ , \quad a_t = \exp(\alpha_t) \quad \text{and} \quad E(a_t) = 1 \quad \text{or} \quad \theta = 0 \ , \tag{5.1}$$

where

$$\alpha_t = \mu + \sum_{i=1}^{p} \rho_i \alpha_{t-i} + \eta_t \ ,$$

and α_t is a stationary stochastic process, thus generalizing (2.3) to be an AR(p) process. (The STUR model may also be extended in the current context by adding lags of x_t, as in (4.1), for example.) Of course, for the sake of estimation, we will not assume that $E(a_t) = 1$ or that $\theta = 0$, as we do not wish to impose that the process must be STURA or STURB according to (2.15) or (2.16). Another possible variation of the above STUR process is the addition of a time varying intercept, say c_t, to the measurement equation. In this way, the upward trending behavior of many economic series could, potentially, be better modeled. As we will see below, such generalizations pose no problem for the approximate maximum likelihood (AML) estimation technique.

Assume that the ε_t and the η_t are normally distributed with zero means, and variances σ_ε^2 and σ_η^2, respectively, and are independent. Thus, a_t is distributed as a lognormal random variable. The unknown parameters are gathered into a vector, $\Omega = (\mu, \rho, \sigma_\varepsilon^2, \sigma_\eta^2)$, where $\rho = (\rho_1, \rho_2, ..., \rho_p)$. In order to construct a likelihood function, use the well known property that $f(x, a \mid \Omega) = f(x \mid a, \Omega) f(a \mid \Omega)$, where $f(\cdot)$, is a joint density. Then, noting that likelihood is proportional to probability, assuming a value of unity for the constant of proportionality, carrying the conditioning argument one step further, and introducing a randomization parameter, K yields

$$L(x \mid \Omega) = \int L(x, a \mid \Omega) da = \int L(x \mid a, \Omega) f(a \mid \Omega) da = E_{a \mid \Omega} L(x \mid a, \Omega)$$

and,

$$\hat{E}_{a \mid \Omega} L(x \mid a, \Omega) = \hat{L}(x \mid \Omega) = \frac{1}{K} \sum_{k=1}^{K} \left(\frac{1}{\sqrt{2\pi}} \right)^T \sigma_\varepsilon^{-T} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^{T} \hat{\varepsilon}_{t,k}^2} \tag{5.2}$$

$$\hat{\varepsilon}_{t,k} = x_t - \hat{a}_{k,t} x_{t-1} \tag{5.3}$$

where by $E_{a \mid \Omega}$ we mean the expectation over a conditional on the parameter Ω, and assuming that the density of x_1 is the same as the conditional density of x_t, given x_{t-1}, and that $x_0 = 0$. This is a standard result, except that the a_t are unobserved so that we construct an estimate, $\hat{L}(\cdot)$ by randomly sampling across different distributions of a_t.

In order to calculate values of the likelihood function for various points in the parameter space, various sequences, $a_k = \{a_{k,t}\}$, are generated for each $\{\mu, \rho_i, i=1,...,p, \sigma_\eta^2\}$ in the parameter space. Thus, the approximate maximum likelihood function given as (5.2) is a random variable, and is an estimator of $E_{a|\Omega}L(x \mid a, \Omega)$. Maximization of (5.2) proceeds by first generating K sequences $\{\eta_{t,k}\}$, t=1,...,T. Then, $\hat{L}(\cdot \mid \cdot)$ and its derivatives are calculated for each point in the parameter space $\hat{\Omega} = (\hat{\mu}, \hat{\rho}_i, i=1,...,p, \hat{\sigma}_\varepsilon^2, \hat{\sigma}_\eta^2)$. This iterative procedure is carried out until a $\hat{\Omega}$ is found which optimizes $\hat{L}(\cdot \mid \cdot)$, according to some convergence criterion. Following accepted techniques from the simulation-based estimation literature, the *same* K sequences, $\{\eta_{t,k}\}$, t=1,...,T, are used throughout the optimization. Approximate maximum likelihood methods are also considered in the disequilibrium literature. For example, Laroque and Salanie (1994) show that simulated pseudo ML estimators have good finite sample properties, even for values of *K* as low as 15. It should be noted that σ_ε^2 cannot be concentrated out of the approximate likelihood function (since the sum of logs is not equal to the log of sums).

The finite sample properties of the AML method were examined using a series of Monte Carlo experiments. STUR processes were generated according to (5.1) with p=1, $\rho_1 = 0.6$, $E(a_t) = 1$, $\eta_t \sim$ *i.i.d.* N(0, 0.0001) and N(0, 0.000001), and $\varepsilon_t \sim$ *i.i.d.* N(0, 1.0). Parameters were estimated using the GAUSS computer program and the Broyden, Fletcher, Goldfarb, Shanno (BFGS) algorithm. Table 3 contains average parameter estimates and standard errors for 1000 trials, for sample sizes of 100, 250, and 500 observations, and for randomization parameter, *K*, values of 10, 25, 100, and 250. A number of results emerge from inspection of Table 3. First, the AML method appears to produce fairly reasonable average point estimates of the STUR parameters, in particular for the larger sample sizes of 250 and 500, and for the larger *K* values of 100 and 250. For example, when *K*=10 and 25 the average point estimates of σ_η^2 are not as close to their true values as when higher randomization parameters are used. Interestingly, selecting *K*=250 does not generally result in substantial improvement over the *K*=100 case. Second, σ_ε^2 is estimated very well, even for small K and T. Finally, it should be noted that the parameters are usually estimated fairly imprecisely, particularly in Panel (B), where a smaller value of σ_η^2 is used. This suggests that AML estimates should be used with caution, in practical applications. Overall, though, the AML

technique is intuitively appealing, works fairly well in simulations, and we feel that the estimation method is worthy of further investigation.

A second estimation technique was also considered, and consists simply of minimizing the sum squared residuals, the $\hat{\varepsilon}_t$, in (5.3). In this formulation, α_t is assumed to be the AR(1) process given by (2.3). Using a procedure very similar to the AML method, K fixed sequences, $\{\eta_{k,t}\}$, t=1,...,T are first generated. Then, K sequences of $\{\hat{a}_{k,t}\}$, t=1,...,T are generated for a point, $\hat{\Omega}$, in the parameter space. From these, a function value

$$\frac{1}{K}\sum_{k=1}^{K}\sum_{t=1}^{T}(x_t - \hat{a}_{k,t}x_{t-1})^2$$

is calculated. This approximate minimum sum squares (AMSS) is calculated for various different $\hat{\Omega}$ values, and some numerical optimization method is used, yielding a final estimate, $\hat{\Omega}^*$. For the AMSS method, a natural estimator for σ_ε^2 is

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\frac{1}{K}\sum_{k=1}^{K}\sum_{t=1}^{T}(x_t - \hat{a}_{k,t}^{*}x_{t-1})^2$$

where $\hat{a}_{k,t}^{*}$ uses the "optimal" estimates, $\hat{\Omega}^*$. Of note is that estimates from the AML and AMSS methods can be used to construct estimates the STURB condition, (2.15), where

$$\hat{sc} = \hat{m} + \tfrac{1}{2}\hat{\sigma}_\alpha^2 = \frac{\hat{\mu}}{(1-\hat{\rho}_1)} + \frac{\hat{\sigma}_\eta^2}{(1-\hat{\rho}_1^2)}$$

Thus, the STUR condition, (2.15), can easily be checked, although we do not provide standard errors for these estimates.

Unfortunately, AML only simplifies to AMSS if either: (i) the randomization parameter, *K*, is unity (which results in wild estimates for obvious reasons); or (ii) $\ln L(x \mid \Omega) \approx E_{a|\Omega}\ln L(x \mid a, \Omega)$ (Jensen's inequality is actually an equality, i.e. $E_{a|\Omega}\ln L(x \mid a, \Omega) = \ln(E_{a|\Omega}L(x \mid a, \Omega))$). For example, it can be shown (using a bit of algebra) that AML reduces to AMSS when the following conditions hold:

$$\frac{1}{K}\left(\frac{\sum_{k=1}^{K} e^{\frac{-1}{2\tilde{\sigma}_\varepsilon^2}\sum_{t=1}^{T}\hat{\varepsilon}_{t,k}^2}\sum_{t=1}^{T}\hat{\varepsilon}_{t,k}^2}{\sum_{k=1}^{K} e^{\frac{-1}{2\tilde{\sigma}_\varepsilon^2}\sum_{t=1}^{T}\hat{\varepsilon}_{t,k}^2}}\right) = \frac{1}{K}\sum_{k=1}^{K}\sum_{t=1}^{T}\hat{\varepsilon}_{t,k}^2 \qquad (5.4)$$

where $\tilde{\sigma}_\varepsilon^2$ is the variance estimator which arises when one attempts to concentrate σ_ε^2 from $L(x \mid \cdot)$, and when terms of the form

$$\left(e^{\frac{-1}{2\tilde{\sigma}_\varepsilon^2}\sum_{t=1}^{T}\hat{\varepsilon}_{t,i}^2} \right) \left(\sum_{t=1}^{T}\hat{\varepsilon}_{t,j}^2 \right) \quad ,i,j = 1, \ \cdots \ ,K \quad ,i \neq j \tag{5.5}$$

are identically zero. [3] Thus, while intuitively appealing, there is no obvious reason why AMSS should provide reasonable estimates. In order to further examine AMSS, a number of simulations analogous to those carried out for the AML method were done. Perhaps not surprisingly, the AMSS method produced wild estimates much of the time, and failed to converge at all for many of the simulations. Thus, at this juncture, we include discussion of AMSS primarily as a curio, acknowledging that further examination of the method may be of some interest, particularly given that it is computationally inexpensive.

Other estimation techniques may also be of some interest for STUR, but for the sake of brevity we have limited our introductory examination to two. Clearly, there is also a lot of other interesting work which remains to be done in this area in the future. For example, state space models which incorporate STUR conditions, and which employ the Kalman filter are of potential interest. In the next section STUR models, which are estimated using AML, are examined from an empirical perspective. One model which is compared to STUR has a_t distributed as an AR(1) process. The model is estimated using the Kalman filter, and differs from the STUR model in that $a_t \sim$ AR(1) rather than $a_t = \exp(\alpha_t)$ with $\alpha_t \sim$ AR(1).

6. Forecasting and Test Results

One clear way to judge the relevance of a model is to ask how well it performs compared to other models when analyzing actual data. In this section the relative forecasting performance of a simple STUR model is compared to a linear autoregressive model with constant parameters, an standard autoregressive model with a time-varying parameter which is estimated using the Kalman filter algorithm, and a random walk model. Of note is that the Kalman filter is a natural estimation technique for time varying parameter models, while AML may be a natural method when the log of the AR parameter follows an AR

---

[3] A full derivation of these conditions is available upon request from the authors, and follows by taking the log of the likelihood in (5.2), and attempting to concentrate out σ_ε^2. This in turn yields conditions which are sufficient to ensure that the AML estimator is the same as the standard OLS estimator (i.e. conditions whereby we can use the standard argument that the likelihood is maximized when the sum of squared residuals is minimized). It is these conditions that are reported above as equations (5.4) and (5.5).

process. Along these lines we emphasize that our analysis focuses on which *model* best represents the data, and does not attempt to compare the performance of the different estimation techniques. Many more sophisticated forecasting models are available for the series, so that the forecast comparison below is meant just as a guide to the usefulness of the four models studied, from the perspective of multi-step ahead forecasting.

The series analyzed are listed in Table 4, Panel (A). The in-sample period is 1955:1-1987:12, and the out-of-sample period - used to evaluate the forecasting models - is 1988:1-1991:10. All data are monthly. The data are a potpourri of monthly U.S. macroeconomic series, and are meant to depict a somewhat wide range of different characteristics which are associated with economic time series.

In practice, model builders often work with log (x_t) and with growth rates, or "rates of return", defined by $r_t = (x_t - x_{t-1})/x_{t-1}$ which are usually very well approximated by $\Delta\log(x_t)$. However, note that if x_t is actually generated by (2.1), (2.2) and is positive then

$$\alpha_t = \Delta\log(x_t) + \log[1 - \varepsilon_t/x_t] \approx \Delta\log(x_t) - \varepsilon_t/x_t \qquad (6.1)$$

As usual, ε_t is thought to be small compared to x_t. It follows that α_t is approximately equal to r_t. Consequently, if r_t is well modeled by a stationary series, as is often assumed, then $\log(x_t)$ will be an exact unit root process and x_t will be approximately STUR. This has implications for the original level variable which are rarely discussed. The original variable will not have a perfect unit root but something like a STUR model is more appropriate. Of course, (6.1) illustrates that estimating STUR processes using growth rates introduces heteroskedasticity into the model. For this reason, we estimate the models in this section using levels data. This is not meant, however, to imply that using log data isn't preferred, and future research should shed more light on the issue. Rather, we keep the current forecasting exercise simple by focusing on STUR models using levels data, and ignore heteroskedasticity in our empirical analysis.

The series were all found to be "generic" I(1), using augmented Dickey-Fuller tests (see Table 5). However, as discussed above, the ADF test has low power against some STUR alternatives. Thus, the series were also tested using the $\hat{Z}_T$ statistic. Table 5 contains test results for the series. The results

suggest that the null of difference stationarity should be rejected for 6 of the 10 series at a 1% level. This suggests that some macro-economic series might be *better* specified when the pure unit root restriction is relaxed, although the caveats mentioned in Section 4 certainly apply. Furthermore, the particular form of non-linear model which is appropriate will naturally vary for each series, and is an currently an area of considerable research in economics. However, since the test considered has some power against STUR, these types of models should be of some interest. [4]

For each of the actual series, forecasting models were formed as follows:

(a) RW: standard random walk with drift models.

(b) AR(p): standard autoregressive models of order p, AR(p), where the order is selected using a BIC criterion and all coefficients, including a constant, are estimated by OLS with no constraints on roots of the model.

(c) Kalman: a time varying parameter model, $x_t = a_t\, x_{t-1} + \varepsilon_t$, where a_t is assumed to evolve according to an AR(1) process is fitted to each series. The parameters of this model are determined in-sample by a grid search procedure. The parameter a_t is estimated recursively by using the Kalman filter algorithm, as described in Anderson and Moore (1979) and in Harvey and Phillips (1982). Two alternative multi-step forecasting procedures were considered and compared, and are discussed in Appendix I. In the comparisons, the more successful technique, called here the "chain-rule" method is used.

(d) STUR: models of the type given in (2.1), (2.2) and (2.3) are fitted, with no constants placed on the parameters, using the AML method discussed above. Parameter estimates are given in Table 4, Panel (B), assuming that α_t follows an AR(1) process. Forecasts using the AMSS estimates are simulation based, analogous to the methods discussed in Section 5. The number of randomization steps, K, is set at 100. The h period ahead forecasts are constructed as

$$x^f_{t+h} \;=\; \frac{1}{K}\sum_{k=1}^{K}\Big(\prod_{i=1}^{h} a^f_{t+i,k}\Big)x_t$$

[4] Tests of non-linearity such as the one suggested above usually have linearity as the null. (See Luukkonen, Saikkonen, and Teräsvirta (1988) and Tsay (1989), for example.) In many cases, though, it may be preferable to specify a specific non-linear model as the null. For the STUR case, this is left to future research.

Two candidate AML forecasting techniques were considered:

(d.1): Use the parameter estimates, $\hat{\Omega}$, to generate K sequences $\{\hat{\varepsilon}_{t,k}\}$ and $\{\hat{\eta}_{t,k}\}$, t=1,...,T. Then construct K sequences $\{\hat{a}_{t,k}\} = \left\{\frac{x_t - \hat{\varepsilon}_{t,k}}{x_{t-1}}\right\}$, and $\{a^f_{t+h,k}\} = \left\{e^{\hat{\mu}}(a^f_{t+h-1,k})^{\hat{\rho}}e^{\hat{\eta}_{t,k}}\right\}$, where $\hat{a}_{t,k}$ is the initial value for the recursion on a^f.

(d.2): Use $\{\hat{\eta}_{t,k}\}$, t=1,...,T to construct K sequences $\{\alpha^f_{t+h,k}\} = \left\{\hat{\mu} + \hat{\rho}\alpha^f_{t+h-1,k} + \hat{\eta}_{t,k}\right\}$, and $\{a^f_{t+h,k}\} = \left\{e^{\alpha^f_{t+h,k}}\right\}$, where the same $\hat{\eta}_{t,k}$ are used for all h in (d.1) and (d.2).

Guyton et al. (1986), alternatively, use Bayes' Rule to generate the conditional distribution of $\{a\}$ given $\{x\}$. The technique is complicated, though, and any simplification relies heavily on an assumption that the a_t are normally distributed. Experimentation using in-sample data indicated that (d.1) yielded reasonable forecasts. For this reason, (d.1) is used as the STUR forecasting method. To simplify our analysis of models (a)-(d), fixed autoregressive parameters are not updated as new sample information becomes available. We leave the examination of evolving AR(p) and STUR models (see Swanson and White (1995), for example) to future work.

In order to reduce possible biases in any of the forecasts, a regression was run (in-sample), of the form

$$x_{t+h} = \alpha_h + \beta_h x^f_{t+h}\,, \tag{6.2}$$

where x^f_{t+h} is one of the four forecasts described above, and α_h, β_h are the *bias adjustment* parameters used to calculate the *final* forecasts. There is no obvious reason to believe that the constant parameter AR(p), or Kalman forecasts are biased, but they were put through the same procedure so that a direct comparison could be made.

To ask if one method is statistically significantly better than another, suppose that at time t, the two forecasting methods produce errors u_t, v_t, from $s_t = u_t + v_t$, $d_t = u_t - v_t$. The null hypothesis to be tested is that var(u) = var(v), which is equivalent to $r = corr\,(s_t\,,\, d_t) = 0$. To perform this particular test, a stan-

dard z-statistic for correlation is used where $z = \frac{1}{2}\log(1+r)/(1-r)$ (Anderson (1984)). Under the null $\sqrt{n-1}\,z$ is asymptotically distributed as a $N(0\,,\,1)$ random variable where n is the post-sample size. A low p-value indicates rejection of the null hypothesis. Diebold and Mariano (1995) discuss various versions of this test.

Table 6 summarizes the forecasting comparison results. For the 1-step, 5-step and 10-step ahead forecast horizons, the first four columns of figures contain out-of-sample mean squared errors, while the last 4 columns contain the results of the comparisons. In particular, "ranks" from 1 (best) to 4 (worst) are assigned to each model for each variable. Total ranks for each model at each horizon are also given, with the lowest corresponding to the "best" score, in a crude sense. Whenever a model was found to be statistically superior to another model, for a given series, a * (5%) or @ (10%) was placed beside the models' rank score. The total number of "wins" based on *s and @s is also given for each model, again offering a crude indication of model superiority.

Various provocative results are apparent. For example, the AR model ties for the lowest number of "wins" at all horizons, except for forecast horizon, h=1, when the AR model "wins" more frequently that STUR. Furthermore, 7 of AR's 16 "wins" are for models of unemployment. This agrees with a hypothesis that unemployment is best modeled as a stationary process, at least based on our limited analysis. For h=1, Kalman has the best "rank" sum, while STUR has the worst rank sum. Furthermore, STUR has no "wins". However, for h=5 and h=10, the results change dramatically. Interestingly, the constant parameter AR model "ranks" worst for both multi-step horizons. At h=5, STUR "ranks" second, behind RW, while for h=10, STUR clearly ranks first. Also, for both multi-step horizons, STUR ranks second in terms of "wins", while RW wins for h=5 and Kalman wins for h=10.

Overall, the results are rather mixed. All of the models appear to offer useful forecasts, at least for some of the series. However, the constant parameter AR model appears to under-perform the other models for most of the series, most of the time. The STUR model, on the other hand, performs very well outside of the one-step ahead case. These STUR results are encouraging, and suggest that further analysis

of such models may be useful. Nevertheless, the issue of why STUR performs so poorly for $h=1$ remains a quandry. More light should be shed on the issue, as more sophisticated estimation and testing techniques become available.

In general, much care needs to be taken when constructing forecasts for *any* given economic time series. Thus, the four forecasting methods considered here are necessarily limited in scope. Further experiments, for instance, might compare STUR forecasts with more complicated alternative forecasts. Nevertheless, the testing and forecasting results of this Section suggest that STUR processes offer a complementary modeling option, and, as such, should be entertained when forecasting economic time series. This is, perhaps, not a surprising result, as it is hard to find many macroeconomic theories which suggest that the *perfect* unit root model is preferable over more general, *near* unit root models.

7. Bivariate Generalizations and Conclusions

If x_t, y_t are a pair of series having a decomposition

$$\begin{aligned} x_t &= Aw_t + \tilde{x}_t \\ y_t &= w_t + \tilde{y}_t \end{aligned} \tag{7.1}$$

where $\tilde{x}_t$, $\tilde{y}_t$ are both stationary and w_t is STUR, then $z_t = x_t - Ay_t$ will be stationary, and a standard cointegration situation occurs in that x_t , y_t will both be extended memory in mean but there exists a linear combination that is stationary. However, a more interesting situation occurs from considering the logs of a pair of economic variables, lx_t , ly_t. If, as is often assumed, the changes of these variables are stationary, then lx , ly have perfect unit roots, and, as sometimes indicated by tests, lx_t and ly_t are cointegrated, so that $lz_t = lx_t - cly_t$ is I(0), then lx_t , ly_t can be thought of as being generated by a (linear) error-correction model

$$\begin{aligned} \Delta lx_t &= \gamma_1 lz_{t-1} + \text{lags}\Delta lx_t \text{ , } \Delta ly_t + u_{xt} \\ \Delta ly_t &= \gamma_2 lz_{t-1} + \text{lags}\Delta lx_t \text{ , } \Delta ly_t + u_{yt} \end{aligned} \tag{7.2}$$

where all terms in this model are I(0). It follows that there is a decomposition

$$lx_t = clw_t + l\tilde{x}_t$$
$$ly_t = lw_t + l\tilde{y}_t$$

when lw_t has a perfect unit root, and the last terms are each I(0). Gonzalo and Granger (1995) discuss how the terms in this decomposition can be estimated from (7.2) giving, essentially $l\tilde{x}_t = (1+c)lz_t$, $l\tilde{y}_t = lz_t$. It is also frequently observed that for cointegration involving the logs of variables, $c = 1$, and this value will be assumed (see Engle and Granger (1987) for an introductory discussion of cointegration and error correction). Taking exponentials gives

$$x_t = \exp(lw_t)\exp(2lz_t)$$

and,

$$y_t = \exp(lw_t)\exp(lz_t)$$

suggesting that, in practice x_t , y_t will be cointegrated with a stochastic parameter, in that $x_t - y_t\exp(lz_t)$ will be I(0). As nothing has been estimated about the properties of lz_t, except that it is stationary and has zero mean, $\exp(lz_t)$ can have virtually any positive expected value. However, if the variance of lz_t is small, then the characteristics of $\exp(lz_t)$ match those of the stochastic unit roots discussed above. Thus, even a procedure as innocuous as finding that two logged series are cointegrated does have potentially *important* implications for the stochastic behavior of the series. Even this simplistic analysis suggests that random coefficient cointegration may be a consequence of the analysis of the logs of variables, in many cases. Of course, the error-correction model (7.2) could also have stochastic coefficients, leading to deeper questions concerning testing for, and defining appropriate forms of cointegration. These questions, and other multivariate extensions, in the area of *stochastic cointegration*, for instance, are the subject of future research.

The results presented in this introduction to stochastic unit root processes are, we believe, sufficiently promising for further discussion and analysis of these and similar processes to occur. They allow the theorist and the econometrician to get away from the very tight constraints on a process imposed by a perfect unit root and to generate data of a more realistic kind. It is seen that STUR models have the potential to provide improved forecasts for macroeconomic time series, at least relative to

AR(p), time varying parameter, and random walk models. Furthermore, a simple test indicates that non-linear models such as STUR processes may be preferred to difference stationary models in many cases.

Future research directions include the estimation of STUR processes, as well as an analysis of the economic theories in which stochastic unit roots arise naturally. Another important area for development is stochastic coefficient cointegration.

References

Andel, J., 1976, Autoregressive series with random parameters, Math. Operationsforsch. u. Statis. 7, 735-41.

Anderson, B.D.O. and J.B. Moore, 1979, Optimal filtering (Prentice-Hall, Englewood Cliffs).

Anderson, T.W., 1984, An introduction to multivariate statistical analysis (John Wiley & Sons, New York).

Box, G.E.P. and G.M. Jenkins, 1970, Time series analysis, forecasting and control (Holden Day, San Francisco).

Brandt, A., 1986, The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients, Advances In Applied Probability 18, 211-220.

Diebold, F.X. and R.S. Mariano, 1995, Comparing predictive accuracy, Journal of Business and Economic Statistics 13, 253-263.

Engle, R.F. and C.W.J. Granger, 1987, Co-integration and error correction: representation, estimation, and testing, Econometrica 55, 251-76.

_____ and _____, 1991, Long-run economic relationships: readings in cointegration (Oxford University Press, Oxford).

Franses, P.H., 1994, A multivariate approach to modeling univariate seasonal time series, Journal of Econometrics 63, 133-151.

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

Gonzalo, J. and C.W.J. Granger, 1995, Estimation of common long-memory components in cointegrated systems, Journal of Business and Economic Statistics 13, 27-35.

Granger, C.W.J., 1993, What are we learning about the long-run?, Economic Journal 103, 107-17.

_____ and J. Hallman, 1991, Nonlinear transformations of integrated time series, Journal of Time Series Analysis 12, 207-224.

_____ and P. Newbold, 1986, Forecasting economic time series (Academic Press, San Diego).

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, 105-15.

Hall, Robert E., 1978, Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence, Journal of Political Economy 96, 971-87.

Harvey, A.C. and G.D.A. Phillips, 1982, The estimation of regression models with time-varying parameters, in M Deistler, E. Fürst, and G. Schwödiauer, eds., Games, Economic Dynamics, and Time Series

Analysis (Physica-Verlag, Wien-W*ü*rzburg) 306-321.

Harvey, A.C., 1989, Forecasting, structural time series models and the Kalman filter (Cambridge University Press, Melksham).

Hamilton, J., 1995, State space models, Chapter 50, pages 3041-3080, Handbook of Econometrics, Vol 4, R.F. Engle and D.L. McFadden, eds., North-Holland, Amsterdam.

Koopmans, L.H., 1974, The spectral analysis of time series (Academic Press, New York).

Laroque, G. and B. Salanie, 1994, Estimating the canonical disequilibrium model: asymptotic theory and finite sample properties, Journal of Econometrics 62, 165-210.

Leybourne, S.J., B.P.M. McCabe and A.R. Tremayne, 1994, Can economic time series be differenced to stationarity?, mimeo, University of Nottingham.

Luukkonen, Rita, Pentti Saikkonen and Timo Teräsvirta, 1988, Testing linearity against smooth transition autoregressive models, Biometrika 75, 491-9.

Maybeck, P.S., 1982, Stochastic models, estimation, and control (Academic Press, New York, NY).

McCabe, B.P.M. and A.R. Tremayne, 1995, Testing a series for difference stationarity, Annals of Statistics 23, 1015-1028.

Osborn, D.R., 1988, Seasonality and habit persistence in a life cycle model of consumption, Journal Of Applied Econometrics 3, 255-266.

Pourahmadi, M., 1986, On stationarity of the solution of a doubly stochastic model, Journal Of Time Series Analysis 7, 123-131.

_____, 1988, Stationarity of the solution of $X_t = A_t X_{t-1} + \varepsilon_t$ and analysis of non-Gaussian dependent random variables, Journal of Time Series Analysis 9, 225-230.

Rosenberg, B., 1973, Random coefficients models: the analysis of a cross section of time series by stochastically convergent parameter regression, Annals of Economic and Social Measurement 2/4, 399-428.

Swanson, N.R. and H. White, 1995, A model selection approach to assessing the information in the term structure using linear models and artificial neural networks, Journal of Business and Economic Statistics 13, 265-275.

Tjøstheim, D.,1986, Some doubly stochastic time series models, Journal of Time Series Analysis 7, 51-72.

Tsay, R.S., 1989, Testing and modeling threshold autoregressive processes, Journal of the American Statistical Society 84, 231-40.

Table 1: Results of Augmented Dickey Fuller Test Monte Carlo Study[1]

Panel (A): σ_η = 0.01

	DF	ADF(1)	ADF(2)	ADF(5)	ADF(10)
CASE 1 ($\beta_0 = \beta_1 = 0$)					
Average $\hat{\beta}_2$	-0.006	-0.007	-0.007	-0.007	-0.007
Average $t_{\hat{\beta}_2}$	-0.345	-0.379	-0.379	-0.372	-0.350
$t_{\hat{\beta}_2} > -1.95$	97.2	96.8	96.8	95.6	95.6
$t_{\hat{\beta}_2} > 1.29$	6.8	6.8	6.8	7.2	6.8
CASE 2 ($\beta_1 = 0$)					
Average $\hat{\beta}_2$	-0.021	-0.021	-0.021	-0.021	-0.021
Average $t_{\hat{\beta}_2}$	-1.510	-1.532	-1.520	-1.508	-1.473
$t_{\hat{\beta}_2} > -2.88$	95.6	95.6	96.0	94.8	96.0
$t_{\hat{\beta}_2} > -0.06$	5.6	5.2	5.2	6.4	6.4
CASE 3 (no constraints)					
Average $\hat{\beta}_2$	-0.040	-0.041	-0.042	-0.043	-0.045
Average $t_{\hat{\beta}_2}$	-2.155	-2.199	-2.199	-2.186	-2.147
$t_{\hat{\beta}_2} > -3.43$	92.8	92.4	92.8	94.1	92.4
$t_{\hat{\beta}_2} > -0.92$	6.4	4.8	4.0	4.4	5.2

Panel (B): σ_η = 0.001

CASE 1 ($\beta_2 = \beta_1 = 0$)					
Average $\hat{\beta}_2$	-0.007	-0.007	-0.007	-0.007	-0.007
Average $t_{\hat{\beta}_2}$	-0.385	-0.395	-0.381	-0.359	-0.340
$t_{\hat{\beta}_2} > -1.95$	96.4	96.4	96.0	96.0	96.4
$t_{\hat{\beta}_2} > 1.29$	6.0	6.8	5.6	6.4	5.6
CASE 2 ($\beta_1 = 0$)					
Average $\hat{\beta}_2$	-0.022	-0.022	-0.022	-0.022	-0.022
Average $t_{\hat{\beta}_2}$	-1.584	-1.596	-1.578	-1.551	-1.522
$t_{\hat{\beta}_2} > -2.88$	94.1	93.6	93.2	94.0	94.8
$t_{\hat{\beta}_2} > -0.06$	4.0	4.0	4.0	3.6	5.2
CASE 3 (no constraints)					
Average $\hat{\beta}_2$	-0.043	-0.044	-0.044	-0.044	-0.046
Average $t_{\hat{\beta}_2}$	-2.248	-2.263	-2.244	-2.211	-2.175
$t_{\hat{\beta}_2} > -3.43$	93.2	91.6	92.0	93.2	94.0
$t_{\hat{\beta}_2} > -0.92$	3.2	3.2	2.8	3.6	5.2

[1] The STUR model estimated is, $x_t = a_t x_{t-1} + \varepsilon_t$, where $a_t = exp(\alpha_t)$, and, $\alpha_t = \mu + \rho\alpha_{t-1} + \eta_t$. The parameter, ρ, is set at 0.6, while μ is fixed so that $E(a_t)=1$. The two error processes are simulated as independent, each being distributed i.i.d. normal with $\sigma^2_\varepsilon = 1.0$ and $\sigma^2_\eta = 0.0001$, and 0.000001, respectively, for Panels (A) and (B). Results are shown for the Augmented Dickey Fuller tests with an intercept and a trend using Δx_t as the dependent variable when the null hypothesis is $x_t \sim I(1)$. The regressions were run using various augmentations (lags) of the dependent variable. The the number of lags used, (p), the average value of the coefficient $\hat{\beta}_2$ in (3.1), and the corresponding average $t_{\hat{\beta}_2}$ value aer all shown. In each case, 10000 simulations were performed. The sample size was set at 250 observations, and the critical values used correspond to a 5% and a 95% cumulative probability under the null hypothesis, and are given in Fuller (1976).

Table 2: The Power of the $\hat{Z}_T$ Test Against STUR Alternatives[1]

T	$\sigma_\eta^2 = 0.0$	$\sigma_\eta^2 = 0.0001$	$\sigma_\eta^2 = 0.001$	$\sigma_\eta^2 = 0.01$	$\sigma_\eta^2 = 0.1$
Panel (A): Nominal Size = 1%					
50	0.009	0.018	0.016	0.093	0.330
100	0.014	0.019	0.058	0.249	0.639
250	0.016	0.041	0.150	0.494	0.852
500	0.011	0.103	0.354	0.813	0.955
1000	0.014	0.213	0.614	0.939	0.987
Panel (B): Nominal Size = 5%					
50	0.052	0.061	0.069	0.187	0.485
100	0.046	0.079	0.135	0.376	0.795
250	0.047	0.121	0.233	0.618	0.926
500	0.056	0.185	0.504	0.892	0.978
1000	0.060	0.356	0.718	0.964	0.993
Panel (C): Nominal Size = 10%					
50	0.100	0.116	0.126	0.252	0.576
100	0.097	0.134	0.186	0.444	0.854
250	0.094	0.188	0.312	0.693	0.949
500	0.106	0.255	0.577	0.921	0.984
1000	0.121	0.427	0.767	0.972	0.995

[1] Results are shown for $\hat{Z}_T = T^{\frac{-3}{2}} \hat{\sigma}^{-2} \hat{\kappa}^{-1} \sum_{t=p+3}^{T} \left(\sum_{j=p+2}^{t-1} \hat{\varepsilon}_j \right)^2 (\hat{\varepsilon}_t^2 - \hat{\sigma}_\varepsilon^2)$. Critical values for sample sizes 50, 100, 500, and 1000 are from Leybourne et al. (1994) and are {.349, .320, .278, .261} for the 1% level, {.215, .192, .161, .149} for the 5% level and {.161, .142, .114, .104} for the 10% level. As the critical values for T=250 are not available, we calculated them using 10000 simulations, under the null hypotheses, $a_t = 1$ for all t. The values are {.289, .168, .122} at the 1%, 5%, and 10% levels respectively. The STUR alternative used in the power simulations ($\sigma_\eta^2 > 0$) is $x_t = a_t x_{t-1} + \varepsilon_t$, where $a_t = exp(\alpha_t)$, $\alpha_t = \mu + \rho\alpha_{t-1} + \eta_t$, and $E(a_t) = 1$. The two error processes are assumed to be independent, each being distributed i.i.d. normal with respective variances, σ_ε^2 and σ_η^2. ρ=0.60, and σ_ε^2= 0.01, 0.1, 1.0, 10.0 for σ_η^2= 0.0001, 0.001, 0.01, 0.1, respectively. The parameter μ is fixed so that $E(a_t) = 0$. 10000 simulations were used for the Monte Carlo experiment.

Table 3: AML Estimates for Various Randomization Steps, K, and Sample Sizes[1]

T	K	$\hat{\mu}$	$\hat{\rho}$	$\hat{\sigma}_\eta$	$\hat{\sigma}_\varepsilon^2$
Panel (A): σ_η = 0.01					
100	10	-0.0038	0.5952	0.0084	0.9909
		(0.0067)	(0.2102)	(0.0098)	(0.0724)
	25	-0.0086	0.5993	0.0092	0.9966
		(0.0136)	(0.0017)	(0.0051)	(0.0189)
	100	-0.0047	0.6004	0.0117	0.9955
		(0.0092)	(0.0188)	(0.0088)	(0.0275)
	250	-0.0057	0.6006	0.0111	0.9955
		(0.0131)	(0.0226)	(0.0077)	(0.0369)
250	10	-0.0026	0.6005	0.0092	0.9937
		(0.0054)	(0.2371)	(0.0094)	(0.0395)
	25	-0.0027	0.6050	0.0099	1.0026
		(0.0046)	(0.0320)	(0.0080)	(0.0121)
	100	-0.0025	0.6018	0.0100	0.9993
		(0.0041)	(0.0159)	(0.0035)	(0.0090)
	250	-0.0024	0.5996	0.0101	1.0000
		(0.0053)	(0.0016)	(0.0036)	(0.0125)
500	10	-0.0013	0.5855	0.0069	1.0129
		(0.0031)	(0.2337)	(0.0052)	(0.0349)
	25	-0.0013	0.5828	0.0087	1.0031
		(0.0022)	(0.1210)	(0.0042)	(0.0281)
	100	-0.0013	0.6001	0.0099	1.0000
		(0.0025)	(0.0041)	(0.0031)	(0.0074)
	250	-0.0010	0.5967	0.0098	1.0002
		(0.0025)	(0.0327)	(0.0026)	(0.0044)
Panel (B): σ_η = 0.001					
100	10	-0.0030	0.5678	0.0053	0.9955
		(0.0072)	(0.3007)	(0.0065)	(0.0721)
	25	-0.0034	0.5910	0.0029	1.0027
		(0.0082)	(0.1352)	(0.0046)	(0.0547)
	100	-0.0039	0.5970	0.0020	0.9919
		(0.0086)	(0.1515)	(0.0043)	(0.0302)
	250	-0.0040	0.6058	0.0018	0.9942
		(0.0070)	(0.0353)	(0.0035)	(0.0340)
250	10	-0.0027	0.6353	0.0056	0.9989
		(0.0047)	(0.1160)	(0.0083)	(0.0257)
	25	-0.0025	0.6006	0.0042	0.9963
		(0.0045)	(0.0341)	(0.0069)	(0.0239)
	100	-0.0023	0.6089	0.0019	0.9994
		(0.0049)	(0.0560)	(0.0037)	(0.0148)
	250	-0.0021	0.6049	0.0016	0.9956
		(0.0044)	(0.0245)	(0.0029)	(0.0258)
500	10	-0.0012	0.5818	0.0035	0.9996
		(0.0027)	(0.1694)	(0.0057)	(0.0236)
	25	-0.0015	0.5687	0.0016	0.9968
		(0.0025)	(0.2096)	(0.0023)	(0.0177)
	100	-0.0011	0.6015	0.0018	0.9992
		(0.0021)	(0.0604)	(0.0034)	(0.0082)
	250	-0.0012	0.6043	0.0012	0.9986
		(0.0026)	(0.0294)	(0.0012)	(0.0049)

[1] The parameters listed are averages based on 1000 simulations using the approximate maximum likelihood method discussed above. Bracketed values are standard errors across replications. The data is simulated using the STUR model: $x_t = a_t x_{t-1} + \varepsilon_t$, where $a_t = exp(\alpha_t)$, and, $\alpha_t = \mu + \rho\alpha_{t-1} + \eta_t$. The parameter, ρ, is set at 0.6, while μ is fixed so that $E(a_t)=1$. The two error processes are simulated as independent, each being distributed i.i.d. normal with $\sigma_\varepsilon^2 = 1.0$ and $\sigma_\eta^2 = 0.0001$, and 0.000001. Sample sizes of T=100, 250, and 500 are considered.

Table 4: Series Definitions, Summary Statistics and STUR Estimates[1]

Panel (A): Series Definitions

Series Used		Description and CITIBASE Mnemonic
Money Stock - M2	(FM2)	Money Stock: M2, Billions 1982 (FM2D82)
Stock Price Index	(SPI)	S&P's Common Stock Price Index * (FSPCOM)
Long Interest Rate	(RL)	US Treasury Composite 10 yr +, %/yr. * (FYGL)
Short Interest Rate	(RS)	US T-Bills Auction Avg., 3-mo., %/yr. * (FYGN3)
Medium Interest Rate	(RM)	US Treasury Cons Maturity, 1-yr., %/yr. * (FYGT1)
Industrial Production	(IP)	Industrial Production, total index (IP)
Nominal Inventory	(NI)	Mfg. & Trade Inventories: Tot, Billions (IVMT)
Unemployment Rate	(UR)	Unemp. Rate. All Workers 16+ yrs. (LHUR)
CPI - Wage & Cler.	(CPI)	CPI all wage earners and cler., all items (PRNEW)
Prod. Price Index	(PPI)	PPI: all commodities * (PW)

Panel (B): Summary Statistics and STUR Parameter Estimates

Series	Mean	Variance	$\hat{\mu}$	$\hat{\rho}$	$\hat{\sigma}_\eta^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{sc}$
FM2	1581	196651	0.0026	-0.846	0.0001	70.0	0.0014
SPI	103	2809	-0.0036	0.025	0.0586	9.92	-0.0019
RL	6.58	8.33	-0.0012	-0.302	0.0001	0.07	-0.0008
RS	5.77	9.40	-0.0060	-0.328	0.0004	0.33	-0.0043
RM	6.39	10.3	-0.0040	-0.169	0.0003	0.28	-0.0032
IP	64.2	424.6	0.0002	-0.992	0.0000	0.38	0.0001
NI	279	41888	0.0002	0.960	0.0019	7.46	0.0049
UR	5.97	2.68	-0.0009	-0.376	0.0000	0.05	-0.0006
CPI	54.3	826	0.0052	-0.184	0.0000	0.10	0.0043
PPI	55.4	793	0.0048	-0.902	0.0001	0.17	0.0025

[1] All variables are monthly with in-sample: 1955:1-1987:12 and out-of-sample: 1988:1-1991:10, and are seasonally adjusted unless marked with an asterisk. Means and variances are calculated using the entire sample. The STUR parameter estimates are based on the approximate maximum likelihood method discussed above, with starting values gotten from the minimization of $\sum_{k=1}^{K}\sum_{t=1}^{T}\hat{\varepsilon}_{k,t}^2$, where $x_t = a_t x_{t-1} + \varepsilon_t$, and $a_t = exp(\alpha_t)$, $\alpha_t = \mu + \rho\alpha_{t-1} + \eta_t$. In order to account for potential trending behavior, the series were demeaned before estimating the STUR parameters tabulated above. To facilitate an ex-post forecast comparison, STUR estimates are based on the sample 1955:1-1987:12. The STURA condition, (2.15), is estimated as $\hat{sc} = \hat{\mu}(1-\hat{\rho})^{-1} + \frac{1}{2}\hat{\sigma}_\eta^2(1-\hat{\rho}^2)^{-1}$.

Table 5: Test of Monthly Series for STUR[1]

	STUR Tests		*ADF Unit Root Tests*		
Series	Lags of Δx_t	$\hat{Z}_T$	p	$\hat{\beta}_2$	$\hat{t}_2$
FM2	9	0.204*	9	-0.014	-3.06
SPI	4	0.439**	5	-0.017	-1.95
RL	12	0.763**	8	-0.024	-2.48
RS	6	1.386**	7	-0.026	-1.78
RM	13	1.075**	7	-0.026	-2.02
IP	1	0.041	1	-0.027	-2.99
NI	13	0.511**	10	-0.002	-1.16
UR	12	-0.017	4	-0.024	-3.08
CPI	9	0.667**	10	-0.001	-1.67
PPI	12	0.167	10	-0.003	-1.85

[1] Sample is 1955:1-1987:12 for all series. All data is monthly (see Table 6.1). The 2nd and 3rd columns report STUR test results. Lags of Δx_t denotes the order p of the autoregression in Δx_t used to 'whiten' the original series. The order, q, was chosen by analysis of the first 15 estimated autocorrelations using the region $(-2/T^{1/2}, +2/T^{1/2})$. ** denotes rejection of the difference stationary null hypothesis at the 1% level, and * denotes rejection at the 5% level using the critical values of 0.279 and 0.171 from Table 2. The test assumes that the series are non-stationary, and has H_0: difference stationarity, H_A: STUR. All series were previously found to be "generic" I(1), as evidenced by examining the last 3 columns in the table. In particular, the null hypothesis of I(1) fails to reject for all ten series, where the ADF test statistics were calculated from the regression model given as (3.1) above, as were the number of augmentations, p.

Table 6: Forecasting Comparison with Bias Adjustment[1]

Panel (A): One-Step Ahead Forecasts

Series	Mean Squared Forecast Errors				Ranks and Significance Tests			
	RW	AR	Kalman	STUR	RW	AR	Kalman	STUR
FM2	108	55.7	55.9	199	3*	1**	2**	4
SPI	114	126	107	113	3	4	1@	2
RL	0.043	0.040	0.040	0.044	2	1	1	3
RS	0.043	0.041	0.030	0.062	3*	2*	1**@	4
RM	0.074	0.080	0.060	0.080	2	3	1***	3
IP	0.421	0.351	0.351	1.095	3*	1*	1*	4
NI	20.1	15.1	14.7	19.5	4	2	1*	3
UR	0.017	0.020	0.030	0.020	1*	2*	3	2
CPI	0.100	0.082	0.093	0.110	3	1	2	4
PPI	0.807	0.860	0.740	1.570	2*	3*	1*	4
Rank Sum					26	20	14	33
No. Wins					6	6	12	0

Panel (B): Five-Step Ahead Forecasts

FM2	1614	1006	1086	1678	3	1**	2**	4
SPI	822	1091	1077	825	1@	4	3	2@
RL	0.251	0.300	0.250	0.240	2@	4	2*	1*
RS	0.660	0.902	0.643	0.710	2**	4	1*	3*
RM	0.701	0.771	0.740	0.710	1*	4	3	2@
IP	4.18	4.03	4.01	4.51	3	2	1	4
NI	299	234	252	247	4	1@	3*	2
UR	0.141	0.091	0.502	0.150	2@*	1***	4	3*
CPI	0.770	1.36	1.09	0.810	1@*	4	3	2*
PPI	6.07	8.70	6.82	5.76	2*	4	3*	1*
Rank Sum					21	29	25	24
No. Wins					10	6	6	7

Panel (C): Ten-Step Ahead Forecasts

FM2	4191	2940	2742	4093	4	2	1**	3
SPI	1162	1366	1284	1291	1	4	2	3
RL	0.450	0.570	0.390	0.370	3*	4	2@*	1**
RS	2.08	1.87	1.91	1.89	4	1	3*	2
RM	1.97	1.78	2.01	1.70	3	2*	4	1*
IP	6.17	7.72	6.39	6.26	1@	4	3	2
NI	1522	1026	674	725	4	3	1**	2*
UR	0.371	0.302	0.520	0.384	2	1***	4	3
CPI	1.82	3.31	1.55	1.07	3@	4	2*	1@*
PPI	7.54	15.2	8.88	7.16	2*	4	3*	1*
Rank Sum					27	29	25	19
No. Wins					4	4	9	7

[1] The mean squared forecast errors (MSFEs) are reported for the out of sample period, 1988:1-1991:10. Forecasting models were constructed using the period, 1955:1-1987:12. RW corresponds to a random walk model, AR to a AR(p) model (with p chosen using a BIC), Kalman to a state space model estimated using the Kalman filter (see Appendix I), and STUR is a STUR model with an AR(1) process specified for the stochastic unit root part of the process. STUR and random walk models were estimated with a drift, while AR and Kalman models included constants. The number of *s (5% level) and @s (10% level) indicate how many models were outperformed by the given model using a standard z-test for correlation, where $z = \frac{1}{2}\log(1+r)/(1-r)$. Under the null hypothesis that $var(u) = var(v)$, $\sqrt{n-1}\,z \sim N(0\,,\,1)$ where n is the ex-post sample size, and u and v are the out-of-sample forecast errors for the two forecasting models being compared. More than one * (@) indicates that the particular model was superior to more than one other model. For example, in the 5-step ahead case (see Panel (B)), the AR model performs significantly better than each of the other three models, at a 5% level, since three *s are associated with the 1st ranked AR model. Ranks range from 1 to 4, where 4 is associated with the highest MSFE and 1 corresponds to the lowest MSFE. The Rank Sum indicates the total rank score for all of the series forecast. The No. Wins indicates the total number of times that each model was significantly "better" than some other model.

Appendix I

The estimation of the time-varying regression model in this paper uses a version of the model developed by Harvey and Phillips (1982).

$$x_t = \bar{a}_1 + a_t x_{t-1} + \varepsilon_t \tag{A.1}$$

$$a_t - \bar{a}_2 = \rho(a_{t-1} - \bar{a}_2) + \eta_t \tag{A.2}$$

When the normality assumption is added, (A.1),(A.2) is known as a conditionally Gaussian model. This is because the estimators constructed using the Kalman filter are the mean and covariance of the state vector conditioned on information available at time t-1. For a detailed explanation of the Kalman filter see Harvey (1989) and Hamilton (1995). The Kalman filter is applied to the following state-space representation of (A.1) and (A.2):

$$x_t = z'_t \alpha_t + \varepsilon_t \tag{A.3}$$

$$\alpha_t = \Phi\alpha_{t-1} + e_3\eta_t \tag{A.4}$$

where $z'_t = (1\ x_{t-1}\ x_{t-1})$, $\alpha'_t = (\bar{a}_{1t}\ \bar{a}_{2t}\ \delta_{2t})$, $\Phi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \phi \end{bmatrix}$, $\delta_{2t} = a_t - \bar{a}_{2t}$, $e_3 = (0\ 0\ 1)'$, and α_0 is distributed as $N(a_0, P_0)$. The disturbances are assumed to be serially uncorrelated. For further details see Harvey and Phillips (1982). The one-step ahead forecast of x_t using (A.3) and (A.4) is

$$E(x_{t+1} \mid x_t) = z_{t+1} E(\alpha_{t+1} \mid x_t)$$

That is, $E(x_{t+1} \mid x_t) = \bar{a}_1 + (c + \rho E(a_{t+1} \mid x_t))x_t$ where (A.2) can be written as $a_t = c + \rho a_{t-1} + \eta_t$. This is very similar to the one-step ahead AR forecast; and a natural multi-step predictor is:

$$\hat{E}(x_{t+h} \mid x_t) = \bar{a}_1 + \bar{a}_1 \hat{E}(a_{t+1} \mid x_t) + \cdots + \bar{a}_1 \hat{E}(a_{t+1}^{h-1} \mid x_t) + \hat{E}(a_{t+1}^{h} \mid x_t) x_t \tag{A.5}$$

This method is here called the conditional mean forecast, even though $E(a_{t+1}^h \mid x_t)$ is replaced with $(E(a_{t+1} \mid x_t))^h$. Another technique is to recursively substitute previous forecasts and to form:

$$\hat{E}(x_{t+h} \mid x_t) = \hat{E}(z_{t+h} \mid x_t)\Phi^h \alpha_t \tag{A.6}$$

where $\hat{E}(z_{t+h} \mid x_t) = (1\ \ \hat{E}(z_{t+h-1} \mid x_t)\Phi^{h-1}\alpha_t\ \ \hat{E}(z_{t+h-1} \mid x_t)\Phi^{h-1}\alpha_t)$, h>1. This method is similar to the chain rule method in Harvey (1989).

Both multi-step forecasting techniques are clear approximations of the true conditional mean; and as such, were subjected to the bias adjustment discussed in Section 5. The bias adjusted forecasts exhibited lower MSFEs then their unadjusted counterparts at least two thirds of the time using both methods and for h=5,10. Further, the chain rule method resulted in lower MSFEs than the conditional mean method for 12 of the 18 monthly series for both h=5 and h=10.

Appendix II

Table A.1: Augmented Dickey-Fuller Unit Root Tests[1]

Series	p	$\hat{\beta}_2$	$\hat{t}_2$
FM2	9v-0.014v-3.66		
SPI	5v-0.017v-1.95		
RL	8v-0.024v-2.48		
RS	7v-0.026v-1.78		
RM	7v-0.026v-2.02		
IP	1v-0.027v-2.99		
NI	10v-0.002v-1.16		
UR	4v-0.024v-3.08		
CPI	10v-0.001v-1.67		
PPI	10v-0.003v-1.85		

[1] The null hypothesis of I(1) fails to reject for all ten series. The test statistics were calculated from the regression model given as (3.1) above, as were the number of augmentations, p.