Simulation and Prediction Evidence On the Usefulness of Seasonal Unit Root Models

Norman R. Swanson[1,2] and Richard Urbach[2]
[1] Rutgers University
[2] DFA Capital Management, Inc.

May 2005

Abstract

In this paper, we provide new evidence on the empirical usefulness of seasonal unit root (SUR) models. This is done using a series of simulation and prediction experiments, as well as via discussion of the stochastic properties of SUR models. Additionally, we summarize a new testing methodology for comparing the simulated distributions of multiple alternative models. Two key features of the testing methodology are that all models are assumed to be (possibly) misspecified, thus allowing for the fact that all models should be viewed as approximations of some underlying unknown data generating process, and that use of the block bootstrap allows for construction of asymptotically valid critical values that properly account for the effects of parameter estimation error, simulation error, and the time series structure of the data. Based on a simulation comparison of periodic autoregression (PAR), deterministic seasonality, SUR, and random walk models, we find that SUR models perform very poorly when used to construct simulated distributions, in all cases other than when S^{th} (log) differences are formed, where S is the number of seasons. On the other hand, PAR models perform very well under all standard transformations of economic variables. In the context of ex-ante prediction, SUR models yield the lowest mean square forecast error in only 5 of 156 times, when used to forecast a variety of economic variable at various forecast horizons. Furthermore, the SUR model *never* "wins" when used to predict either 1st or 12th (log) differences, even though the data used are measured at a monthly frequency, so that S=12. Finally, all of the above empirical findings are shown to be consistent with theoretical properties of SUR models. We thus conclude that some varieties of SUR models have serious theoretical and empirical shortcomings, and may fall within the "empty box" category discussed in Granger (1999).

JEL classification: C13, C22, C52, C53.
Keywords: seasonal unit root, periodic autoregression, difference stationary, prediction, simulation.

* Norman R. Swanson (nswanson@econ.rutgers.edu): Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA and DFA Capital Management, Inc. 100 Manhattanville Road, Suite 11, Purchase, NY 10577. Richard Urbach (ru@dfa.com): DFA Capital Management, Inc. 100 Manhattanville Road, Suite 11, Purchase, NY 10577. The authors are grateful to John Chao, Valentina Corradi and Eric Ghysels for comments made during discussions on some of the findings in this paper. The authors have also benefited from the excellent research assistance provided by Geetesh Bhardwaj.

1 Introduction

One of the most important characteristics of econometric models is that they are generally viewed as approximations to some unknown underlying data generating process. Furthermore, the extent to which these "approximations" are adequate is often assessed by evaluation of their prediction and simulation properties. For example, insurance companies and banks are often interested in simulating future economic scenarios in order to evaluate measures of capital adequacy, possibly under a variety of alternative starting conditions, product pricing rules, and regulatory environments. Governments, on the other hand, often utilize macroeconometric forecasting models in order to predict economic variables of interest, such as interest rates and inflation, both of which play a crucial role in policy analysis. Our objective in this paper is to assess the adequacy of a number of simple seasonal unit root type models, in terms of prediction and simulation. This is done in part via the use of well known econometric tools, and in part via the use of new tests that have been recently developed to assess model simulation performance. Our assessment, in turn, will be used to shed light upon the question of whether seasonal unit root (SUR) models should fall into the "empty box" category discussed by Granger (1999).[1]

Since the seminal paper of Hylleberg, Engle, Granger and Yoo (HEGY: 1990), there have been numerous papers and books written on the topic of seasonal unit roots. From an empirical perspective, these have been largely concerned with documenting the prevalence of seasonal unit roots in the data, and evaluating the predictive performance of seasonal unit root models. Important contributions in this area include Beaulieu and Miron (1993), Franses (1996), Hylleberg (1994), Miron (1996), Miron and Beaulieu (1996), Franses, Hoek and Paap (1997), Paap, Franses, and Hoek (1997), Franses and Vogelsang (1998), Osborn, Heravi and Birchenhall (1999), Ghysels and Osborn (2001), Franses and Paap (2002), Osborn (2002), Franses and van Dijk (2005), and the references cited therein. The evidence presented in these papers is mixed. For example, in a seminal paper on the subject, Beaulieu and Miron (1993) find very little if any evidence of seasonal unit roots in U.S. data. Osborn et al. (1999), on the other hand, find evidence that predictions based on seasonal differences are at least as accurate as those based on models with deterministic seasonality. Additionally, most of the papers above contain evidence based on SUR tests and on

[1]Granger (1999) presents a variety of theoretical and empirical evidence suggesting that many fractional I(d) processes have stochastic properties which simply do not mimic the dynamical properties of most historical data.

point predictions, with no discussion of simulation properties. In this paper we add to the discussion by: constructing a wide variety of SUR tests; carrying out ex-ante prediction experiments for various data transformations and forecast horizons; and carefully examining the performance of various models when they are used to form simulated future economic scenarios and associated simulation based distributions. Given that our simulation distributions are constructed using simulations initialized at time period T, say, and use models estimated with data up until time period T, they can be viewed as predictive densities for periods beyond T, our simulation experiments can be viewed as generalizing the point prediction results of previous studies to the case of predictive density evaluation of SUR models.

The test that we discuss in this paper for assessing model simulation performance is based on a measure of "goodness of fit" of econometric models that is related to standard notions of Kolmogorov distance. However, in contrast to most Kolmogorov type testing approaches, our approach is not based on the common practice of testing for the correct specification of aspects of a given candidate model. Instead, we evaluate the overall distributional fit of our candidate models, assuming that all models are potentially misspecified, as elucidated in Corradi and Swanson (2004a, 2005a).[2] To be more precise, the testing approach begins by fixing a given econometric model as

[2]To illustrate the importance of allowing for misspecification, assume that we are interested in testing whether the conditional distribution of $y_t|y_{t-1}$ is $N(\alpha_1^{\dagger}y_{t-1}, \sigma_1)$, say. Suppose also that in "true" information set, $\Im_{t-1}$, includes both y_{t-1} and y_{t-2}, and the true conditional model is $y_t|\Im_{t-1} = y_t|y_{t-1}, y_{t-2} = N(\alpha_1 y_{t-1} + \alpha_2 y_{t-2}, \sigma_2)$, where $\alpha_1^{\dagger}$ differs from α_1. In this case, we have correct specification with respect to the information contained in y_{t-1}; but we have dynamic misspecification with respect to y_{t-1}, y_{t-2}. Even without taking account of parameter estimation error, the critical values obtained assuming correct dynamic specification are invalid, thus leading to invalid inference. Stated differently, tests that are designed to have power against both uniformity and independence violations (i.e. tests that assume correct dynamic specification under H_0) will reject; an inference which is incorrect, at least in the sense that the "normality" assumption is *not* false. In summary, if one is interested in the particular problem of testing for correct specification for a given information set, then our approach is appropriate. This is relevant in our case, as all models are assumed to be approximations, so that we indeed have only a particular conditioning information set available, and we wish to find the "best" model, given that information set. Assuming correct specification under the null additionally leads to various related inferential problems. For example, if a sequence of two Kolmogorov-Smirnov type conditional distributional tests of the Andrews (1997) variety are carried out, both of which assume correct specification under the null (as is standard practice), then in additional to the usual sequential test bias problems, one test will have an invalid limiting distribution using standard procedures, as both models surely should not be assumed to be correctly specified under their respective null hypotheses. At most one of them can be true; and in general, both null models are likely misspecified. Thus, the notion of allowing for misspecification of the null

the "benchmark" model, against which all "alternative" models are to be compared. Test statistics are then formed via comparison of the empirical distribution of the historical series and that of the simulated series. These statistics can be viewed as distributional analogs of the mean square error based statistical tests discussed in Diebold and Mariano (1995) and White (2000). The limiting distribution of the test statistics is a Gaussian process with a covariance kernel that reflects the contribution of parameter estimation error, the effect of (possible) dynamic misspecification, and simulation error. This limiting distribution is thus not nuisance parameter free, so that critical values cannot be tabulated. In order to obtain valid asymptotic critical values, we summarize two block bootstrap procedures that depend on the relative rate of growth of the actual and simulated sample sizes, and that are robust to misspecification.

The approach taken in this paper is to consider four very simple econometric models and to examine thirteen different economic variables measuring seasonally unadjusted economic activity. The models used are a seasonal unit root (SUR) model, a deterministic seasonality (DS) model, a periodic autoregression (PAR) model, and a strawman random walk (RW) model, possibly with drift.[3] Simulation models and prediction models are then calibrated. In a series of simulation experiments, we then evaluate the distributional characteristics of data simulated using the four different model types. Thereafter, a series of prediction experiments is carried out in which estimation is done recursively, so that sequences of *ex ante* predictions at various forecast horizons are constructed, including 1-month, 1-quarter, 1-year, and 5-years ahead.

Our findings can be summarized as follows. There is moderate evidence of seasonal unit root and PAR structure in our variables, based upon application HEGY seasonal unit root tests and Boswijk and Franses (1996) PAR tests. However, when estimated SUR models are used to simulate future economic scenarios, the volatility of the simulated data increases rapidly as the length of the simulation period is increased, for the case where 1st differences are simulated. For example, first log difference volatility increases around 400% when comparing simulations of length 60 periods with simulations of length 1200 periods, regardless of which of our thirteen economic variables are used for model calibration. This is clearly a very serious shortcoming of the SUR model, when

hypothesis as well as the alternative hypothesis carries into testing scenarios other than that discussed in this paper.

[3]For an exhaustive discussion of the seasonality models considered in this paper, see the books by Hylleberg (1992), Franses (1996), Miron (1996), and Ghysels and Osborn (2001). For an interesting discussion on the usefulness of seasonal adjustment, see Franses (2001).

comparing with other models. Namely, the other models, such as the RW and DS models are shown to yield reasonable simulated observations, regardless of data transformation. As discussed below, this shortcoming of the SUR models arises naturally, given the structure of the models, and is in part due to the fact that the 12th (log) difference SUR models considered in this paper can be viewed as a sum of 12 different random walks in first differences, all of which should be expected to drift infinitely far apart as the simulation period increases. This feature of SUR models suggests that they can only be expected to reproduce reasonable dynamics for the precise variety of data with which it is calibrated (e.g. 12th (log) difference data when the data are monthly).

Even though the above discussion suggests that the SUR model is limited in its ability to mimic distributional behavior of standard transformations of economic variables, it still may be the case that the SUR model is useful for prediction. This turns out not to be the case, however. For example, in our prediction experiments there are 156 variable/forecast horizon permutations. Across all of these, the SUR model yields the lowest mean square forecast error only 5 times. The PAR model yields the lowest MSFE 23 times, the DS model "wins" 35 times, and the random walk and random walk with drift models win the remaining 88 times (approximately 45 times each). Thus, from the perspective of prediction, the SUR model is inferior at all horizons and for all transformations, including 12th (log) differences. Indeed, the SUR model *never* "wins" when used to predict either 1st or 12th (log) differences. While it is not surprising that the SUR model "loses" for 1st (log) differences, it is interesting that it also loses whenever 12th (log) differences are predicted. Indeed, this fact, taken together with our finding that DS and PAR models win numerous times, suggest that any seasonality present in the data is not of the SUR variety. Not only does the SUR model yield poor simulation results for anything other than 12th (log) differences, but it yields poor predictions for every transformation, *including* 12th (log) differences. Thus, our main conclusion is that some varieties of SUR models may fall within the "empty box" category, based on their ability to mimic the stochastic properties of economic variables, and based on their performance in finite sample prediction and simulation experiments.

The rest of the paper is organized as follows. In Section 2, we outline our empirical methodology and elucidate some relevant analytical properties of SUR models. In Section 3, we describe the data used, and summarize our findings based on a series of simulation and prediction experiments. Concluding remarks are given in Section 4.

2 Empirical Methodology

2.1 Models and Diagnostic Tests

There are a plethora of models from whence the "best" model for a particular time series can be chosen. These include both linear and nonlinear models, parametric and nonparametric models, and stochastic and deterministic models, for example. However, there is ample empirical evidence that parsimonious (linear) models often yield more accurate predictions than more complex (nonlinear) models. Our approach in this paper adheres to the notion that parsimonious linear models are often adequate for prediction and simulation, as we consider a group of the most widely used linear models, and check whether any of these outperform SUR models when used for simulation and prediction. If we find evidence that predictions from seasonal unit root model are less accurate, say, then we have direct evidence against the usefulness of SUR models. Whether such evidence generalizes to the case of more complex (nonlinear) models is left to future research.

In the sequel we shall consider models of the form:

I. $\triangle_1 y_t = \theta_0 + \Gamma(L)\triangle_1 X_t + \varepsilon_t$ – *Random Walk (RW) Model*

II. $\phi(L)\triangle_1 y_t = \Sigma_{s=1}^{S}\theta_s d_{s,t} + \Gamma(L)\triangle_1 X_t + \varepsilon_t$ – *Difference Stationary (DS) Seasonal Model*

III. $\phi(L)\triangle_S y_t = \theta_0 + \Gamma(L)\triangle_S X_t + \varepsilon_t$ – *Seasonal Unit Root (SUR) Model*

IV. $y_t = \theta_s + \Sigma_{i=1}^{p}\theta_{i,s}y_{t-i} + \Sigma_{i=1}^{p_1}\gamma_{1,i,s}x_{1,t-i} + ... + \Sigma_{i=1}^{p_k}\gamma_{k,i,s}x_{k,t-i} + \varepsilon_t$ – *Periodic Autoregression (PAR) Model*

In these models: y_t is the scalar "target" variable; $\triangle_1$ denotes the 1^{st} difference operator (i.e. $(1-L)$); $\triangle_S$ denotes the S^{th} difference operator, where S is the number of seasons presumed to be in the data (i.e. $\triangle_S$ is the seasonal difference operator); $\phi(L) = (1 - \phi_1 L - ... - \phi_p L)$ is a standard lag polynomial of order p, expressed using the lag operator, L; the θ_s denote seasonal intercepts, with associated and conformably defined dummy variables, $d_{s,t}$; $X_t = (x_{1,t}, ..., x_{k,t})'$ are exogenous regressors, $\Gamma(L)$ is a conformably defined vector of lag polynomials of orders $p_1, ..., p_k$ which is associated with the elements of X_t.

Focusing on the univariate properties of these models (i.e. ignoring the X_t), notice that the PAR model generalizes the DS model, as it allows the intercept *and slope* coefficients to vary according to the season. Notice also that both of these models generalize the RW model, which is used as a "strawman" model in our analysis. In all cases, two versions of the random walk were fitted, one with $\theta_0 = 0$ and one with $\theta_0 \neq 0$.

Models were estimated using least squares. For example, note that under quite general conditions, the parameters of the PAR model can be readily estimated via least squares. In particular, consider the PAR model with $p = 1$ (i.e. the PAR(1) model) and $p_1 = ... = p_k = 0$. This model can be estimated via least squares using the equation: $y_t = \Sigma_{s=1}^{S}\theta_s d_{s,t} + \Sigma_{s=1}^{S}\theta_{1,s}d_{s,t}y_{t-1} + \varepsilon_t$. Here, under error normality and given fixed starting values, the least squares estimator is the maximum likelihood estimator, and standard asymptotics pertain (see Franses and Paap (1999) and Franses and van Dijk (2005) for further details). Notice that the PAR(1) model has a unit root when $\theta_{1,1}\theta_{1,2}\theta_{1,3}\cdots\theta_{1,s} = 1$. Clearly, the PAR(1) model nests the simple random walk model where $\triangle_1 y_t = (1-\theta L)y_t = \theta_0 + \varepsilon_t$, with $\theta = 1$. In this case, the characteristic equation is $(1-\theta^S z) = 0$, so that when $\theta = 1$, y_t has a single nonseasonal unit root, corresponding to the simple random walk model. Also, when $\theta = -1$, y_t has a seasonal unit root. Thus, as mentioned above, both seasonal and non-seasonal unit root processes such as those given in our RW, DS, and SUR models, are nested within PAR models (see Hylleberg, Engle, Granger and Yoo (HEGY: 1990) for further details). Nonlinear variants of the above models are discussed in Franses and van Dijk (2005), Franses, de Bruin and van Dijk (2000) and Rodrigues (2001), and the papers cited therein.[4] Boswijk and Franses (1996: BF) outline tests for $\theta_{1,s}\theta_{2,s}\theta_{3,s}\theta_{4,s} = 1$ and for $\theta = 1$ and $\theta = -1$ (see Franses and Paap (2002) for a summary).

Of final note is that X_t variables are selected from amongst the set of all variables which are examined in the paper, and were only used in our prediction experiments, and not in our simulation experiments.

2.2 Comments on Seasonal Unit Root Models

In this sub-section, we summarize some of the most important recognized features of SUR models. First, it is important to recall that seasonal unit root models such as $\triangle_S y_t = \theta_0 + \varepsilon_t$, for example, define time series that are driven by a nonstationary process in each season. Now, these seasons might share a common drift, θ_0, but their evolution is clearly independent from one season to the next, given that the evolution of each season can be written as $y_t^s = \theta_0 + \varepsilon_{s,t}$, say, where $\varepsilon_{s,t} \sim iid$ $N(0, \sigma^2)$, say, for $s = 1, ..., S$. Thus, $\triangle_S y_t$ can be viewed as being comprised of S independent random walks. This feature is noted quite frequently in the literature when it is stated that the

[4]For a discussion of the types of nonlinear models that these authors have adapted to seasonality, see Granger and Teräsvirta (1993) and Teräsvirta (1994, 1998), for example.

model allows "winter to become summer", as discussed in Osborn and Ghysels (2001) and Franses and van Dijk (2005), and suggests that one should expect values of y_t in different seasons to drift infinitely far apart, given enough time. Clearly this is a feature of SUR models which is not in accord with expected behavior of economic series (see below for further discussion of this feature of SUR models).

All models should likely be viewed as approximations. Thus, the above feature may be acceptable, as long as a particular SUR model still yields adequate predictions and simulated observations for a given time series. However, from the perspective of simulation, a serious consequence of the above characteristic is that simulated data from such models cannot be used to provide reasonable simulated data for $\triangle_{S-j}y_t$, say, where $0 < j < S$. This feature is illustrated in Figure 1, where inflation, measured as 1st log difference CPI index values, are constructed using simulated observations from 2 models, each of which is calibrated using monthly U.S. data for the period 1991-2004. The first model specifies $\ln(y_t/y_{t-1}) = \theta_0 + \varepsilon_t$ (i.e. a standard random walk with drift in logs), while the second model specifies $\ln(y_t/y_{t-12}) = \theta_0 + \varepsilon_t$ (i.e. a standard SUR model in logs). Notice that the first log difference data generated via the SUR simulation model are increasing, with ever increasing amplitude, a feature which is clearly not in accord with the historical record, as will be discussed later.[5]

The mechanism underlying the behavior of the series in Figure 1 is simple. First, assume that data are logged prior to model estimation, so that we are actually interested in modeling $\ln y_t$, say, as is often the case when using macroeconomic variables such as those considered in this paper. Without loss of generality, consider the simplest form of the SUR model specified in Section 2.1. Namely, assume that $\triangle_S \ln y_t = \varepsilon_t$. Further, notice that $\triangle_S \ln y_t = \ln\left(\frac{y_t}{y_{t-S}}\right) = \Sigma_{i=0}^{S-1} \ln\left(\frac{y_{t-i}}{y_{t-i-1}}\right)$. Now, each of the twelve first log differences in this sum can be expressed as:

$$\ln\left(\frac{y_{t-i}}{y_{t-i-1}}\right) = \varepsilon_{t-i} + \varepsilon_{t-S-i} + \varepsilon_{t-2S-i} + ... - \varepsilon_{t-i-1} - \varepsilon_{t-S-i-1} - \varepsilon_{t-2S-i-1} - ... ,$$

for $i = 0, ..., S-1$. This follows by assuming that, for ease of exposition, that $\ln\left(\frac{y_t}{y_{t-S}}\right) = \varepsilon_t$, so that

$$\ln y_{t-i} = \ln y_{t-S-i} + \varepsilon_{t-i} = ... = \varepsilon_{t-i} + \varepsilon_{t-S-i} + \varepsilon_{t-2S-i} + ...,$$

[5]Note that the two models fit the data equally well, based on examination of in-sample correlation and residual serial autocorrelation.

where ε_t may be an *iid* random variable, for example. Thus, each first difference in the sum, $\Sigma_{i=0}^{S-1} \ln\left(\frac{y_{t-i}}{y_{t-i-1}}\right)$, evolves independently, as a sum of *iid* errors, as noted above. Now, note that for $\varepsilon_{t-i} >> 0$, there is a large pulse in $\ln\left(\frac{y_{t-i}}{y_{t-i-1}}\right)$, which persists in $\ln\left(\frac{y_{t+nS-i}}{y_{t+nS-i-1}}\right)$, as one simulates the process forward n years, where S is the number of observations in each year. In time, there will clearly occur a sequence of innovations where jumps $\varepsilon_{t-i}, \varepsilon_{t+S-i}, ...$reinforce each other, so that $\ln\left(\frac{y_{t-i}}{y_{t-i-1}}\right)$ has pulses of increasing amplitude at frequency $1/S$, and $\ln\left(\frac{y_{t+1-i}}{y_{t-i}}\right)$ has pulses of increasing amplitude at frequency $1/S$, but with the opposite sign. The result is a series that oscillates with increasing positive and negative amplitude, as demonstrated in Figure 1. Nevertheless, the twelve different first differences that comprise $\ln\left(\frac{y_t}{y_{t-S}}\right)$ are "cointegrated", in the sense that their sum is a stationary process (i.e. $\ln\left(\frac{y_t}{y_{t-S}}\right) = \Sigma_{i=0}^{S-1} \ln\left(\frac{y_{t-i}}{y_{t-i-1}}\right) = \varepsilon_t$). Thus, while the individual first log difference processes that comprise $\ln\left(\frac{y_t}{y_{t-S}}\right)$ can drift infinitely far apart (and should be expected to do so) over time, their sum remains stationary. It is in this sense that the SUR model can only be expected to reproduce reasonable dynamics for the precise variety of data with which it is calibrated (i.e. S^{th} log difference data in the above example). Apparently, when used as an approximation "tool" for unknown DGPs, the SUR model is thus limited. Nevertheless, it remains to see whether these theoretical limitations translate into empirical shortcomings; a topic to which we turn in Section 3.

2.3 A Distribution Comparison Test for Simulated Data

In this section we summarize the simulation based distributional accuracy test discussed in Corradi and Swanson (CS: 2004a) and Corradi and Swanson (2005a) for comparing dynamic stochastic general equilibrium models. The test can also be used to compare simulated distributions from econometric models, such as those used in this paper.

Assume that our objective is to compare the joint distribution of the historical data with the joint distribution of a simulated series. Following CS, and for the sake of simplicity (but without loss of generality), we limit our attention in the section to the evaluation of the joint empirical distribution of (actual and model-based) current and previous period output.

Consider m alternative econometric models, and set model 1 as the benchmark model. We require at least one of the competing models (e.g. model j for $j = 2, ..., m$) to be nonnested with respect to the benchmark, a requirement which is trivially satisfied in many contexts. For the sake of notational ease of expression, let $\Delta \ln z_t$, $t = 1, ..., T$ denote the natural logarithm of an actual historical series of interest, and let $\Delta \ln z_{j,n}$, $j = 1, ..., m$ and $n = 1, ..., N$, denote the same series, simulated under model j, where N denotes the length of the simulated sample.[6] In general, some parameters in the models may be kept fixed (at calibrated values, for example), while others may be estimated.

Along these lines, denote $\Delta \ln z_{j,n}(\widehat{\theta}_{j,T})$, $n = 1, ..., N$, $j = 1, ..., m$ to be a sample of length N drawn (simulated) from model j and evaluated at the parameters estimated, under model j, using the T available historical observations.[7] The reason why we use differences is that stationarity is assumed in our subsequent analysis. For ease of exposition, assume that interest focuses on current and lagged values of the variable of interest, let $y_t = (\ln \Delta z_t, \ln \Delta z_{t-1})$, $y_{j,n}(\widehat{\theta}_{j,T}) = (\Delta \ln z_{j,n}(\widehat{\theta}_{j,T}), \Delta \ln z_{j,n-1}(\widehat{\theta}_{j,T}))$. (In general, one can set y_t to be any stationary vector of economic variables.) Also, let $F_0(u; \theta_0)$ denote the distribution of y_t evaluated at u and $F_j(u; \theta_j^\dagger)$ denote the distribution of $y_{j,n}(\theta_j^\dagger)$, where $\theta_j^\dagger$ is the probability limit of $\widehat{\theta}_{j,T}$, taken as $T \to \infty$, and where $u \in U \subset \Re^2$, possibly unbounded. Accuracy is measured in terms of square error. The squared (approximation) error associated with model i, $i = 1, ..., m$, is measured in terms of the (weighted) average over U of $E\left(\left(F_i(u; \theta_i^\dagger) - F_0(u; \theta_0)\right)^2\right)$. Thus, the rule is to choose Model 1 over Model 2 if

$$\int_U E\left(\left(F_1(u; \theta_1^\dagger) - F_0(u; \theta_0)\right)^2\right) \phi(u) du < \int_U E\left(\left(F_2(u; \theta_2^\dagger) - F_0(u; \theta_0)\right)^2\right) \phi(u) du$$

where $\int_U \phi(u) du = 1$ and $\phi(u) \geq 0$ for all $u \in U \subset \Re^2$. For any evaluation point, this measure defines a norm and it implies a standard goodness of fit measure (see Corradi and Swanson (2005a) for further discussion of this loss measure). The hypotheses of interest are:

$$H_0 : \max_{j=2,\ldots,m} \int_U E\left(\left(F_0(u; \theta_0) - F_1(u; \theta_1^\dagger)\right)^2 - \left(F_0(u) - F_j(u; \theta_j^\dagger)\right)^2\right) \phi(u) du \leq 0$$

versus

$$H_A : \max_{j=2,\ldots,m} \int_U E\left(\left(F_0(u; \theta_0) - F_1(u; \theta_1^\dagger)\right)^2 - \left(F_0(u) - F_j(u; \theta_j^\dagger)\right)^2\right) \phi(u) du > 0.$$

Thus, under H_0, no model can provide a better approximation (in square error sense) to the distribution of Y_t than the approximation provided by model 1. If interest focuses on confidence intervals, so that the objective is to "approximate" $\Pr(\underline{u} \leq y_t \leq \overline{u})$, then the null and alternative hypotheses

[6]The variable of interest need not be logged.

[7]CS establish their results using the quasi maximum likelihood estimator for θ.

can be written in a form similar to that above, and appropriate statistics can be constructed, as discussed in Corradi and Swanson (2005a,b). In order to test H_0 versus H_A, the relevant test statistic is $\sqrt{T}Z_{T,N}$, where:

$$Z_{T,N} = \max_{j=2,\ldots,m} \int_U Z_{j,T,N}(u)\phi(u)du, \tag{1}$$

and

$$\begin{aligned} Z_{j,T,N}(u) &= \frac{1}{T}\sum_{t=1}^{T}\left(1\{y_t \le u\} - \frac{1}{N}\sum_{n=1}^{N} 1\{y_{1,n}(\widehat{\theta}_{1,T}) \le u\}\right)^2 \\ &\quad -\frac{1}{T}\sum_{t=1}^{T}\left(1\{y_t \le u\} - \frac{1}{N}\sum_{n=1}^{N} 1\{y_{j,n}(\widehat{\theta}_{j,T}) \le u\}\right)^2, \end{aligned}$$

where $\widehat{\theta}_{j,T}$ is an estimator of $\theta_j^{\dagger}$.

From equation (1), it is immediate to see that the computational burden increases with the dimensionality of U, that is with the number of variables and/or lagged values we are considering. In fact, we need to approximate the integral by taking an average over a fine grid of U.[8] Unfortunately, Monte Carlo integration techniques, such as the importance sampling or one of its accelerated versions cannot be used. This is because $Z_{j,T,N}(u)\phi(u)$ is not a joint density and can be either negative or positive. A possibility would be to compute instead the statistic $\widetilde{Z}_{T,N} = \max_{j=2,\ldots,m} \frac{1}{T}\sum_{i=k}^{T} Z_{j,T,N}(y_i)$, where k denotes the highest lag order. If y_t were an *iid* vector-valued process, then $\sqrt{T}Z_{T,N}$ and $\sqrt{T}\widetilde{Z}_{T,N}$ are asymptotically equivalent, as shown in Andrews (1997). However, in our case, y_t is a dependent process and the argument used in Andrews' proof, based on his Lemma A6, does not apply. Nevertheless, as T gets large $(y_k, y_{k+1}, ...y_T)$ will become a dense subset in U, and so we conjecture that $\sqrt{T}Z_{T,N}$ and $\sqrt{T}\widetilde{Z}_{T,N}$ are asymptotically equivalent even in the dependent case, though a formal proof of this is not a trivial task.

As outlined in CS, under some weak assumptions, $Z_{T,N}$ converges in distribution to a functional of a Gaussian process with a covariance kernel that reflects contributions by parameter estimation error, simulation error, when present, and the time series structure of the data. Asymptotically

[8] For example, if U is a two-dimensional subset of $\Re^2$, and ϕ is uniform on U, then

$$Z_{T,N} = \frac{1}{N_1 N_2} \max_{j=2,\ldots,m} \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} Z_{j,T,N}(u_{i,j}).$$

valid critical values for the test statistic can be constructed as follows.[9]

Begin by resampling b blocks of length l, $bl = T-1$, from the actual sample. Let $y_t^* = (\ln \Delta z_t^*, \ln \Delta z_{t-1}^*)$ be the resampled series, such that $y_2^*, ..., y_{l+1}^*, y_{l+2}^*, ..., y_{T-l+2}^*, ..., y_T^*$ is equal to $y_{I_1+1}, ..., y_{I_1+l}, y_{I_2+1}, ..., y_{I_b+1}, ..., y_{I_b+l}$, where I_i, $i = 1, ..., b$ are independent, discrete uniform on $1, ..., T-l+1$, that is $I_i = i$, $i = 1, ..., T-l$ with probability $1/(T-l)$. We use the resampled series y_t^* to compute the bootstrap estimator $\widehat{\theta}_{j,T}^*$ for $j = 1, ..., m$.

We now use $\widehat{\theta}_{j,T}^*$ to simulate samples under model j, $j = 1, ..., m$; let $y_{j,n}(\widehat{\theta}_{j,T}^*)$, $n = 2, ..., N$ be the series simulated under model j. At this point, we need to distinguish between the case of $\delta = 0$ (vanishing simulation error) and $\delta > 0$ (nonvanishing simulation error). In the former case, we do not need to resample the simulated series, as there is no need of mimicking the contribution of simulation error to the covariance kernel. On the other hand, in the latter case, we do need to resample the simulated series. More precisely, we draw $\widetilde{b}$ blocks of length $\widetilde{l}$, with $\widetilde{b}\widetilde{l} = N-1$, let $y_{j,n}^*(\widehat{\theta}_{j,T}^*)$, $j = 1, ..., m$, $n = 2, ..., N$ denote the resample series under model j. Notice that $y_{j,2}^*(\widehat{\theta}_{j,T}^*), ..., y_{j,l+1}^*(\widehat{\theta}_{j,T}^*), ..., y_{j,N}^*(\widehat{\theta}_{j,T}^*)$ is equal to $y_{j,\widetilde{I}_1}(\widehat{\theta}_{j,T}^*), ..., y_{j,\widetilde{I}_1+l}(\widehat{\theta}_{j,T}^*), ..., y_{j,\widetilde{I}_{\widetilde{b}}+l}(\widehat{\theta}_{j,T}^*)$, where $\widetilde{I}_i$, $i = 1, ..., \widetilde{b}$ are independent discrete uniform on $1, ..., N-\widetilde{l}$. Notice that, for each of the m models, and for each bootstrap replication, we draw $\widetilde{b}$ discrete uniform $\widetilde{I}_i$ on $1, ..., N-\widetilde{l}$, draws are independent across models, we have just suppressed the dependence of $\widetilde{I}_i$ on j, for notational simplicity.

We consider two different bootstrap analogs of $Z_{T,N}$, the first of which is valid when $T/N \to \delta > 0$ and the second of which is valid when $T/N \to \delta = 0$. Notice that in the second version, simulation error vanishes so that $y_{j,n}^*(\widehat{\theta}_{j,T}^*)$ in the first statistic is replaced with $y_{j,n}(\widehat{\theta}_{j,T}^*)$, $j = 1, ..., m$. Define:

$$Z_{T,N}^{**} = \max_{j=2,\ldots,m} \int_U Z_{j,T,N}^{**}(u)\phi(u)du,$$

[9] Recall that all candidate models are potentially misspecified under both hypotheses in the test outlined above. Thus, the parametric bootstrap is not generally applicable in our context. Put another way, if observations are resampled from one of the candidate models, then we cannot ensure that the resampled statistic has the same limiting distribution as that of $Z_{T,N}$. Hence, we use the nonparametric block bootstrap.

where

$$
\begin{aligned}
Z_{j,T,N}^{**}(u) &= \frac{1}{T}\sum_{t=1}^{T}\left(\left(1\{y_t^* \leq u\} - \frac{1}{N}\sum_{n=1}^{N}1\{y_{1,n}^*(\widehat{\theta}_{1,T}^*) \leq u\}\right)^2\right. \\
&\quad \left. - \left(1\{y_t \leq u\} - \frac{1}{N}\sum_{n=1}^{N}1\{y_{1,n}(\widehat{\theta}_{1,T}) \leq u\}\right)^2\right) \\
&\quad -\frac{1}{T}\sum_{t=1}^{T}\left(\left(1\{y_t^* \leq u\} - \frac{1}{N}\sum_{n=1}^{N}1\{y_{j,n}^*(\widehat{\theta}_{j,T}^*) \leq u\}\right)^2\right. \\
&\quad \left. - \left(1\{y_t \leq u\} - \frac{1}{N}\sum_{n=1}^{N}1\{y_{j,n}(\widehat{\theta}_{j,T}) \leq u\}\right)^2\right)
\end{aligned}
$$

and

$$
Z_{T,N}^{*} = \max_{j=2,\dots,m}\int_U Z_{j,T,N}^{*}(u)\phi(u)du,
$$

where

$$
\begin{aligned}
Z_{j,T,N}^{*}(u) &= \frac{1}{T}\sum_{t=1}^{T}\left(\left(1\{y_t^* \leq u\} - \frac{1}{N}\sum_{n=1}^{N}1\{y_{1,n}(\widehat{\theta}_{1,T}^*) \leq u\}\right)^2\right. \\
&\quad \left. - \left(1\{y_t \leq u\} - \frac{1}{N}\sum_{n=1}^{N}1\{y_{1,n}(\widehat{\theta}_{1,T}) \leq u\}\right)^2\right) \\
&\quad -\frac{1}{T}\sum_{t=1}^{T}\left(\left(1\{y_t^* \leq u\} - \frac{1}{N}\sum_{n=1}^{N}1\{y_{j,n}(\widehat{\theta}_{j,T}^*) \leq u\}\right)^2\right. \\
&\quad \left. - \left(1\{y_t \leq u\} - \frac{1}{N}\sum_{n=1}^{N}1\{y_{j,n}(\widehat{\theta}_{j,T}) \leq u\}\right)^2\right).
\end{aligned}
$$

CS prove the first order validity of critical values constructed using the above bootstrap statistics. They thus suggest proceeding in the following manner. For any bootstrap replication, compute the bootstrap statistic, $\sqrt{T}Z_{T,N}^{**}$ $\left(\sqrt{T}Z_{T,N}^{*}\right)$. Perform B bootstrap replications (B large) and compute the quantiles of the empirical distribution of the B bootstrap statistics. Reject H_0 if $\sqrt{T}Z_{T,N}$ is greater than the $(1-\alpha)th$-quantile. Otherwise, do not reject. Now, for all samples except a set with probability measure approaching zero, $\sqrt{T}Z_{T,N}$ has the same limiting distribution as the corresponding bootstrapped statistic, when $\int_U \left((F_0(u)-F_1(u))^2-(F_0(u)-F_j(u))^2\right)\phi(u)du = 0$ for all $j=2,...,m$. In this case, the above approach ensures that the test has asymptotic size equal to α. On the other hand, when one (or more) competing models is (are) strictly dominated by the benchmark, the approach ensures that the test has an asymptotic size between 0 and α. Finally, under the alternative, $Z_{T,N}$ diverges to (plus) infinity, while the corresponding bootstrap statistic has a well defined limiting distribution. This ensures unit asymptotic power. Note that the suggested bootstrap procedure mimics the limiting distribution of the statistics in the least favorable case for the null, thus leading to conservative inference (see Corradi and Swanson (2004a) for further discussion and alternatives approaches for inference using $Z_{T,N}$).

2.4 Predictive Accuracy Tests

There are many predictive accuracy and predictive model selection tests available in the current literature, such as those which are: designed to be valid when parameter uncertainty does not vanish asymptotically (see e.g. West (1996)); designed to compare multiple horizon prediction models (see e.g. Clark and McCracken (2001, 2004); designed to be generically consistent against (non)linear alternatives (see e.g. Corradi and Swanson (2004b)); designed for comparing many models (see e.g. Sullivan, Timmerman and White (1999,2001) and White (2000)); designed for comparing predictive density models (see e.g. Clements and Smith (2000,2002) and Corradi and Swanson (2005b)); and designed to relax loss function differentiability assumptions (see e.g. Diebold and Mariano (1995) and McCracken (2004)), for example. Indeed, there are so many predictive accuracy test that it would require too much time to list them all. Many of them, however, are discussed in the forthcoming handbook of forecasting, two papers from which are listed in the references as Corradi and Swanson (2004c) and West (2004).[10]

Some of the tests that have received the most attention in recent theoretical literature include encompassing tests, such as that due to Clark and McCracken (CM: 2004) and Harvey, Leybourne and Newbold (1997) and the predictive accuracy tests, such as that due to Diebold and Mariano (DM: 1995). In this paper, we use the DM predictive accuracy test, which is defined as follows:

$$DM = \sqrt{P}\frac{\frac{1}{P}\sum_{t=R}^{T}\widehat{d}_{t+h}}{\frac{1}{P-h+1}\sum_{j=-\bar{j}}^{\bar{j}}\sum_{t=R+j}^{T-h}K\left(\frac{j}{M}\right)\left(\widehat{d}_{t+h}-\overline{d}\right)\left(\widehat{d}_{t+h-j}-\overline{d}\right)},$$

where $\widehat{d}_{t+h} = \widehat{u}_{1,t+h}^2 - \widehat{u}_{2,t+h}^2$ under mean square error loss, $\overline{d} = \frac{1}{P-h+1}\sum_{t=R}^{T-\tau}\widehat{d}_{t+h}$, and P is the ex ante forecast period. Of note is that loss functions other than mean square error loss can be used, although in this paper we focus exclusively on mean square forecast error loss. The limiting

[10]For further references and discussion see Clements and Hendry (1999a,b), and the references cited therein.

distributions of the DM statistic is given in Theorems 3.1 and 3.2 in Clark and McCracken (2004), and for $h > 1$ contains nuisance parameters so that critical values cannot be directly tabulated, and hence Clark and McCracken (2004) use the Kilian (1999) parametric bootstrap to obtain critical values (see e.g. McCracken and Saap (2004) for discussion). Of interest in our case, though, is the fact that for $h = 1$, when models are nonnested, and assuming that parameter uncertainty vanishes asymptotically, the standard normal distribution applies in the case of the DM test, even when the heteroskedasticity and autocorrelation consistent standard error given in the numerator of DM is replaced with $\frac{1}{P-h+1}\sum_{t=R+j}^{T-h}\left(\widehat{d}_{t+h}-\overline{d}\right)^2$. Furthermore, nonstandard critical values which obtain in other cases are generally larger, so that rejection of the null hypothesis using percentiles of the normal distribution often implies rejection using nonstandard critical values. Finally, nonstandard critical values are usually quite close, in absolute magnitude, to standard normal values. For these reasons, we use standard normal rejection regions when we report significance of DM test statistics. These should, of course, only be taken as a rough guide.[11]

3 Empirical Results

The variables examined in this section include: Industrial Production (IP: total index, 1997=100, NSA), various interest rates including the 1-Year (R1) and 10-year (R2) Treasury Constant Maturity Rates and the Effective Federal Funds Rate (R3); various measures of the money stock (billions of dollars, NSA), including M1, M2, and M3; and various measures of the Consumer Price Index for All Urban Consumers (index, 1982-84=100, NSA), including All Items (CPI1), All Items Less Energy (CPI2), All Items Less Food (CPI3), All Items Less Food & Energy (CPI4), Energy (CPI5), and Food (CPI6). All data are monthly U.S. observations for the period 1959:1-2004:12. In the proceeding sub-sections, we summarize results based upon: (i) a basic data analysis, (ii) simulation experiments, and (iii) prediction experiments.

[11] Another reason to use the normal distribution to obtain critical values is that it is not clear whether the Kilian parametric bootstrap is asymptotically valid under estimation schemes such as recursive and rooling estimation (see Corradi and Swanson (2004d) for further discussion).

3.1 Basic Data Analysis

Each of the 4 types of models is estimated using least squares, and lags are selected using the Schwarz Information Criterion (SIC). For each estimated model, summary measures are reported, including adjusted R^2, Durbin-Watson (DW) first order autocorrelation, and Lagrange multiplier (LM) first order autocorrelation test statistics. Recall that two versions of the random walk were fitted, one with $\theta_0 = 0$ and one with $\theta_0 \neq 0$. For our basic data analysis and our simulation experiments, we report results for the case where $\theta_0 \neq 0$, called "RW". However, for our prediction experiments, we report both cases, called "RW" ($\theta_0 = 0$) and "RW-D" ($\theta_0 \neq 0$).

In addition to estimation summary statistics, a number of diagnostic statistics are discussed. These include mean, variance, skewness and kurtosis statistics, as well as Jarque-Bera, augmented Dickey-Fuller (ADF) test (with lags selected via use of the SIC), BF periodic autoregression tests, and HEGY seasonal unit root tests (see Beaulieu and Miron (1993) and Osborn and Rodrigues (2002) for a detailed discussion of the HEGY test). All of these results are gathered in Tables 1-4.

A number of conclusions can be drawn upon examination of these tables. First, note that in Table 1, 12 of 13 series fail to reject the null of a unit root, based on application of the ADF test. However, no series exhibit evidence of seasonal unit roots (see HEGY test results in Table 1 and Table 2), while only about 1/2 of the series exhibit evidence of a periodic autoregressive structure (see BF test results in Table 1). Interestingly, there is more evidence of PAR when our shorter subsample from 1991:1-2004:12 is used. In particular, note in Table 4 that 8 of 13 series exhibit evidence of PAR structure. Furthermore, based on this shorter sub-sample, one series (M1) exhibits evidence of a seasonal unit root, and various other series (including IP, M2, M3, R3, CPI2, CPI4, and CPI6) are near to accepting the seasonal unit root null hypothesis at conventional significance levels. Thus, we have some evidence that SUR and PAR models may be useful for approximating the DGPs of a variety of our series. Second, and in accord with our first conclusion, note from Tables 3 and 4 that there is little evidence of error serial correlation when the posited models are DS or SUR models (see DW and LM test results), regardless of whether the longer (1959:1-2004:12) or shorter subsamples are used for estimation. In addition, less than 1/2 of the series have remaining serial correlation after being filtered using the PAR model, based upon the shorter subsample (see 10th and 11th columns of entries in Tables 3 and 4). Finally, it is clear from examination of adjusted R^2 values that there is little to choose between the different models based on this measure.

3.2 Simulation Experiments

In order to evaluate the performance of the different models from the perspective of simulation, each estimated model was used to simulate N observations. The starting values for the simulations are fixed to be the last observation of the historical sample used in estimation. Note that the models used for simulation are the models outlined above, *including the error term*, where the error is assumed to be *iid* normal, and where the variance of the error term is estimated using the residuals of the fitted model. Thus, simulation models take forms such as $\triangle_1 y_t = 1.66 + \varepsilon_t$, for example, where $\varepsilon_t \sim iidN(0, \sigma_\varepsilon^2)$, and σ_ε^2 is calibrated from the data. Results were also collected for the case where errors were drawn randomly from the empirical distribution of the residuals. Findings in this case were qualitatively similar to those reported below, and hence have been omitted for the sake of brevity (complete results are available upon request). We consider simulations of length $N = \{60, 300, 1200\}$. Given that the historical data are monthly, and given that the initialization point of the simulations is our last historical observation, the simulated path lengths correspond to 5, 25, and 100 year future trajectories. All results are based on the analysis of 100 simulation paths, and a number of summary measures are gathered in Tables 5a-8c. Corresponding results based upon application of the distribution comparison test (i.e. $Z_{T,N}$) are gathered in Tables 9a-9d. Finally, kernel density plots comparing simulated with historical distributions are given in Figures 2a-5b.

Turning first to Tables 5a-8c, note that these tables are organized as follows. Each table contains 4 panels, one for each of the 4 model varieties. Entries in the tables are mean, standard deviation, skewness, kurtosis, minimum and maximum values based upon the two historical sub-samples examined in this paper (1959:1-2004:12 and 1991:1-2004:12), as well as based upon data simulated using the 4 different models.[12] Finally, note that tabulated entries based upon simulated data are averages across 100 simulation paths, each with length given in the first row of entries in the table. Our findings based upon examination of these tables can be summarized as follows. First, note that Tables 5a, 6a, 7a, and 8a report results for IP, M1, R1, and CPI1 in (log) levels. Thus,

[12] Of note is that all simulation models are calibrated using data estimated with the larger subsample of historical data. It is also of interest to estimate the simulation models with the shorter sample period, and results based upon such models are summarized in the next subsection of the paper, where predictive accuracy is discussed.

the various summary measures reported in the tables should be expected to increase in absolute magnitude with the simulated sample size. This is indeed the case. However, conditional on this fact, it is noteworthy that PAR models perform as well or better than all competitors for almost every summary distributional measure, and across all 4 variables. Nevertheless, it is also the case that summary distributional measures based upon simulation of 1st and 12th (log) differences using the alternative models indicate that the PAR model sometimes yields simulations characterized by outliers. This can be seen by inspecting the remaining tables from amongst Tables 5a-8c. One of the reasons that this occurs is because we are estimating our models with the entire historical dataset, and the PAR models are sensitive to outliers and structural breaks. This can be seen to be the case, as data simulated using models parameterized with the shorter subsample of historical observations yield much more reasonable distributional summary measures in these cases (results are available upon request). Additionally, note that in the next subsection we construct forecasts using the different models, all of which are estimated using data in the shorter subsample; and in these prediction experiments, the PAR models often outperform the other models, regardless of data transformation.

Second, notice that the volatility of the simulation data under the SUR models increases quite rapidly with the simulation sample size when 1st differences are simulated. For example, regardless of which of the four variables are simulated, first log difference volatility increases around 400% when comparing results from 60 simulated values (3rd column of entries) with 1200 simulated values (9th column of entries) in Tables 5b, 6b, 7b, and 8b. As should be expected, this feature of SUR models pertains to any transformation other that 12th (log) differences. Thus, if the only target variable of interest is a 12th (log) difference variable, then the SUR model is useful. This is evidenced by the fact that the SUR models compare favorably when simulating 12th (log) differences (see Tables 5c, 6c, 7c, and 8c). This is clearly a very serious shortcoming of the SUR model, when compared with other models. Namely, the other models, such as the RW and DS models yield reasonable simulated observations, regardless of data transformation. This shortcoming of the SUR models is clearly in part due to the fact that the models can essentially be viewed as a sum of 12 different random walks in first differences, all of which can clearly drift infinitely far apart over time (see above discussion).

Given the above discussion, an important issue concerns how long it takes before the SUR model begins to yield unacceptable simulation observations for anything other than twelfth (log) differences. The evidence reported here suggests that problems crop up quite quickly (for further evidence, see the next subsection, which reports prediction findings).

Third, notice that inspection of Tables 5a-8c suggests that the RW, DS, and PAR models all perform well in a variety of cases (i.e. for various variables and distributional summary measures). Thus, it is difficult to select amongst the alternative models. However, the $Z_{T,N}$ test offers a convenient method for selecting amongst alternative models.[13]

As mentioned above, Tables 9a-d contain $Z_{T,N}$ test results, in which the benchmark (set equal to RW) is compared against the other three models. The tables are organized as follows. The first column gives S, the length of the simulation sample used, and l, the block length used in the construction of test critical values. The second column of entries reports the numerical values of the test statistic ($Z = Z_{T,S}$) discussed above, while the next four columns report 5% and 10% bootstrap critical values based on a bootstrap statistics constructed allowing for parameter estimation error ($Z^{**} = Z^{**}_{T,S}$) and assuming that parameters estimation error vanishes asymptotically ($Z^{*} = Z^{*}_{T,S}$). The last four columns report the Corradi and Swanson (2004a) distributional loss measure associated with model i, $i = 1, 2, 3$, (i.e. $\int_U \frac{1}{T}\sum_{t=1}^{T}\left(1\{Y_t \leq u\} - \frac{1}{N}\sum_{n=1}^{N} 1\{Y_{i,n}(\widehat{\theta}_{1,T}) \leq u\}\right)^2 \phi(u)du$) - see Section 2 for further details. As noted elsewhere, T denotes the historical sample size. For the (unreported) case where $N = T$, we set the block length used in the bootstrap to be $l = 2$, 4,6, and 12. For all other cases, where $N = aT$, say, we set l equal to 'a' times the corresponding value of l when $N = T$. All statistics are based on grids of 20x20 values for u, distributed uniformly across the historical data ranges of the variables being compared. Bootstrap empirical distributions are constructed using 100 bootstrap replications. Finally, note that, in contrast with the results reported in Tables 5a-8c, models are estimated using the shorter subsample of data, thus addressing the issue of outliers and structural breaks that arises upon inspection of the PAR model results discussed above.

Conclusions based upon examination of Tables 9a-d can be drawn, as follows. First, consider the distributional loss measures reported in the last four columns of entries of the tables. It is interesting to note that the only time the SUR model yields the lowest loss measure is for CPI1, when comparing 12th log differences. In all other cases across the four tables, the SUR model is outperformed by a variety of the other models. Thus, we again have evidence of the limited

[13]Simulation results are only reported for a subset of the 13 variables, as results based upon the other variables are qualitatively similar. Complete results are available upon request from the authors.

usefulness of the SUR model. Also as might be expected, the PAR model does best when comparing levels data. Furthermore, there appears little to choose between the RW, DS and PAR models when simulating 1st or 12th (log) differences, the exception being CPI1, for which RW and DS appear to be superior.

Second, inspection of the $Z_{T,N}$ test statistics reported in the 2nd column of the table with the bootstrapped critical values reported in the 3rd-6th columns of entries confirms that the benchmark model is never rejected in favor of any alternative model. Thus, while there are clearly numerous cases for which the RW yields higher distributional loss than some of the competitors, the differences are not only small in absolute magnitude, but are also insignificant. Of course, it is also the case that when the SUR model is used as the benchmark, there are numerous test rejections (results are not reported for the sake of brevity). In summary, we have evidence that all of the models are potentially useful, with the exception of the SUR model, which doesn't even appear to dominate when the target variable of interest is 12th (log) differences.

More evidence in favor of the above conclusions is summarized in Figures 2a-5b, where simulated distributions for first (log) difference data are plotted against the historical record. Notice that the simulated distributions associated with the SUR model are much further away from the historical distribution for all variables/transformations. The exception to this is PAR, which yields inferior distributions when the longer subsample is used, but which yields markedly improved distributions when the shorter subsample is used (as expected, given our above discussion). Notice also that simulated distributions based on the SUR model become progressively worse as one increases the length of the simulation sample, as can be seen by comparing Panels 1 with Panel 2 or Panel 2 with Panel 3 in any of Figures 2a-5b. As expected, this is clearly not the case, however, for RW, DS, and PAR models, again underscoring our earlier findings concerning the nonrobustness of SUR models.

In the next subsection, we discuss the predictive ability of the different models.

3.3 Prediction Experiments

In our prediction experiments h-step ahead forecasts are constructed using recursively estimated versions of each of the 4 models, beginning with a window of 60 observations, and ending with a window of $T - h$ observations, where h is the forecast horizon. After each successive forecasting model is estimated, an $h-$step ahead prediction is constructed, and an associated ex-ante prediction error is formed by comparing the predicted value with the historical value. Overall, this results in sequences of $T-h-60$ ex-ante predictions and prediction errors, where h is set equal to 1, 3, 12, and 60. Thus, the longest predictions are 5- years ahead, and the shortest are 1-month ahead. The samples used in this exercise are 1959:1-2004:12 and 1991:1-2004:12. However, results for the longer sample are poor relative to those for the shorter period, in accord with our findings based on simulation experiments (see e.g. Figures 1-4)). Thus, results are only reported for the shorter sample period.

In the construction of the prediction models, X_t variables are selected from amongst the set of all variables which are examined in the paper. In particular, in the prediction experiments, there are two cases considered, one where all four models are specified without any exogenous variables, and one where the four models are specified with the X_t variables. In the latter case, the exogenous variables are themselves modelled as univariate processes, with the same generic specification as the y_t variable. Of note, however, is that the qualitative findings reported in this section are not dependent upon whether or not X_t variables are included in the prediction models, although prediction is generally worse for all models when exogenous variables are included. For these reasons, we only report results based upon prediction using univariate models.

Results from this experiment are gathered in Tables 10a-10c, where mean square forecast errors (MSFEs) are reported, and in Tables 11a-11c, where DM test statistics based on MSFE loss are reported. A number of conclusions can be drawn from examination of these tables. First, note that there are 156 variable/forecast horizon permutations summarized in Tables 10a-10c. Across all of these, the SUR model yields the lowest MSFE only 5 times. The PAR model yields the lowest MSFE 23 times, the DS model "wins" 35 times, and the RW and RW-D models win the remaining 88 times (approximately 45 times each). Thus, from the perspective of prediction, the SUR model is inferior at all horizons and for all transformations, including 12th (log) differences. The other models all perform similarly, and are all clearly useful, depending upon the particular permutation being considered.

Second, the SUR model never "wins" when used to predict either 1st or 12th (log) differences. While it is not surprising that the SUR model "loses" for 1st (log) differences, it is interesting that it also loses whenever 12th (log) differences are predicted. Given that the DS and PAR models win numerous times, we have clear evidence that any seasonality present in the data is not of the SUR variety. The SUR model yields poor simulation results for anything other than 12th (log) differences, and yields poor predictions for every transformation, *including* 12th (log) differences.

Third, the prevalence of RW and RW-D "wins" increases as the forecast horizon increases, from 1-month ahead to 5-years ahead (see Panels 1-4 of the tables). This suggests that greater model complexity may not be as useful for longer as for shorter horizons; a finding which is not too surprising, given the well known difficulty in predicting h-steps ahead, for h large. Thus, when simulations and/or prediction models are used to form a picture of the distant future, parsimonious models may be preferred to more complicated models.[14]

Fourth, notice in Tables 11a-c (where the benchmark is set equal to RW), that many reported DM test statistics are negative. This is in accord with the fact that the RW model often yields lower MSFEs; positive values correspond to the case where the "competitor" model yields lower MSFE. Recall also that starred entries denote rejection of the null hypothesis based upon the percentiles of the standard normal distribution. As many of our models are nested, the critical values used are thus only valid indicators when the test rejects, in which case we know it would reject under valid critical values, with nominal level less than the 10% used in the table (see discussion above). Given this feature of our setup, it is interesting to note that there are far fewer test rejections based upon (log) level and 12th (log) difference forecasts than for 1st (log) differences, suggesting that there is often little to choose between the models, based on standard significance level tests. However, the RW model clearly does well for many variables when 1st (log) differences are being predicted, particularly for $h > 1$.

In summary, our prediction experiments lend credence to our simulation results. Namely, it appears that the SUR models are inferior to other simpler (deterministic) seasonal models, more complicated (but clearly more realistic) PAR models, and to strawman random walk models. This results appears to hold regardless of data transformation, forecast horizon, or variable.

4 Concluding Remarks

Seasonal unit root (SUR) models are examined via a series of simulation and prediction experiments, and using a new simulation based testing methodology for comparing alternative models. We find that the SUR model is extremely limited in its ability to mimic distributional behavior of standard

[14]This conclusion is only preliminary, as much work still needs to be done using nonlinear models, for example, before definitive evidence on the connection between forecast horizon and model complexity can be given.

transformations of economic variables, when used for simulation. In addition, the SUR model yields the lowest mean square forecast error only 5 of 156 times when used for prediction with a variety of economic variable/forecast horizon combinations. Furthermore, the SUR model *never* "wins" when used to predict either 1st or 12th (log) differences. We conclude that some varieties of SUR models may fall within the "empty box" category discussed in Granger (1999). Of final note is that evidence against the usefulness of SUR models for prediction and simulation could be made even stronger were similar results shown to hold using nonlinear versions of the models considered in this paper. Such an analysis is left to future research.

5 References

Andrews, D.W.K., (1997), A Conditional Kolmogorov Test, *Econometrica*, 65, 1097-1128.

Beaulieu, J.J. and J. Miron, (1993), Seasonal Unit Roots and Deterministic Seasonals in Aggregate U.S. Data, *Journal of Econometrics*, 55, 305-328.

Boswijk, H.P. and P.H. Franses, (1996), Unit Roots in Periodic Autoregressions, *Journal of Time Series Analysis*, 17, 221-245.

Chao, J.C., V. Corradi, and N.R. Swanson (2001), An Out of Sample Test for Granger Causality", *Macroeconomic Dynamics,* v.5, 598-620.

Clark, T.E., and M.W., McCracken, (2001), Tests of Equal Forecast Accuracy and Encompassing for Nested Models, *Journal of Econometrics* 105, 85-110.

Clark, T.E., and M.W., McCracken (2004), Evaluating Long-Horizon Forecasts, Working Paper, University of Missouri-Columbia.

Clements, M.P. and D.F. Hendry, (1999a), *Forecasting Economic Time Series: The Zeuthen Lectures on Economic Forecasting,* MIT Press: Cambridge.

Clements, M.P. and D.F. Hendry, (1999b), On Winning Forecasting Competitions in Economics, *Spanish Economic Review*, 1, 123-160.

Clements, M.P. and J. Smith, (2000), Evaluating the Forecast Densities of Linear and Nonlinear Models: Applications to Output Growth and Unemployment, *Journal of Forecasting,* 19, 255-276.

Clements, M.P. and J. Smith, (2002), Evaluating Multivariate Forecast Densities: A Comparison of Two Approaches, *International Journal of Forecasting,* 18, 397-407.

Corradi, V., (1999), Deciding between $I(0)$ and $I(1)$ via FLIL-Based Bounds, *Econometric Theory,* 15, 643-663.

Corradi, V. and N.R. Swanson, (2002), A Consistent Test for Out of Sample Nonlinear Predictive Ability, *Journal of Econometrics,* 110, 353-381.

Corradi, Valentina and Norman R. Swanson, (2004a), Evaluation of Dynamic Stochastic General Equilibrium Models Based on Distributional Comparison of Simulated and Historical Data, *Journal of Econometrics*, forthcoming.

Corradi, Valentina and Norman R. Swanson, (2004b), Some Recent Developments in Predictive Accuracy Testing With Nested Models and (Generic) Nonlinear Alternatives, *International Journal of Forecasting*, 20, 185-199.

Corradi, V. and N.R. Swanson, (2004c), Predictive Density Evaluation, in: *Handbook of Economic Forecasting*, eds. Clive W.J. Granger, Graham Elliot and Allan Timmerman, Elsevier, Amsterdam, forthcoming.

Corradi, Valentina and Norman R. Swanson, (2004d), Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes, discussion paper, Rutgers University.

Corradi, V. and N.R. Swanson, (2005a), A Test for Comparing Multiple Misspecified Conditional Interval Models, *Econometric Theory*, forthcoming.

Corradi, V. and N.R. Swanson, (2005b), Predictive Density and Conditional Confidence Intervals Accuracy Tests, *Journal of Econometrics,* forthcoming.

Diebold, F.X., and R.S. Mariano, (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253-263.

Diebold, F.X., L.E. Ohanian, and J. Berkowitz, (1998), Dynamic Equilibrium Economies: A Framework for Comparing Models and Data, *Review of Economic Studies,* 65, 433-451.

Franses, P.H., (1996), *Periodicity and Stochastic Trends in Economic Time Series*, Oxford, Oxford University Press.

Franses, P.H., (1996), Recent Advances in Modelling Seasonality, *Journal of Economic Surveys*, 10, 299-345.

Franses, P.H., (2001), Some Comments on Seasonal Adjustment, *Revista de Economia del Rosaro*, 4, 9-16.

Franses, P. H., P. De Bruin, and D. van Dijk (2002), Seasonal Smooth Transition Autoregression, Econometric Institute Report 2000-06/A, Erasmus University.

Franses, P.H., H. Hoek, and R. Paap, (1997), Bayesian Analysis of Seasonal Unit Roots and Seasonal Mean Shifts, *Journal of Econometrics,* 78, 359-380.

Franses, P.H. and R. Paap, (2002), Forecasting with Periodic Autoregressive Time Series Models, in: *A Companion to Economic Forecasting,* eds. M.P. Clements and D.F. Hendry, Oxford, Blackwell Publishers, pp. 432-452.

Franses, P.H. and D. van Dijk, (2005), The Forecasting Performance of Various Models for Seasonality and Nonlinearity for Quarterly Industrial Production, *International Journal of Forecasting, 21, 87-102.*

Franses, P.H. and T. Vogelsang, (1998), On Seasonal Cycles, Unit Roots, and Mean Shifts, *Review of Economics and Statistics*, 80, 231-240.

Ghysels E. and D.R. Osborn, (2001), *The Econometric Analysis of Seasonal Time Series*, Cambridge, MA, Cambridge University Press.

Granger, C.W.J., (1999), Aspects of Research Strategies for Time Series Analysis, Presentation to the conference on New Developments in Time Series Economics, Yale University.

Granger C.W.J. and T. Teräsvirta, (1993), *Modelling Nonlinear Economic Relationships*, Oxford, Oxford University Press.

Harvey, D.I., S.J. Leybourne and P. Newbold, (1997), Tests for Forecast Encompassing, *Journal of Business and Economic Statistics*, 16, 254-259.

Hylleberg, S., (1992), *Modelling Seasonality*, Oxford, Oxford University Press.

Hylleberg, S., (1994), Modelling Seasonal Variation, in: *Nonstationary Time Series Analysis and Cointegration*, eds. C.P. Hargreaves, Oxford, Oxford University Press.

Hylleberg, S., R.F. Engle, C.W.J. Granger and B.S. Yoo, (1990), Seasonal Integration and Cointegration, *Journal of Econometrics,* 44, 215-238.

Kilian, L., (1999), Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?, *Journal of Applied Econometrics*, 14, 491-510.

McCracken. M.W., and S. Sapp, (2004), Evaluating the Predictive Ability of Exchange Rates Using Long Horizon Regressions: Mind your p's and q's. *Journal of Money, Credit and Banking*, forthcoming.

Miron, J.A., (1996), *The Economics of Seasonal Cycles*, Cambridge, MA, MIT Press.

Miron, J.A. and J.J. Beaulieu, (1996), What Have Macroeconomists Learned About Business Cycles from the Study of Seasonal Cycles, *Review of Economics and Statistics*, 78, 54-66.

Osborn, D.R., (2002), Unit Roots versus Deterministic Representations of Seasonality for Forecasting, in: *A Companion to Economic Forecasting,* eds. M.P. Clements and D.F. Hendry, Oxford, Blackwell Publishers, pp. 409-431.

Osborn, D.R., S. Heravi, and C.R. Birchenhall, (1999), Seasonal Unit Roots and Forecasts of Two-Digit European Industrial Production, *International Journal of Forecasting*, 15, 27-47.

Osborn, D.R. and P.M.M. Rodrigues, (2002), Asymptotic Distributions of Seasonal Unit Root Tests: A Unifying Approach, *Econometric Reviews,* 21, 221-241.

Paap, R., P.H. Franses, and H. Hoek, (1997), Mean Shifts, Unit Roots, and Forecasting Seasonal Time Series, *International Journal of Forecasting, 13*, 357-368.

Rodrigues, P.M.M., (2001), Near Seasonal Integration, *Econometric Theory, 17*, 70-86.

Sullivan, R., A. Timmerman and H. White, (1999), Data-Snooping, Technical Trading Rules and the Bootstrap, *Journal of Finance,* 54, 1647-1692.

Sullivan, R., A. Timmerman and H. White, (2001), Dangers of Data-driven Inference: The Case of Calendar Effects in Stock Returns, *Journal of Econometrics,* 104, 249-286.

Stock, J.H. and M.W. Watson, (1999), Business Cycle Fluctuations in US Macroeconomic Time Series, in J. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics*, Amsterdam, The Netherlands: Elsevier.

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, 265-279.

Teräsvirta, T., (1994), Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models, *Journal of the American Statistical Association*, 89, 208-218.

Teräsvirta, T., (1998), Modelling Economic Relations with Smooth Transition Regressions, in: *Handbook of Applied Economic Statistics*, New York, Marcel Dekker, pp. 507-552.

Watson, M.W., (1993), Measure of Fit for Calibrated Models, *Journal of Political Economy,* 101, 1011-1041.

West, K., (1996), Asymptotic Inference About Predictive Ability, *Econometrica*, 64, 1067-1084.

West, K., (2004), Forecast Evaluation, in:*Handbook of Economic Forecasting*, eds. Clive W.J. Granger, Graham Elliot and Allan Timmerman, Elsevier, Amsterdam, forthcoming.

White, H., (2000), A Reality Check for Data Snooping, *Econometrica*, 68, 1097-1126.

Table 1: Summary Statistics $^{(*)}$

Series	Mean	Std. Dev.	Skewness	Kurtosis	Jarque Bera	ADF	BF	HEGY
IP	67.32	25.39	0.41	2.24	29.09	-1.45*	40.12	-5.90
M1	583.73	397.19	0.43	1.62	60.70	-0.82*	5.43*	-4.47
M2	2224.39	1719.47	0.67	2.38	49.94	-1.94*	28.97	-4.48
M3	2918.99	2497.17	0.89	2.89	73.45	-1.66*	26.17	-5.88
R1	6.17	2.92	1.00	4.30	131.33	-1.61*	0.25*	-10.02
R2	7.02	2.55	0.98	3.62	97.52	-1.52*	0.07*	-7.86
R3	6.12	3.35	1.18	4.94	214.55	-2.55	4.03*	-7.27
CPI1	93.93	53.99	0.25	1.59	51.52	-1.36*	13.49	-7.20
CPI2	96.79	56.20	0.28	1.59	52.65	-1.60*	10.10	-5.95
CPI3	93.92	54.31	0.26	1.59	52.15	-1.02*	11.12	-9.02
CPI4	97.53	57.20	0.30	1.59	53.69	-1.15*	11.36	-6.60
CPI5	72.11	40.18	0.00	1.56	47.78	-0.55*	5.45*	-8.19
CPI6	94.12	52.05	0.20	1.60	48.48	-0.90*	1.08	-6.40

$^{(*)}$ Various summary measures for the 13 series examined are presented in the table. The series are: Industrial production (IP); three money stock variables including M1, M2 and M3; three interest rate variables including the 1-year treasury, constant maturity rate (R1), the 10-year treasury constant maturity rate (R2) and the effective federal funds Rate); and six CPI series including CPI for all urban consumers: all items (CPI1), all items less energy (CPI2), all items less food (CPI3), all items less food and energy (CPI4), energy (CPI5) and food (CPI6). Data are monthly U.S. figures for the period 1959:1-2004:12. Mean, Std. Dev. Skewness Kurtosis and Jarque-Bera statistics are calculated for the levels series. Unit root tests are carried out for log levels except for the interest rates. For the unit root tests the starred entries represent the cases where we fail to reject the null of unit root at 10% nominal level. HEGY test statistics are reported for seasonal frequency π - for tests at all the other seasonal frequencies and non seasonal unit root see Table 2. Further details are contained in Section 2.

Table 2: HEGY Test Statistics $^{(*)}$

	IP	M1	M2	M3	R1	R2	R3	CPI1	CPI2	CPI3	CPI4	CPI5	CPI6
π_1	-2.12*	-0.97*	-1.80*	-1.50*	-2.29*	-1.83*	-2.61*	-1.24*	-1.45*	-1.19*	-1.44*	-0.67*	-0.93*
π_2	-5.90	-4.47	-4.48	-5.88	-10.02	-7.86	-7.27	-7.20	-5.95	-9.02	-6.60	-8.19	-8.19
π_3	-4.55	-3.30	-0.36*	-0.80*	0.66*	-2.82	0.18*	-4.70	-3.78	-5.14	-4.24	-5.50	-4.91
π_4	-4.13	-6.00	-5.78	-8.64	-8.43	-8.96	-12.37	-7.79	-9.66	-8.00	-8.59	-8.44	-8.25
π_5	-6.26	-3.85	-4.35	-3.02	-11.54	-8.08	-6.42	-7.16	-6.35	-7.69	-7.47	-5.33	-6.63
π_6	4.26	1.16*	1.17*	2.84	6.92	7.18	6.19	5.65	4.02	7.83	4.28	9.38	5.21
π_7	-1.11*	-2.38*	0.08*	0.43*	-2.93*	-4.98	0.63*	-1.65*	-2.25*	-1.39*	0.05*	-3.15	-2.86*
π_8	-4.41	-6.70	-7.01	-9.24	-9.86	-9.14	-13.17	-8.37	-7.87	-7.83	-6.63	-8.48	-8.17
π_9	-6.14	-5.40	-4.77	-6.53	-9.15	-8.48	-10.03	-8.70	-7.83	-8.98	-8.51	-8.59	-8.30
π_{10}	1.50*	4.12	5.73	6.42	6.96	6.96	6.25	0.63*	2.34	2.35	3.16	4.29	2.11*
π_{11}	0.78*	0.20*	0.87*	2.54*	-2.98*	-3.27	-2.14*	-0.94*	-1.00*	0.20*	0.79*	-1.22*	-2.49*
π_{12}	-7.58	-6.43	-7.17	-6.77	-7.24	-9.06	-7.91	-8.46	-7.91	-8.10	-7.65	-7.88	-8.15
$F_{3,4}$	18.96	23.49	16.79	37.67	35.84	44.15	76.49	41.34	53.77	45.16	45.90	50.70	46.06
$F_{5,6}$	29.13	8.00	10.13	8.62	90.64	61.04	39.77	42.55	28.35	63.50	37.01	63.68	36.26
$F_{7,8}$	10.26	25.03	24.55	43.30	52.87	53.94	86.96	35.83	33.10	31.26	22.03	40.13	37.05
$F_{9,10}$	20.00	24.53	29.69	47.98	70.30	72.02	71.00	37.88	33.52	43.01	41.94	48.61	36.58
$F_{11,12}$	29.63	20.80	26.90	27.78	30.01	47.18	32.60	35.95	31.59	33.01	30.10	31.60	36.51

$^{(*)}$ See Table 1. Starred entries represent cases where we fail to reject the null of a unit root at the 10% nominal level. In order to test the hypothesis of a unit root at various seasonal frequencies, the equation to be estimated is $\theta(L)\, y_{13t} = \sum_{k=1}^{12} \pi_k y_{k,t-1} + \epsilon_t$, where $\theta(L)$ is a polynomial in the lag operator L, and $y_{13t} = \left(1 - B^{12}\right) y_t$. For our monthly data, the frequencies for seasonal unit roots are π, $\pm\frac{\pi}{2}$, $\pm\frac{2\pi}{3}$, $\pm\frac{\pi}{3}$, $\pm\frac{5\pi}{6}$, and $\pm\frac{\pi}{6}$. Thus, testing for the presence of seasonal unit root using the HEGY test involves testing the null that $\pi_k = 0$, for $k > 1$. Testing the null that $\pi_1 = 0$ corresponds to a test for the null of a non seasonal unit root. To test the presence of a unit root at any seasonal frequency, we first need to test $\pi_k = 0$, for $k = 2$ against the alternative that $\pi_k < 0$. For other roots, we test $\pi_k = 0$ for even $k\,(> 2)$ against a two sided alternative. If the null fails to reject for even k, we test $\pi_{k-1} = 0$ against the alternative that $\pi_{k-1} < 0$. As we estimate the equation using LS, the tests can be carried out as a sequence of t-tets, or a set of F-tests for the even/odd pairs $\pi_k = \pi_{k-1} = 0$. Estimation details and critical values are reported in Beaulieu and Miron (1993). Note that the results in this table further confirm the findings of Beaulieu and Miron (1993) that the data reject the null of seasonal unit roots at most of frequencies.

Table 3: Estimation Results Using Data for the Period 1959:1-2004:12 (*)

	RW			*DS*			*SUR*			*PAR*	
	DW	*LM*	$\overline{R}^2$	*DW*	*LM*	$\overline{R}^2$	*DW*	*LM*	$\overline{R}^2$	*DW*	*LM*
IP	2.49	32.85*	0.75	1.95	5.96*	0.93	2.02	1.63	0.99	1.71	11.50*
M1	2.19	5.20*	0.76	2.01	0.81	0.96	2.01	0.63	0.99	1.74	9.30*
M2	1.94	0.38	0.62	2.02	0.50	0.98	2.00	0.03	0.99	2.05	1.46
M3	1.50*	32.76*	0.62	1.97	5.84*	0.99	1.98	0.13	0.99	2.14	7.66*
R1	1.30*	67.70*	0.18	1.99	0.04	0.88	1.98	1.06	0.98	1.99	0.18
R2	1.37*	53.83*	0.15	1.96	3.51*	0.90	1.99	0.02	0.99	1.90	9.96*
R3	1.24*	79.31*	0.15	2.01	0.50	0.89	2.02	2.63	0.98	2.00	0.01
CPI1	0.87*	173.99*	0.47	2.01	1.53	0.99	1.98	1.62	0.99	2.06	4.84*
CPI2	0.88*	172.07*	0.49	2.00	1.26	0.99	2.01	0.01	0.99	2.10	10.16*
CPI3	0.73*	219.26*	0.52	2.01	1.36	0.99	2.07	5.34*	0.99	2.29	26.30*
CPI4	0.75*	214.07*	0.53	2.05	4.52*	0.99	2.01	0.15	0.99	2.08	8.75*
CPI5	1.15*	98.83*	0.28	1.99	0.05	0.95	2.00	0.20	0.99	1.87	9.88*
CPI6	1.38*	51.76*	0.17	2.01	1.49	0.96	2.02	1.73	0.99	2.07	4.88*

(*) See notes to Table 1. Entries in this table include adjusted R^2 values, Durbin-Watson (DW) statistics and first order serial correlation LM test statistics (LM), for all four estimated models (i.e. RW, DS, SUR, and PAR) and for each series (for further details see Section 2 of the paper). Starred entries correspond to cases where the null hypothesis of no serial correlation is rejected at the nominal 10% level.

Table 4: Estimation Results for the Period 1991:1-2004:12 (*)

	RW			*DS*			*SUR*			*PAR*		*ADF*	*BF*	*HEGY*
	DW	*LM*	$\overline{R}^2$	*DW*	*LM*	$\overline{R}^2$	*DW*	*LM*	$\overline{R}^2$	*DW*	*LM*			
IP	2.95*	37.34*	0.90	1.94	4.20	0.93	2.07	0.26	0.99	2.21	1.88	-1.58*	0.09*	-3.79
M1	2.12	0.86	0.71	2.03	0.52	0.96	1.91	0.33	0.99	1.70	3.54	-2.07*	6.32*	-2.45*
M2	1.99	0.00	0.61	2.09	3.17	0.97	2.04	1.43	0.99	1.40*	13.80*	-0.71*	0.30*	-2.98
M3	1.32*	19.36*	0.49	2.13	3.02	0.98	1.93	1.73	0.99	0.98*	40.18*	-0.51*	3.62*	-2.82
R1	1.11*	31.81*	0.22	1.99	0.27	0.95	2.06	0.20	0.98	1.04*	35.98*	-1.71*	0.04*	-5.03
R2	1.44*	12.92*	0.10	1.92	3.33	0.89	2.01	0.02	0.95	1.4*	14.10*	-1.91*	0.58*	-4.88
R3	0.87*	49.728	0.36	2.08	5.84	0.98	2.06	0.22	0.99	2.24	6.97*	-1.85*	0.14*	-3.33
CPI1	1.46*	10.89*	0.43	1.94	1.60	0.87	2.01	1.19	0.99	1.73	3.10	-1.43*	10.07	-4.53
CPI2	1.42*	12.43*	0.74	2.01	2.53	0.95	2.05	0.68	0.99	2.10	2.05	-1.58*	84.94	-3.34
CPI3	1.43*	11.85*	0.45	1.93	2.35	0.88	2.03	1.50	0.99	1.76	2.34	-1.61*	76.39	-4.29
CPI4	1.35*	15.65*	0.74	2.03	4.97	0.96	2.15	2.56	0.99	2.28	3.65	-1.59*	96.32	-3.46
CPI5	1.36*	15.93*	0.33	1.88	4.61	0.9	2.02	0.13	0.98	1.45*	11.88*	-0.36*	1.22*	-4.09
CPI6	2.01	0.02	0.17	1.99	1.65	0.83	1.88	0.84	0.99	2.14	1.71	0.39*	109.35	-3.93

(*) See notes to Table 3. Note that ADF, BF, and HEGY test statistics are also reported, as those reported in Table 1 are for the period 1959:1-2004:12.

Table 5a: Simulation Experiment Results I [*]

Target Simulation Variable is $\ln(IP_t)$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
				$Panel\ 1: RW$					
Mean	4.13	4.58	4.82	4.97	5.13	5.30	5.45	5.13	6.32
Std. Dev.	0.40	0.15	0.13	0.22	0.29	0.33	0.41	0.37	1.08
Skewness	-0.26	-0.41	0.77	0.46	0.36	0.08	0.14	0.81	0.41
Kurtosis	2.29	1.64	4.37	3.88	2.84	2.78	2.95	3.41	2.42
Min.	3.30	4.31	4.47	4.33	4.42	4.40	4.25	4.25	4.25
Max.	4.78	4.78	5.54	5.86	5.98	6.28	6.59	6.59	9.64
				$Panel\ 2: DS$					
Mean	4.13	4.58	4.87	5.05	5.25	5.47	5.67	5.26	6.75
Std. Dev.	0.40	0.15	0.10	0.14	0.17	0.20	0.22	0.33	1.19
Skewness	-0.26	-0.41	0.52	0.24	-0.08	-0.24	-0.09	0.35	0.12
Kurtosis	2.29	1.64	2.89	2.93	2.83	3.40	2.81	2.29	1.96
Min.	3.30	4.31	4.59	4.67	4.76	4.79	4.95	4.59	4.59
Max.	4.78	4.78	5.23	5.53	5.76	6.10	6.41	6.41	9.73
				$Panel\ 3: SUR$					
Mean	4.13	4.58	4.85	5.01	5.20	5.37	5.55	5.20	6.45
Std. Dev.	0.40	0.15	0.09	0.16	0.20	0.23	0.27	0.32	1.05
Skewness	-0.26	-0.41	0.71	0.45	0.17	0.21	0.12	0.57	0.22
Kurtosis	2.29	1.64	3.85	3.19	2.95	2.89	2.72	2.71	2.10
Min.	3.30	4.31	4.62	4.63	4.61	4.72	4.83	4.61	4.61
Max.	4.78	4.78	5.29	5.60	5.88	6.11	6.45	6.45	9.71
				$Panel\ 4: PAR$					
Mean	4.13	4.58	5.68	5.68	5.67	5.65	5.65	5.66	5.67
Std. Dev.	0.40	0.15	0.18	0.19	0.18	0.18	0.18	0.18	0.17
Skewness	-0.26	-0.41	-1.02	-0.74	-0.74	-0.81	-0.76	-0.80	-0.46
Kurtosis	2.29	1.64	7.00	5.51	4.88	5.20	5.32	5.54	4.38
Min.	3.30	4.31	4.74	4.81	4.82	4.83	4.79	4.74	4.74
Max.	4.78	4.78	6.27	6.25	6.16	6.09	6.13	6.27	6.27

[*] Models given under Panels 1-4 are used to form simulations of the "target simulation variable". Simulation models are estimated using the full data sample (i.e. 1959:1-2004:12). Columns 2 and 3 report summary statistics based on the historical data. The rest of the columns report summary statistics for various subsamples of simulated samples of length 1200 observations, averaged across 100 simulation paths. For further details, see Sections 2 and 3.

Table 5b: Simulation Experiment Results II $^{(*)}$

Target Simulation Variable is $\ln(IP_t) - \ln(IP_{t-1})$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
				$Panel\ 1: RW$					
Mean	0.003	0.003	0.003	0.003	0.003	0.002	0.003	0.003	0.003
Std. Dev.	0.023	0.021	0.023	0.023	0.023	0.023	0.023	0.023	0.023
Skewness	-0.545	0.167	0.053	-0.024	0.046	-0.020	0.004	0.011	-0.001
Kurtosis	3.462	3.025	2.976	2.938	2.955	2.957	2.930	2.954	3.014
Min.	-0.069	-0.045	-0.069	-0.080	-0.078	-0.086	-0.085	-0.086	-0.112
Max.	0.053	0.052	0.083	0.083	0.089	0.080	0.093	0.093	0.098
				$Panel\ 2: DS$					
Mean	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Std. Dev.	0.023	0.021	0.022	0.022	0.022	0.022	0.022	0.022	0.022
Skewness	-0.545	0.167	-0.510	-0.478	-0.516	-0.496	-0.482	-0.496	-0.487
Kurtosis	3.462	3.025	3.144	3.016	3.192	3.077	3.117	3.110	3.108
Min.	-0.069	-0.045	-0.068	-0.075	-0.074	-0.074	-0.078	-0.078	-0.085
Max.	0.053	0.052	0.062	0.061	0.060	0.068	0.063	0.068	0.074
				$Panel\ 3: SUR$					
Mean	0.003	0.003	0.003	0.003	0.002	0.003	0.003	0.003	0.003
Std. Dev.	0.023	0.021	0.028	0.039	0.048	0.055	0.061	0.048	0.088
Skewness	-0.545	0.167	0.017	-0.085	-0.058	0.005	0.024	-0.003	0.054
Kurtosis	3.462	3.025	3.079	2.980	2.891	2.946	3.003	3.592	3.950
Min.	-0.069	-0.045	-0.091	-0.169	-0.170	-0.188	-0.208	-0.208	-0.369
Max.	0.053	0.052	0.106	0.134	0.165	0.190	0.201	0.201	0.428
				$Panel\ 4: PAR$					
Mean	0.003	0.003	0.015	0.000	0.000	0.000	0.000	0.003	0.001
Std. Dev.	0.023	0.021	0.120	0.027	0.027	0.027	0.027	0.059	0.038
Skewness	-0.545	0.167	7.336	0.307	0.300	0.259	0.260	12.741	12.378
Kurtosis	3.462	3.025	59.663	2.597	2.593	2.577	2.578	212.187	321.879
Min.	-0.069	-0.045	-0.073	-0.068	-0.07	-0.075	-0.066	-0.075	-0.078
Max.	0.053	0.052	1.371	0.076	0.078	0.079	0.076	1.371	1.371

$^{(*)}$ See notes to Table 5a.

Table 5c: Simulation Experiment Results III $^{(*)}$

Target Simulation Variable is $\ln(IP_t) - \ln(IP_{t-12})$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
				Panel 1 : RW					
Mean	0.031	0.032	0.027	0.032	0.032	0.031	0.030	0.030	0.032
Std. Dev.	0.046	0.030	0.076	0.083	0.076	0.081	0.077	0.079	0.079
Skewness	-0.758	-1.100	0.083	-0.06	-0.041	-0.080	0.001	-0.021	-0.005
Kurtosis	3.732	3.818	3.243	2.977	2.985	2.978	2.938	3.028	3.001
Min.	-0.135	-0.058	-0.246	-0.241	-0.233	-0.259	-0.245	-0.259	-0.300
Max.	0.131	0.083	0.333	0.335	0.299	0.343	0.284	0.343	0.359
				Panel 2 : DS					
Mean	0.031	0.032	0.037	0.041	0.041	0.044	0.038	0.040	0.040
Std. Dev.	0.046	0.030	0.044	0.049	0.05	0.048	0.05	0.048	0.048
Skewness	-0.758	-1.100	0.047	-0.113	-0.008	0.098	0.035	0.012	0.017
Kurtosis	3.732	3.818	3.289	2.833	2.727	2.858	3.05	2.950	2.965
Min.	-0.135	-0.058	-0.108	-0.145	-0.115	-0.114	-0.118	-0.145	-0.167
Max.	0.131	0.083	0.217	0.192	0.197	0.223	0.250	0.250	0.250
				Panel 3 : SUR					
Mean	0.031	0.032	0.034	0.034	0.032	0.033	0.036	0.034	0.034
Std. Dev.	0.046	0.030	0.042	0.047	0.046	0.047	0.044	0.045	0.046
Skewness	-0.758	-1.100	-0.121	-0.04	0.018	-0.108	0.023	-0.047	-0.024
Kurtosis	3.732	3.818	3.078	2.754	3.012	2.933	2.872	2.938	2.991
Min.	-0.135	-0.058	-0.124	-0.136	-0.119	-0.158	-0.124	-0.158	-0.158
Max.	0.131	0.083	0.16	0.194	0.184	0.188	0.199	0.199	0.236
				Panel 4 : PAR					
Mean	0.031	0.032	0.182	0.001	-0.004	-0.002	0.002	0.036	0.009
Std. Dev.	0.046	0.030	0.377	0.033	0.034	0.032	0.034	0.186	0.098
Skewness	-0.758	-1.100	1.609	-0.032	0.087	-0.047	0.050	4.680	8.590
Kurtosis	3.732	3.818	3.814	3.216	2.953	2.767	2.875	24.342	86.424
Min.	-0.135	-0.058	-0.114	-0.121	-0.126	-0.102	-0.119	-0.126	-0.126
Max.	0.131	0.083	1.303	0.136	0.113	0.104	0.113	1.303	1.303

$^{(*)}$ See notes to Table 5a.

Table 6a: Simulation Experiment Results IV [(*)]

Target Simulation Variable is $\ln(M1_t)$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
				Panel 1 : RW					
Mean	6.10	6.99	7.37	7.62	7.87	8.14	8.39	7.88	9.76
Std. Dev.	0.77	0.09	0.11	0.16	0.20	0.24	0.25	0.41	1.49
Skewness	-0.08	-1.11	0.63	0.16	0.12	0.17	0.20	0.31	0.07
Kurtosis	1.49	4.41	2.92	2.61	2.89	2.66	2.88	2.27	1.91
Min.	4.92	6.71	7.09	7.18	7.32	7.53	7.68	7.09	7.09
Max.	7.24	7.18	7.84	8.13	8.52	8.86	9.22	9.22	13.41
				Panel 2 : DS					
Mean	6.10	6.99	7.35	7.62	7.89	8.17	8.45	7.9	9.96
Std. Dev.	0.77	0.09	0.09	0.11	0.12	0.14	0.16	0.41	1.60
Skewness	-0.08	-1.11	0.46	0.16	0.09	0.25	-0.01	0.10	0.02
Kurtosis	1.49	4.41	2.53	2.76	2.76	3.05	2.5	1.94	1.83
Min.	4.92	6.71	7.17	7.32	7.54	7.73	8.00	7.17	7.17
Max.	7.24	7.18	7.66	8.05	8.30	8.72	8.87	8.87	13.3
				Panel 3 : SUR					
Mean	6.10	6.99	7.34	7.58	7.81	8.04	8.28	7.81	9.58
Std. Dev.	0.77	0.09	0.11	0.20	0.31	0.36	0.40	0.44	1.50
Skewness	-0.08	-1.11	0.99	0.41	0.20	-0.02	-0.27	0.61	0.25
Kurtosis	1.49	4.41	3.99	2.96	2.66	2.60	3.00	2.59	2.11
Min.	4.92	6.71	7.09	7.12	7.07	7.05	6.96	6.96	6.96
Max.	7.24	7.18	7.87	8.42	8.77	9.04	9.45	9.45	14.08
				Panel 4 : PAR					
Mean	6.10	6.99	18.00	18.00	18.01	18.01	18.02	18.01	18.05
Std. Dev.	0.77	0.09	1.47	1.44	1.42	1.40	1.37	1.42	1.25
Skewness	-0.08	-1.11	-5.31	-5.31	-5.3	-5.3	-5.33	-5.31	-5.33
Kurtosis	1.49	4.41	34.04	34.06	33.94	33.92	34.34	34.15	35.23
Min.	4.92	6.71	7.22	7.45	7.62	7.79	7.99	7.22	7.22
Max.	7.24	7.18	19.03	19.04	18.99	18.97	18.99	19.04	19.08

[(*)] See notes to Table 5a.

Table 6b: Simulation Experiment Results V [*]

Target Simulation Variable is $\ln(M1_t) - \ln(M1_{t-1})$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
				Panel 1 : RW					
Mean	0.004	0.003	0.004	0.004	0.004	0.004	0.004	0.004	0.004
Std. Dev.	0.015	0.014	0.015	0.015	0.015	0.015	0.015	0.015	0.015
Skewness	-0.392	-0.045	-0.028	0.018	-0.058	0.06	-0.018	-0.005	-0.009
Kurtosis	2.952	2.659	3.070	3.019	3.027	2.992	2.947	3.011	2.999
Min.	-0.043	-0.029	-0.055	-0.050	-0.053	-0.05	-0.049	-0.055	-0.066
Max.	0.045	0.042	0.058	0.057	0.068	0.059	0.064	0.068	0.073
				Panel 2 : DS					
Mean	0.004	0.003	0.005	0.004	0.005	0.005	0.005	0.005	0.005
Std. Dev.	0.015	0.014	0.015	0.015	0.015	0.015	0.015	0.015	0.015
Skewness	-0.392	-0.045	-0.423	-0.384	-0.350	-0.385	-0.387	-0.386	-0.376
Kurtosis	2.952	2.659	2.898	2.851	2.925	2.884	2.857	2.884	2.865
Min.	-0.043	-0.029	-0.046	-0.052	-0.044	-0.048	-0.045	-0.052	-0.052
Max.	0.045	0.042	0.049	0.046	0.051	0.045	0.046	0.051	0.054
				Panel 3 : SUR					
Mean	0.004	0.003	0.004	0.004	0.005	0.004	0.004	0.004	0.004
Std. Dev.	0.015	0.014	0.018	0.025	0.029	0.034	0.037	0.029	0.055
Skewness	-0.392	-0.045	-0.181	-0.125	-0.176	-0.088	-0.055	-0.102	-0.086
Kurtosis	2.952	2.659	3.058	3.124	3.279	3.153	3.160	3.749	3.824
Min.	-0.043	-0.029	-0.070	-0.086	-0.102	-0.121	-0.124	-0.124	-0.241
Max.	0.045	0.042	0.074	0.096	0.098	0.114	0.133	0.133	0.278
				Panel 4 : PAR					
Mean	0.004	0.003	0.179	0.000	0.000	0.000	0.000	0.036	0.009
Std. Dev.	0.015	0.014	1.373	0.048	0.048	0.048	0.048	0.620	0.313
Skewness	-0.392	-0.045	7.643	-0.287	-0.293	-0.278	-0.290	17.312	33.823
Kurtosis	2.952	2.659	59.666	2.998	2.992	2.982	2.966	303.338	1176.083
Min.	-0.043	-0.029	-0.124	-0.123	-0.123	-0.120	-0.120	-0.124	-0.132
Max.	0.045	0.042	11.607	0.105	0.102	0.102	0.098	11.607	11.607

[*] See notes to Table 5a.

Table 6c: Simulation Experiment Results VI (*)

Target Simulation Variable is $\ln(M1_t) - \ln(M1_{t-12})$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
			Panel 1 : RW						
Mean	0.050	0.034	0.053	0.051	0.049	0.050	0.051	0.051	0.051
Std. Dev.	0.038	0.048	0.051	0.053	0.054	0.054	0.051	0.053	0.053
Skewness	-0.006	0.276	-0.136	-0.020	0.071	-0.041	0.040	-0.018	0.009
Kurtosis	3.282	2.149	3.196	3.038	3.192	2.836	2.832	3.025	3.006
Min.	-0.053	-0.053	-0.180	-0.135	-0.158	-0.138	-0.134	-0.180	-0.180
Max.	0.162	0.135	0.224	0.235	0.242	0.223	0.207	0.242	0.278
			Panel 2 : DS						
Mean	0.050	0.034	0.054	0.053	0.056	0.055	0.052	0.054	0.054
Std. Dev.	0.038	0.048	0.028	0.027	0.029	0.027	0.029	0.028	0.028
Skewness	-0.006	0.276	0.020	0.023	-0.094	-0.049	-0.092	-0.042	-0.027
Kurtosis	3.282	2.149	2.784	2.856	3.122	2.961	2.917	2.943	2.984
Min.	-0.053	-0.053	-0.034	-0.040	-0.052	-0.033	-0.060	-0.060	-0.060
Max.	0.162	0.135	0.145	0.147	0.161	0.145	0.162	0.162	0.185
			Panel 3 : SUR						
Mean	0.050	0.034	0.050	0.049	0.046	0.050	0.050	0.049	0.049
Std. Dev.	0.038	0.048	0.033	0.036	0.038	0.039	0.037	0.036	0.037
Skewness	-0.006	0.276	0.332	-0.088	0.245	0.037	0.312	0.155	0.124
Kurtosis	3.282	2.149	3.457	2.806	3.108	3.144	2.781	3.068	3.090
Min.	-0.053	-0.053	-0.053	-0.077	-0.071	-0.076	-0.051	-0.077	-0.082
Max.	0.162	0.135	0.175	0.153	0.173	0.180	0.178	0.180	0.230
			Panel 4 : PAR						
Mean	0.050	0.034	2.165	0.002	0.000	0.000	0.001	0.434	0.109
Std. Dev.	0.038	0.048	4.379	0.024	0.024	0.024	0.024	2.141	1.087
Skewness	-0.006	0.276	1.545	0.064	-0.058	0.078	0.192	4.772	9.991
Kurtosis	3.282	2.149	3.410	3.104	2.977	3.041	3.196	23.851	101.167
Min.	-0.053	-0.053	-0.074	-0.075	-0.098	-0.080	-0.079	-0.098	-0.098
Max.	0.162	0.135	11.757	0.097	0.089	0.094	0.094	11.757	11.757

(*) See notes to Table 5a.

Table 7a: Simulation Experiment Results VII $^{(*)}$

Target Simulation Variable is $R1_t$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
				Panel 1 : RW					
Mean	6.17	4.50	2.42	2.16	2.46	2.28	2.13	2.29	2.56
Std. Dev.	2.92	1.59	2.44	4.11	5.39	6.22	6.71	5.21	11.94
Skewness	1.00	-0.72	-0.17	-0.15	-0.10	-0.06	0.12	-0.03	-0.14
Kurtosis	4.29	2.39	3.61	2.59	2.68	2.81	3.10	3.69	4.30
Min.	1.01	1.01	-7.45	-10.23	-15.16	-16.38	-18.32	-18.32	-44.21
Max.	16.72	7.14	10.56	12.34	19.37	21.81	23.22	23.22	50.44
				Panel 2 : DS					
Mean	6.17	4.50	2.62	2.62	2.20	1.87	1.67	2.2	3.26
Std. Dev.	2.92	1.59	2.49	4.46	6.05	7.42	8.86	6.28	12.5
Skewness	1.00	-0.72	-0.06	0.25	0.14	0.00	-0.03	-0.09	0.10
Kurtosis	4.29	2.39	3.40	3.27	4.01	3.44	3.07	4.70	3.49
Min.	1.01	1.01	-5.39	-11.98	-20.85	-21.65	-24.54	-24.54	-47.07
Max.	16.72	7.14	10.87	19.43	23.24	26.93	28.98	28.98	48.70
				Panel 3 : SUR					
Mean	6.17	4.50	2.90	3.01	2.98	3.63	3.74	3.25	4.35
Std. Dev.	2.92	1.59	3.55	6.42	8.23	9.70	11.01	8.22	17.56
Skewness	1.00	-0.72	-0.22	-0.12	0.01	-0.16	-0.25	-0.12	-0.29
Kurtosis	4.29	2.39	4.47	2.89	2.53	2.74	2.80	3.60	4.67
Min.	1.01	1.01	-13.95	-16.69	-18.76	-27.50	-34.64	-34.64	-80.31
Max.	16.72	7.14	14.88	23.55	28.26	30.20	33.86	33.86	75.92
				Panel 4 : PAR					
Mean	6.17	4.50	5.30	5.26	5.90	5.72	5.90	5.62	5.71
Std. Dev.	2.92	1.59	3.34	3.34	3.14	3.44	3.2	3.31	3.34
Skewness	1.00	-0.72	0.11	0.22	0.09	-0.04	-0.03	0.06	0.01
Kurtosis	4.29	2.39	2.95	3.05	3.45	2.79	2.62	2.94	2.95
Min.	1.01	1.01	-4.68	-5.21	-3.65	-5.53	-4.28	-5.53	-6.9
Max.	16.72	7.14	15.58	17.23	19.27	17.22	15.54	19.27	19.5

$^{(*)}$ See notes to Table 5a.

Table 7b: Simulation Experiment Results VIII $^{(*)}$

Target Simulation Variable is $R1_t - R1_{t-1}$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
				Panel 1 : RW					
Mean	-0.005	-0.011	0.000	0.001	-0.005	0.000	-0.009	-0.003	-0.001
Std. Dev.	0.472	0.227	0.474	0.465	0.469	0.467	0.467	0.468	0.469
Skewness	-1.368	-0.243	-0.023	0.003	0.014	0.02	0.022	0.007	-0.005
Kurtosis	17.314	3.914	3.034	3.073	3.053	3.029	2.916	3.022	2.994
Min.	-3.910	-0.790	-1.722	-1.800	-1.857	-1.706	-1.723	-1.857	-1.939
Max.	1.900	0.600	1.701	1.619	1.686	1.621	1.730	1.730	2.117
				Panel 2 : DS					
Mean	-0.005	-0.011	0.00	0.008	-0.005	-0.010	0.001	-0.001	-0.002
Std. Dev.	0.472	0.227	0.465	0.463	0.479	0.468	0.477	0.471	0.469
Skewness	-1.368	-0.243	0.015	-0.046	-0.076	-0.038	0.023	-0.025	-0.012
Kurtosis	17.314	3.914	2.931	3.056	2.979	2.937	3.050	2.995	2.985
Min.	-3.910	-0.790	-1.647	-1.689	-1.882	-1.888	-1.819	-1.888	-1.888
Max.	1.900	0.600	1.736	1.933	1.934	1.582	2.024	2.024	2.052
				Panel 3 : SUR					
Mean	-0.005	-0.011	0.015	0.017	0.006	-0.004	0.011	0.009	0.000
Std. Dev.	0.472	0.227	1.207	1.927	2.412	2.833	3.182	2.414	4.687
Skewness	-1.368	-0.243	0.096	0.081	0.004	-0.041	-0.012	-0.010	-0.082
Kurtosis	17.314	3.914	3.587	3.249	3.120	2.897	2.937	3.804	3.966
Min.	-3.910	-0.790	-4.435	-7.302	-9.248	-11.962	-11.61	-11.962	-22.027
Max.	1.900	0.600	5.908	9.024	9.509	9.636	12.569	12.569	21.119
				Panel 4 : PAR					
Mean	-0.005	-0.011	0.056	-0.001	-0.001	0.007	-0.004	0.011	0.003
Std. Dev.	0.472	0.227	0.750	0.478	0.474	0.499	0.481	0.547	0.499
Skewness	-1.368	-0.243	4.621	0.031	0.044	0.036	-0.138	2.428	0.786
Kurtosis	17.314	3.914	45.726	3.928	3.434	3.896	3.842	34.858	15.085
Min.	-3.910	-0.790	-3.556	-2.209	-1.792	-2.438	-2.298	-3.556	-3.556
Max.	1.900	0.600	9.932	2.345	1.816	2.765	2.056	9.932	9.932

$^{(*)}$ See notes to Table 5a.

Table 7c: Simulation Experiment Results IX (*)

Target Simulation Variable is $R1_t - R1_{t-12}$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
				$Panel\ 1: RW$					
Mean	-0.052	-0.306	0.095	0.005	-0.019	-0.036	-0.066	-0.004	-0.021
Std. Dev.	1.773	1.398	1.620	1.618	1.636	1.609	1.689	1.636	1.619
Skewness	0.034	0.109	0.022	0.002	0.018	-0.003	-0.027	0.000	0.014
Kurtosis	4.734	3.162	2.888	2.932	2.972	2.997	2.957	2.956	2.984
Min.	-6.060	-3.910	-5.349	-5.307	-6.389	-5.645	-5.670	-6.389	-6.879
Max.	7.070	3.530	5.868	5.427	5.661	5.682	5.334	5.868	6.957
				$Panel\ 2: DS$					
Mean	-0.052	-0.306	0.175	0.029	-0.084	-0.141	-0.078	-0.020	0.002
Std. Dev.	1.773	1.398	1.724	1.871	1.815	1.777	1.740	1.790	1.818
Skewness	0.034	0.109	-0.120	0.012	0.036	0.082	-0.025	-0.002	0.019
Kurtosis	4.734	3.162	2.922	2.917	2.769	3.070	2.808	2.894	2.950
Min.	-6.060	-3.910	-5.781	-6.586	-6.060	-6.756	-5.827	-6.756	-8.051
Max.	7.070	3.530	5.228	6.563	5.865	5.397	5.972	6.563	7.660
				$Panel\ 3: SUR$					
Mean	-0.052	-0.306	0.120	0.105	-0.009	-0.038	0.005	0.037	-0.009
Std. Dev.	1.773	1.398	1.761	1.801	1.806	1.886	1.770	1.806	1.777
Skewness	0.034	0.109	0.060	-0.078	-0.053	-0.091	0.020	-0.035	-0.013
Kurtosis	4.734	3.162	3.112	2.939	3.112	3.400	2.947	3.128	3.028
Min.	-6.060	-3.910	-5.772	-6.191	-5.896	-7.713	-7.207	-7.713	-7.713
Max.	7.070	3.530	6.984	5.787	7.590	7.173	6.417	7.590	7.590
				$Panel\ 4: PAR$					
Mean	-0.052	-0.306	0.945	-0.118	-0.064	-0.023	0.038	0.155	0.045
Std. Dev.	1.773	1.398	2.881	1.977	1.915	1.880	1.916	2.185	1.977
Skewness	0.034	0.109	0.952	-0.003	0.038	-0.003	-0.135	0.639	0.205
Kurtosis	4.734	3.162	4.269	3.011	2.874	2.847	3.029	4.936	3.915
Min.	-6.060	-3.910	-6.796	-6.414	-6.576	-6.002	-8.798	-8.798	-10.45
Max.	7.070	3.530	13.827	8.037	7.122	7.708	7.282	13.827	13.827

(*) See notes to Table 5a.

Table 8a: Simulation Experiment Results X (*)

Target Simulation Variable is $\ln(CPI1_t)$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
				Panel 1 : RW					
Mean	4.34	5.07	5.35	5.56	5.77	5.98	6.18	5.77	7.31
Std. Dev.	0.66	0.09	0.06	0.07	0.07	0.07	0.08	0.30	1.19
Skewness	-0.19	-0.10	0.12	0.10	0.11	0.09	0.11	0.02	0.01
Kurtosis	1.46	1.83	1.99	2.23	2.40	2.44	2.54	1.83	1.81
Min.	3.36	4.90	5.25	5.40	5.59	5.78	5.98	5.25	5.25
Max.	5.25	5.22	5.55	5.76	5.99	6.18	6.40	6.40	9.66
				Panel 2 : DS					
Mean	4.34	5.07	5.66	6.93	8.36	9.78	11.21	8.39	19.03
Std. Dev.	0.66	0.09	0.30	0.43	0.45	0.47	0.48	2.02	8.20
Skewness	-0.19	-0.10	0.38	0.11	0.04	0.07	0.09	0.08	0.01
Kurtosis	1.46	1.83	1.97	2.03	2.16	2.28	2.34	1.77	1.80
Min.	3.36	4.90	5.25	5.98	7.25	8.57	10.04	5.25	5.25
Max.	5.25	5.22	6.43	7.97	9.52	10.97	12.55	12.55	34.56
				Panel 3 : SUR					
Mean	4.34	5.07	5.35	5.56	5.76	5.96	6.16	5.76	7.22
Std. Dev.	0.66	0.09	0.08	0.15	0.22	0.29	0.36	0.37	1.25
Skewness	-0.19	-0.10	0.91	0.53	0.36	0.19	0.15	0.77	0.30
Kurtosis	1.46	1.83	3.54	3.54	3.33	2.52	2.45	3.01	2.22
Min.	3.36	4.90	5.23	5.22	5.21	5.29	5.37	5.21	5.21
Max.	5.25	5.22	5.66	6.23	6.56	6.84	7.29	7.29	11.04
				Panel 4 : PAR					
Mean	4.34	5.07	9.84	9.84	9.85	9.85	9.86	9.85	9.86
Std. Dev.	0.66	0.09	0.59	0.57	0.55	0.54	0.53	0.56	0.47
Skewness	-0.19	-0.10	-5.06	-4.91	-4.73	-4.51	-4.43	-4.77	-4.21
Kurtosis	1.46	1.83	37.14	35.16	33.66	31.8	31.24	34.38	31.92
Min.	3.36	4.90	5.25	5.47	5.61	5.83	5.96	5.25	5.25
Max.	5.25	5.22	10.59	10.55	10.59	10.57	10.6	10.6	10.72

(*) See notes to Table 5a.

Table 8b: Simulation Experiment Results XI (*)

Target Simulation Variable is $\ln(CPI1_t) - \ln(CPI1_{t-1})$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
				$Panel\ 1: RW$					
Mean	0.003	0.002	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Std. Dev.	0.003	0.002	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Skewness	0.802	-0.038	0.060	0.024	0.013	0.028	-0.012	0.022	0.007
Kurtosis	4.162	3.155	3.043	3.019	3.039	3.057	3.084	3.051	2.992
Min.	-0.005	-0.004	-0.007	-0.008	-0.009	-0.008	-0.009	-0.009	-0.010
Max.	0.018	0.008	0.019	0.016	0.015	0.015	0.017	0.019	0.019
				$Panel\ 2: DS$					
Mean	0.003	0.002	0.016	0.023	0.023	0.023	0.024	0.022	0.023
Std. Dev.	0.003	0.002	0.006	0.003	0.003	0.003	0.003	0.005	0.004
Skewness	0.802	-0.038	-0.417	0.044	0.033	-0.120	-0.008	-1.152	-0.840
Kurtosis	4.162	3.155	2.576	2.991	2.908	3.082	2.836	5.214	5.782
Min.	-0.005	-0.004	-0.004	0.010	0.012	0.010	0.011	-0.004	-0.004
Max.	0.018	0.008	0.031	0.036	0.035	0.035	0.034	0.036	0.037
				$Panel\ 3: SUR$					
Mean	0.003	0.002	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Std. Dev.	0.003	0.002	0.006	0.009	0.011	0.013	0.014	0.011	0.022
Skewness	0.802	-0.038	-0.064	-0.024	-0.005	0.043	-0.066	-0.013	0.034
Kurtosis	4.162	3.155	3.514	3.012	2.998	3.131	3.047	3.835	3.967
Min.	-0.005	-0.004	-0.02	-0.028	-0.037	-0.042	-0.050	-0.050	-0.107
Max.	0.018	0.008	0.027	0.032	0.038	0.052	0.054	0.054	0.113
				$Panel\ 4: PAR$					
Mean	0.003	0.002	0.076	0.000	0.000	0.000	0.000	0.015	0.004
Std. Dev.	0.003	0.002	0.585	0.007	0.007	0.007	0.007	0.263	0.132
Skewness	0.802	-0.038	7.649	0.062	0.057	0.035	0.077	17.437	34.892
Kurtosis	4.162	3.155	59.777	2.865	2.755	2.764	2.777	306.5	1226.815
Min.	-0.005	-0.004	-0.023	-0.022	-0.023	-0.021	-0.022	-0.023	-0.023
Max.	0.018	0.008	5.177	0.022	0.022	0.023	0.022	5.177	5.177

(*) See notes to Table 5a.

Table 8c: Simulation Experiment Results XII [(*)]

Target Simulation Variable is $\ln(CPI_t) - \ln(CPI_{t-12})$

	Historical Data 1959:1 - 2004:12	Historical Data 1991:1 - 2004:12	First 60	61-120	121-180	181-240	241-300	First 300	First 1200
				$Panel\ 1: RW$					
Mean	0.042	0.025	0.039	0.041	0.041	0.041	0.041	0.041	0.041
Std. Dev.	0.028	0.006	0.012	0.011	0.011	0.011	0.012	0.011	0.011
Skewness	1.380	-0.383	0.117	0.079	-0.017	0.065	0.120	0.069	0.016
Kurtosis	4.442	2.249	2.892	2.927	2.884	2.996	2.929	2.928	2.983
Min.	0.007	0.011	0.001	0.002	0.003	0.008	0.001	0.001	-0.006
Max.	0.138	0.037	0.083	0.077	0.077	0.083	0.081	0.083	0.092
				$Panel\ 2: DS$					
Mean	0.042	0.025	0.171	0.273	0.279	0.282	0.282	0.258	0.278
Std. Dev.	0.028	0.006	0.078	0.028	0.030	0.029	0.030	0.061	0.041
Skewness	1.380	-0.383	-0.415	-0.151	0.033	0.044	0.080	-1.865	-2.245
Kurtosis	4.442	2.249	2.019	3.467	2.709	2.628	2.550	6.914	13.207
Min.	0.007	0.011	0.015	0.150	0.184	0.195	0.203	0.015	0.015
Max.	0.138	0.037	0.321	0.371	0.368	0.372	0.365	0.372	0.406
				$Panel\ 3: SUR$					
Mean	0.042	0.025	0.034	0.040	0.039	0.037	0.032	0.037	0.036
Std. Dev.	0.028	0.006	0.021	0.027	0.027	0.031	0.031	0.028	0.028
Skewness	1.380	-0.383	0.249	0.076	0.117	0.214	0.281	0.186	0.034
Kurtosis	4.442	2.249	3.577	3.221	2.610	2.669	3.250	3.137	2.991
Min.	0.007	0.011	-0.036	-0.048	-0.035	-0.051	-0.062	-0.062	-0.069
Max.	0.138	0.037	0.122	0.132	0.126	0.126	0.142	0.142	0.147
				$Panel\ 4: PAR$					
Mean	0.042	0.025	0.933	0.001	-0.002	-0.001	0.001	0.187	0.047
Std. Dev.	0.028	0.006	1.887	0.026	0.025	0.026	0.026	0.923	0.469
Skewness	1.380	-0.383	1.535	0.029	0.052	0.120	-0.201	4.749	9.92
Kurtosis	4.442	2.249	3.382	3.597	2.928	2.827	3.243	23.67	100.064
Min.	0.007	0.011	-0.085	-0.100	-0.082	-0.08	-0.110	-0.110	-0.110
Max.	0.138	0.037	5.256	0.098	0.083	0.076	0.078	5.256	5.256

[(*)] See notes to Table 5a.

Table 9a: Distributional Accuracy Tests Comparing Historical and Simulated $\ln(IP_t)$ Data

Benchmark Model is RW $^{(*)}$

S,l	Z	$Crit.Val.(Z^*)$ 10%	$Crit.Val.(Z^*)$ 5%	$Crit.Val.(Z^{**})$ 10%	$Crit.Val.(Z^{**})$ 5%	CS Distributional Loss RW	CS Distributional Loss DS	CS Distributional Loss SUR	CS Distributional Loss PAR
Panel 1 : Comparing Log Levels									
5T,2	0.022	5.045	5.642	5.044	5.59	5.466	5.474	5.481	5.444
5T,4	0.022	4.732	5.655	4.813	5.677	5.466	5.474	5.481	5.444
5T,6	0.022	5.045	4.824	4.138	4.782	5.466	5.474	5.481	5.444
5T,12	0.022	4.432	5.867	4.527	5.901	5.466	5.474	5.481	5.444
10T,2	-0.133	4.741	5.488	4.775	5.412	5.344	5.478	5.480	5.478
10T,4	-0.133	4.706	5.196	4.813	5.358	5.344	5.478	5.480	5.478
10T,6	-0.133	4.741	5.354	4.761	5.352	5.344	5.478	5.480	5.478
10T,12	-0.133	5.565	6.501	5.672	6.377	5.344	5.478	5.480	5.478
20T,2	-0.070	5.019	5.325	5.046	5.273	5.405	5.478	5.480	5.475
20T,4	-0.070	4.953	5.385	4.973	5.254	5.405	5.478	5.480	5.475
20T,6	-0.070	5.019	5.532	5.022	5.543	5.405	5.478	5.480	5.475
20T,12	-0.070	5.544	6.276	5.625	6.366	5.405	5.478	5.480	5.475
Panel 2 : Comparing First Log Differences									
5T,2	0.013	0.121	0.127	0.131	0.149	1.451	1.438	1.815	1.461
5T,4	0.013	0.198	0.21	0.196	0.224	1.451	1.438	1.815	1.461
5T,6	0.013	0.121	0.276	0.261	0.277	1.451	1.438	1.815	1.461
5T,12	0.013	0.252	0.303	0.268	0.288	1.451	1.438	1.815	1.461
10T,2	0.011	0.429	0.442	0.439	0.452	1.449	1.438	2.151	1.463
10T,4	0.011	0.210	0.235	0.215	0.227	1.449	1.438	2.151	1.463
10T,6	0.011	0.429	0.274	0.237	0.263	1.449	1.438	2.151	1.463
10T,12	0.011	0.257	0.296	0.257	0.288	1.449	1.438	2.151	1.463
20T,2	0.014	0.440	0.455	0.443	0.455	1.451	1.438	2.169	1.462
20T,4	0.014	0.196	0.215	0.200	0.222	1.451	1.438	2.169	1.462
20T,6	0.014	0.440	0.226	0.201	0.229	1.451	1.438	2.169	1.462
20T,12	0.014	0.244	0.272	0.249	0.273	1.451	1.438	2.169	1.462
Panel 3: Comparing Twelvth Log Differences									
5T,2	0.341	0.825	0.849	0.848	0.895	1.754	1.412	1.447	2.386
5T,4	0.341	0.795	0.852	0.802	0.882	1.754	1.412	1.447	2.386
5T,6	0.341	0.825	0.94	0.844	0.909	1.754	1.412	1.447	2.386
5T,12	0.341	0.786	0.833	0.811	0.894	1.754	1.412	1.447	2.386
10T,2	0.290	0.957	0.984	0.973	0.989	1.698	1.494	1.408	2.507
10T,4	0.290	0.934	0.970	0.947	0.986	1.698	1.494	1.408	2.507
10T,6	0.290	0.957	0.996	0.976	1.028	1.698	1.494	1.408	2.507
10T,12	0.290	0.904	0.938	0.924	0.948	1.698	1.494	1.408	2.507
20T,2	0.290	1.024	1.073	1.025	1.063	1.686	1.512	1.396	2.583
20T,4	0.290	1.031	1.065	1.029	1.053	1.686	1.512	1.396	2.583
20T,6	0.290	1.024	1.032	1.002	1.077	1.686	1.512	1.396	2.583
20T,12	0.290	0.975	1.011	1.001	1.034	1.686	1.512	1.396	2.583

$^{(*)}$ $Z_{T,N}$ test statistics (called Z in the table), and associated distributional loss measures (denoted "CS Distributional Loss") are reported, where "N" denotes the length of the simulated data series used in test statistic construction, and "T" is the historical sample length, assumed to be the period 1991:1-2004:12 in this table (note that this historical period is also the period used to estimate the models). These statistics are test statistics for selecting amongst alternative simulation models via comparison of simulated and historical distribtutions associated with the different models. Critical values that are constructed both assuming that T/S approaches $\gamma > 0$ (Z^{**}) and assuming that T/S approaches 0 (Z^*), as T and S increase are reported in the 3th through 6th columns of entries (see Section 2 for complete details). As usual, the models are denoted by RW, DS, SUR, and PAR. The first panel in the table reports results when the "target simulation variable" is (log) levels; the second compares simulated with historical first log differences, and the third panel reports results for twelveth log differences. the benchmark model is RW, against which all other model are compared. The null hypothesis corresponds to the case where no alternative model outperforms the benchmark. Finally, l is the bootstrap block length, and all statistics are based on a grid of 20x20 values for u, distributed uniformly across the historical data range. Bootstrap empirical distributions are constructed using 100 bootstrap replications.

Table 9b: Distributional Accuracy Tests Comparing Historical and Simulated $\ln(M1_t)$ Data

Benchmark Model is RW $^{(*)}$

S,l	Z	$Crit.Val.(Z^*)$ 10%	$Crit.Val.(Z^*)$ 5%	$Crit.Val.(Z^{**})$ 10%	$Crit.Val.(Z^{**})$ 5%	CS Distributional Loss RW	DS	SUR	PAR
				Panel 1 : Comparing Log Levels					
5T,2	0.609	5.003	5.325	5.058	5.301	5.665	5.664	5.664	5.056
5T,4	0.609	4.888	5.572	4.919	5.558	5.665	5.664	5.664	5.056
5T,6	0.609	5.003	5.612	4.969	5.615	5.665	5.664	5.664	5.056
5T,12	0.609	5.372	5.75	5.474	5.777	5.665	5.664	5.664	5.056
10T,2	0.953	4.225	4.884	4.442	4.906	5.609	5.664	5.659	4.656
10T,4	0.953	4.332	4.940	4.301	4.992	5.609	5.664	5.659	4.656
10T,6	0.953	4.225	4.896	4.098	4.935	5.609	5.664	5.659	4.656
10T,12	0.953	4.704	5.454	4.709	5.284	5.609	5.664	5.659	4.656
20T,2	0.863	4.698	5.033	4.804	5.071	5.665	5.659	5.656	4.802
20T,4	0.863	4.303	5.096	4.118	5.208	5.665	5.659	5.656	4.802
20T,6	0.863	4.698	4.313	4.105	4.437	5.665	5.659	5.656	4.802
20T,12	0.863	4.463	5.160	4.395	5.035	5.665	5.659	5.656	4.802
				Panel 2 : Comparing First Log Differences					
5T,2	-0.034	0.156	0.170	0.163	0.185	1.307	1.341	1.728	1.343
5T,4	-0.034	0.215	0.247	0.237	0.251	1.307	1.341	1.728	1.343
5T,6	-0.034	0.156	0.262	0.242	0.265	1.307	1.341	1.728	1.343
5T,12	-0.034	0.287	0.339	0.305	0.342	1.307	1.341	1.728	1.343
10T,2	-0.036	0.611	0.623	0.612	0.623	1.311	1.349	2.228	1.347
10T,4	-0.036	0.354	0.377	0.350	0.397	1.311	1.349	2.228	1.347
10T,6	-0.036	0.611	0.292	0.276	0.291	1.311	1.349	2.228	1.347
10T,12	-0.036	0.279	0.309	0.293	0.322	1.311	1.349	2.228	1.347
20T,2	-0.035	0.484	0.495	0.486	0.495	1.308	1.343	2.106	1.346
20T,4	-0.035	0.235	0.255	0.233	0.258	1.308	1.343	2.106	1.346
20T,6	-0.035	0.484	0.307	0.278	0.303	1.308	1.343	2.106	1.346
20T,12	-0.035	0.256	0.294	0.257	0.283	1.308	1.343	2.106	1.346
				Panel 3 : Comparing Twelvth Log Differences					
5T,2	-0.097	1.026	1.108	1.043	1.108	1.779	2.219	1.876	2.334
5T,4	-0.097	0.890	0.935	0.941	1.043	1.779	2.219	1.876	2.334
5T,6	-0.097	1.026	0.832	0.807	0.912	1.779	2.219	1.876	2.334
5T,12	-0.097	0.689	0.732	0.758	0.809	1.779	2.219	1.876	2.334
10T,2	-0.260	0.938	0.987	0.961	0.987	1.796	2.450	2.056	2.311
10T,4	-0.260	0.810	0.850	0.796	0.849	1.796	2.450	2.056	2.311
10T,6	-0.260	0.938	0.841	0.835	0.860	1.796	2.450	2.056	2.311
10T,12	-0.260	0.721	0.743	0.749	0.794	1.796	2.450	2.056	2.311
20T,2	-0.088	1.003	1.032	0.990	1.028	1.760	2.527	1.847	2.299
20T,4	-0.088	0.860	0.896	0.897	0.916	1.760	2.527	1.847	2.299
20T,6	-0.088	1.003	0.865	0.838	0.896	1.760	2.527	1.847	2.299
20T,12	-0.088	0.84	0.873	0.846	0.869	1.760	2.527	1.847	2.299

$^{(*)}$ See notes to Table 9a

Table 9c: Distributional Accuracy Tests Comparing Historical and Simulated $R1_t$ Data

Benchmark Model is RW $^{(*)}$

S,l	Z	$Crit.Val.(Z^*)$		$Crit.Val.(Z^{**})$		CS $Distributional$ $Loss$			
		10%	5%	10%	5%	RW	DS	SUR	PAR
				Panel 1 : Comparing Levels					
5T,2	2.948	2.878	3.264	2.884	3.258	6.331	6.539	4.734	3.383
5T,4	2.948	3.210	3.656	3.343	3.730	6.331	6.539	4.734	3.383
5T,6	2.948	2.878	3.878	3.516	4.077	6.331	6.539	4.734	3.383
5T,12	2.948	3.473	3.935	3.531	4.183	6.331	6.539	4.734	3.383
10T,2	4.691	2.178	2.915	2.179	2.951	6.655	5.806	4.516	1.963
10T,4	4.691	2.138	2.398	2.139	2.461	6.655	5.806	4.516	1.963
10T,6	4.691	2.178	2.437	1.827	2.471	6.655	5.806	4.516	1.963
10T,12	4.691	1.949	2.466	2.012	2.425	6.655	5.806	4.516	1.963
20T,2	4.351	2.815	3.137	2.815	3.120	6.688	6.807	4.940	2.337
20T,4	4.351	2.621	3.816	2.669	3.836	6.688	6.807	4.940	2.337
20T,6	4.351	2.815	3.201	2.588	3.191	6.688	6.807	4.940	2.337
20T,12	4.351	2.494	3.173	2.548	3.123	6.688	6.807	4.940	2.337
				Panel 2 : Comparing First Log Differences					
5T,2	0.005	0.984	1.010	0.998	1.018	1.054	1.068	2.355	1.049
5T,4	0.005	0.738	0.759	0.742	0.764	1.054	1.068	2.355	1.049
5T,6	0.005	0.984	0.591	0.567	0.599	1.054	1.068	2.355	1.049
5T,12	0.005	0.340	0.366	0.352	0.407	1.054	1.068	2.355	1.049
10T,2	0.001	1.043	1.059	1.042	1.066	1.052	1.059	2.442	1.051
10T,4	0.001	0.765	0.778	0.787	0.825	1.052	1.059	2.442	1.051
10T,6	0.001	1.043	0.624	0.604	0.628	1.052	1.059	2.442	1.051
10T,12	0.001	0.325	0.366	0.341	0.373	1.052	1.059	2.442	1.051
20T,2	0.005	1.245	1.262	1.238	1.255	1.055	1.082	2.664	1.050
20T,4	0.005	0.930	0.966	0.929	0.960	1.055	1.082	2.664	1.050
20T,6	0.005	1.245	0.809	0.751	0.808	1.055	1.082	2.664	1.050
20T,12	0.005	0.457	0.488	0.441	0.495	1.055	1.082	2.664	1.050
				Panel 3 : Comparing Twelvth Log Differences					
5T,2	0.046	0.485	0.516	0.467	0.515	1.308	1.383	1.261	1.370
5T,4	0.046	0.315	0.358	0.327	0.385	1.308	1.383	1.261	1.370
5T,6	0.046	0.485	0.305	0.272	0.300	1.308	1.383	1.261	1.370
5T,12	0.046	0.120	0.130	0.149	0.170	1.308	1.383	1.261	1.370
10T,2	0.040	0.438	0.498	0.455	0.504	1.304	1.264	1.272	1.364
10T,4	0.040	0.28	0.317	0.271	0.323	1.304	1.264	1.272	1.364
10T,6	0.040	0.438	0.261	0.219	0.253	1.304	1.264	1.272	1.364
10T,12	0.040	0.084	0.103	0.082	0.098	1.304	1.264	1.272	1.364
20T,2	0.053	0.425	0.465	0.431	0.464	1.310	1.272	1.258	1.365
20T,4	0.053	0.266	0.287	0.269	0.312	1.310	1.272	1.258	1.365
20T,6	0.053	0.425	0.233	0.198	0.235	1.310	1.272	1.258	1.365
20T,12	0.053	0.066	0.085	0.071	0.091	1.310	1.272	1.258	1.365

$^{(*)}$ See notes to Table 9a

Table 9d: Distributional Accuracy Tests Comparing Historical and Simulated $\ln(CPI1_t)$ Data

Benchmark Model is RW $^{(*)}$

S,l	Z	$Crit.Val.(Z^*)$		$Crit.Val.(Z^{**})$		$CS\ Distributional\ Loss$			
		10%	5%	10%	5%	RW	DS	SUR	PAR
				$Panel\ 1: Comparing\ Log\ Levels$					
5T,2	-0.002	4.562	4.673	4.599	4.669	6.247	6.249	6.249	6.249
5T,4	-0.002	4.869	5.081	4.881	5.080	6.247	6.249	6.249	6.249
5T,6	-0.002	4.562	5.270	4.737	5.260	6.247	6.249	6.249	6.249
5T,12	-0.002	5.051	5.437	5.039	5.464	6.247	6.249	6.249	6.249
10T,2	-0.001	4.362	4.481	4.308	4.491	6.247	6.249	6.249	6.249
10T,4	-0.001	4.392	4.615	4.392	4.600	6.247	6.249	6.249	6.249
10T,6	-0.001	4.362	5.000	4.364	5.000	6.247	6.249	6.249	6.249
10T,12	-0.001	4.550	4.928	4.727	5.014	6.247	6.249	6.249	6.249
20T,2	0.000	4.426	4.779	4.485	4.743	6.248	6.249	6.248	6.248
20T,4	0.000	4.188	4.417	4.145	4.326	6.248	6.249	6.248	6.248
20T,6	0.000	4.426	4.675	4.278	4.675	6.248	6.249	6.248	6.248
20T,12	0.000	4.670	5.199	4.722	5.369	6.248	6.249	6.248	6.248
				$Panel\ 2: Comparing\ First\ Log\ Differences$					
5T,2	-0.002	1.153	1.16	1.159	1.17	1.239	1.241	2.428	1.795
5T,4	-0.002	1.091	1.108	1.103	1.116	1.239	1.241	2.428	1.795
5T,6	-0.002	1.153	1.079	1.072	1.101	1.239	1.241	2.428	1.795
5T,12	-0.002	0.974	0.988	0.982	0.998	1.239	1.241	2.428	1.795
10T,2	-0.001	1.138	1.142	1.141	1.150	1.239	1.240	2.428	1.986
10T,4	-0.001	1.088	1.102	1.095	1.104	1.239	1.240	2.428	1.986
10T,6	-0.001	1.138	1.051	1.056	1.068	1.239	1.240	2.428	1.986
10T,12	-0.001	0.963	0.969	0.962	0.996	1.239	1.240	2.428	1.986
20T,2	-0.003	1.231	1.237	1.233	1.237	1.239	1.241	2.524	2.214
20T,4	-0.003	1.162	1.167	1.169	1.173	1.239	1.241	2.524	2.214
20T,6	-0.003	1.231	1.148	1.137	1.142	1.239	1.241	2.524	2.214
20T,12	-0.003	1.161	1.201	1.168	1.195	1.239	1.241	2.524	2.214
				$Panel\ 3: Comparing\ Twelvth\ Log\ Differences$					
5T,2	0.026	1.693	1.755	1.696	1.731	1.645	1.619	1.628	3.155
5T,4	0.026	1.681	1.742	1.724	1.761	1.645	1.619	1.628	3.155
5T,6	0.026	1.693	1.800	1.699	1.900	1.645	1.619	1.628	3.155
5T,12	0.026	1.649	1.704	1.664	1.723	1.645	1.619	1.628	3.155
10T,2	0.003	3.164	3.201	3.159	3.210	1.618	1.630	1.615	4.645
10T,4	0.003	3.137	3.173	3.150	3.183	1.618	1.630	1.615	4.645
10T,6	0.003	3.164	3.166	3.151	3.170	1.618	1.630	1.615	4.645
10T,12	0.003	3.170	3.205	3.136	3.259	1.618	1.630	1.615	4.645
20T,2	-0.005	4.022	4.053	4.023	4.062	1.627	1.633	1.639	5.545
20T,4	-0.005	4.029	4.057	4.031	4.111	1.627	1.633	1.639	5.545
20T,6	-0.005	4.022	4.048	4.010	4.069	1.627	1.633	1.639	5.545
20T,12	-0.005	4.002	4.051	4.042	4.084	1.627	1.633	1.639	5.545

$^{(*)}$ See notes to Table 9a

Table 10a: Mean Square Forecast Errors for Levels Variables and Various Prediction Horizons(*)

	$\ln IP$	$\ln M1$	$\ln M2$	$\ln M3$	$R1$	$R2$	$R3$	$\ln CPI1$	$\ln CPI2$	$\ln CPI3$	$\ln CPI4$	$\ln CPI5$	$\ln CPI6$
Panel 1: 1-Step Ahead Predictions													
RW	0.4431	0.2352	0.0737	0.0749	0.0565	0.0604	0.0519	0.0126	0.0057	0.0159	0.0067	1.0260	0.0102
RW-D	0.4537	0.2175	0.0463	0.0409	0.0551	0.0600	0.0495	0.0095	0.0031	0.0131	0.0045	1.0234	0.0053
DS	0.0619	0.1983	0.0374	0.0419	0.0406	0.0568	0.0307	0.0056	0.0010	0.0074	0.0016	0.8516	0.0068
SUR	0.0921	0.2352	0.0392	0.0421	0.0889	0.1283	0.0615	0.0107	0.0011	0.0142	0.0017	1.6917	0.0091
PAR	0.0651	0.1685	0.0446	0.0560	0.0989	0.0957	0.0710	0.0065	0.0010	0.0084	0.0014	1.3238	0.0067
Panel 2: 3-Step Ahead Predictions													
RW	0.5841	0.3865	0.3721	0.4685	0.3272	0.2124	0.3698	0.0671	0.0366	0.0777	0.0380	4.1053	0.0574
RW-D	0.6867	0.2386	0.1278	0.1644	0.3156	0.2083	0.3509	0.0366	0.0115	0.0500	0.0168	4.0900	0.0149
DS	0.2604	0.4547	0.1203	0.2672	0.2838	0.2147	0.2515	0.0227	0.0052	0.0302	0.0084	3.9809	0.0157
SUR	0.4120	0.5209	0.1319	0.1945	0.5623	0.4502	0.5180	0.0322	0.0043	0.0439	0.0055	5.9310	0.0261
PAR	0.2590	0.5596	0.1940	0.3388	0.4960	0.3822	0.4823	0.0201	0.0030	0.0253	0.0037	4.2139	0.0189
Panel 3: 12-Step Ahead Predictions													
RW	0.9750	2.8067	4.7167	6.1334	2.8141	0.6980	3.7192	0.5946	0.4738	0.5871	0.4427	12.5540	0.7459
RW-D	2.5826	0.9970	0.8840	1.0331	2.6427	0.5851	3.4898	0.0608	0.0459	0.0825	0.0699	11.9476	0.0873
DS	2.1137	2.0464	0.6353	5.2757	2.4911	0.6047	3.9290	0.0752	0.0517	0.1032	0.0977	12.0869	0.0835
SUR	2.6125	2.0107	0.5452	0.6599	2.7454	0.7070	5.3574	0.0860	0.0326	0.1296	0.0376	16.4207	0.1139
PAR	2.5730	3.9664	1.4114	3.0446	3.1131	0.8649	4.2022	0.0753	0.0240	0.0962	0.0280	20.2150	0.1039
Panel 4: 60-Step Ahead Predictions													
RW	12.9294	18.5684	117.5666	187.4196	10.6222	2.3102	12.0384	14.5594	12.7951	14.6137	12.5522	61.1214	14.1726
RW-D	19.8196	19.5758	38.6793	64.8670	5.9401	0.7436	6.3174	0.2474	0.8870	0.3835	1.4114	64.6100	0.1354
DS	11.2451	389.3332	13.2609	84.0864	3.2412	4.6232	48.0929	0.9171	0.3261	2.7777	2.7029	87.9523	0.2265
SUR	19.5055	9.8409	26.8993	20.0112	4.3034	1.0941	6.2671	0.3000	0.5987	0.4753	0.5496	61.7021	0.1642
PAR	63.4107	12.3900	675.4609	6040.2158	12.7388	2.7090	15.5128	0.9238	0.1159	1.2208	0.2178	137.8179	7.1617

(*) See notes to Table 3. Entries are mean square forecast errors (MSFEs) associated with predictions of the various variables given in the first row of the table. Prediction models are constructed recursively, starting with $R = 120$, and ending with $R = T - h$, where T is the sample size, and h is the forecast horizong (set equal to 1, 3, 12, and 60 months ahead. The sample period used in these experiments is 1991:1-2004:12. All predictions are thus *ex ante*. For all series other than the interest rates, reported values are 1000*MSFE.

Table 10b: Mean Square Forecast Errors for First Differences Variables and Various Prediction Horizons(*)

	$\ln IP$	$\ln M1$	$\ln M2$	$\ln M3$	$R1$	$R2$	$R3$	$\ln CPI1$	$\ln CPI2$	$\ln CPI3$	$\ln CPI4$	$\ln CPI5$	$\ln CPI6$
						Panel 1: 1-Step Ahead Predictions							
RW	1.3111	0.5572	0.0903	0.0609	0.0415	0.0900	0.0242	0.0121	0.0040	0.0170	0.0058	1.3995	0.0105
RW-D	1.3205	0.5613	0.0910	0.0613	0.0418	0.0907	0.0244	0.0122	0.0040	0.0171	0.0058	1.4095	0.0106
DS	0.0945	0.5881	0.0978	0.0931	0.0645	0.1152	0.0504	0.0087	0.0014	0.0119	0.0024	1.4507	0.0150
SUR	0.1571	0.6404	0.0969	0.0832	0.1504	0.2660	0.0985	0.0164	0.0018	0.0222	0.0034	2.9825	0.0165
PAR	0.1167	0.3292	0.0712	0.0557	0.0905	0.1446	0.0426	0.0110	0.0022	0.0148	0.0033	2.1326	0.0135
						Panel 2: 3-Step Ahead Predictions							
RW	0.8031	0.2759	0.0668	0.0574	0.0734	0.1257	0.0473	0.0242	0.0079	0.0327	0.0115	2.2326	0.0093
RW-D	0.8206	0.2818	0.0684	0.0587	0.0751	0.1284	0.0483	0.0247	0.0081	0.0334	0.0117	2.2803	0.0095
DS	0.0821	0.7302	0.1263	0.1588	0.1456	0.1672	0.1791	0.0150	0.0012	0.0198	0.0018	3.0639	0.0118
SUR	0.1168	0.6166	0.0961	0.1269	0.3380	0.3691	0.2355	0.0193	0.0020	0.0242	0.0028	4.0176	0.0109
PAR	0.0852	0.2071	0.0690	0.0812	0.1324	0.2004	0.0609	0.0133	0.0017	0.0163	0.0023	2.4223	0.0109
						Panel 3: 12-Step Ahead Predictions							
RW	0.0891	0.1956	0.0377	0.0471	0.1144	0.1363	0.1203	0.0113	0.0011	0.0149	0.0017	1.9117	0.0087
RW-D	0.1047	0.2174	0.0424	0.0523	0.1255	0.1496	0.1324	0.0124	0.0012	0.0164	0.0018	2.1002	0.0095
DS	0.0874	1.4245	0.2967	0.4242	0.2339	0.2369	0.7071	0.0158	0.0011	0.0197	0.0017	3.2767	0.0096
SUR	0.2335	0.8519	0.1478	0.1757	0.2850	0.2102	0.4560	0.0222	0.0024	0.0314	0.0036	4.0528	0.0127
PAR	0.1071	0.1510	0.0718	0.0862	0.1360	0.1698	0.1464	0.0164	0.0014	0.0216	0.0022	4.2372	0.0121
						Panel 4: 60-Step Ahead Predictions							
RW	0.1006	0.1472	0.0343	0.0506	0.0750	0.0902	0.0570	0.0088	0.0010	0.0113	0.0016	1.3252	0.0103
RW-D	0.4558	0.3687	0.0668	0.0983	0.1457	0.1802	0.0827	0.0116	0.0030	0.0142	0.0038	1.8541	0.0230
DS	0.1427	4.0174	0.6127	3.4489	0.9262	0.9650	2.8204	0.0257	0.0027	0.0613	0.0074	5.2848	0.0239
SUR	0.2788	3.2887	0.3665	1.1902	0.7922	0.2967	1.0579	0.0321	0.0070	0.0562	0.0273	2.0512	0.0309
PAR	3.2522	0.1437	132.2183	756.4395	0.1021	0.0788	0.1834	0.0556	0.0118	0.0523	0.0137	87.4698	0.6326

(*) See notes to Table 10a.

Table 10c: Mean Square Forecast Errors for Twelvth Differences Variables and Various Prediction Horizons(*)

	$\ln IP$	$\ln M1$	$\ln M2$	$\ln M3$	$R1$	$R2$	$R3$	$\ln CPI1$	$\ln CPI2$	$\ln CPI3$	$\ln CPI4$	$\ln CPI5$	$\ln CPI6$
						Panel 1: 1-Step Ahead Predictions							
RW	0.0895	0.1399	0.0391	0.0483	0.0895	0.1592	0.0664	0.0108	0.0012	0.0141	0.0020	1.7635	0.0105
RW-D	0.0920	0.1412	0.0400	0.0496	0.0925	0.1605	0.0701	0.0109	0.0012	0.0142	0.0020	1.7760	0.0106
DS	0.0986	0.2272	0.0494	0.0514	0.0835	0.1625	0.0535	0.0112	0.0013	0.0146	0.0022	1.8409	0.0113
SUR	0.292	0.7128	0.1132	0.1236	0.2713	0.4150	0.1802	0.0315	0.0030	0.0432	0.0044	5.0989	0.0276
PAR	0.1100	0.2020	0.0675	0.0780	0.1510	0.2457	0.0919	0.0157	0.0017	0.0207	0.0027	2.7520	0.0144
						Panel 2: 3-Step Ahead Predictions							
RW	0.3267	0.3307	0.1504	0.2183	0.4327	0.5874	0.4475	0.0399	0.0050	0.0525	0.0067	6.8383	0.0313
RW-D	0.3499	0.3411	0.1590	0.2304	0.4612	0.6010	0.4814	0.0407	0.0050	0.0537	0.0067	6.9707	0.0320
DS	0.3752	0.8618	0.1889	0.2440	0.5181	0.6802	0.4382	0.0438	0.0052	0.0567	0.0064	8.2541	0.0341
SUR	1.3232	1.6419	0.3764	0.5473	1.4804	1.4138	1.3706	0.1018	0.0109	0.1448	0.0135	18.4279	0.0800
PAR	0.4436	0.4057	0.2485	0.3418	0.5931	1.003	0.6351	0.0566	0.0068	0.0751	0.0087	10.0216	0.0441
						Panel 3: 12-Step Ahead Predictions							
RW	1.8649	2.1139	0.5286	0.8415	2.2546	1.4238	5.0882	0.1573	0.0377	0.2216	0.0417	33.1578	0.2176
RW-D	2.2076	2.3589	0.6221	1.0036	2.6131	1.5619	5.7435	0.1726	0.0395	0.2430	0.0422	36.3592	0.2360
DS	2.0516	4.6794	0.9340	1.4771	2.8849	1.7125	8.4573	0.1475	0.0326	0.2002	0.0304	37.1246	0.2402
SUR	6.6580	7.8451	1.3648	1.9668	4.4073	2.0242	14.4416	0.2763	0.0731	0.4376	0.0786	55.1990	0.3242
PAR	2.7354	1.7437	0.8751	1.2833	2.8916	1.5638	6.7677	0.2340	0.0504	0.3280	0.0521	61.9906	0.3031
						Panel 4: 60-Step Ahead Predictions							
RW	3.3782	4.1147	0.7474	1.1833	1.7578	0.8190	2.9628	0.0800	0.0216	0.0949	0.0334	22.8868	0.1028
RW-D	5.3828	11.3733	1.5902	4.2176	2.5256	1.8654	3.7001	0.1600	0.0164	0.1881	0.0272	35.6651	0.1701
DS	5.0067	62.9991	14.5818	46.3097	5.2120	4.2442	29.6034	0.2016	0.0616	0.3853	0.0294	52.6605	0.3645
SUR	6.9389	21.1590	6.3385	14.0659	5.9484	2.3612	13.7812	0.5168	0.0835	0.9113	0.2306	40.0898	0.2805
PAR	28.0142	4.2213	363.9968	3620.4612	1.4944	0.8165	2.3548	1.0753	0.1293	0.8302	0.0808	299.729	6.2199

(*) See notes to Table 10a.

Table 11a: DM Predictive Accuracy Test Results for Levels Variables and Various Prediction Horizons $^{(*)}$

Benchmark Model: RW

	$\ln IP$	$\ln M1$	$\ln M2$	$\ln M3$	$R1$	$R2$	$R3$	$\ln CPI1$	$\ln CPI2$	$\ln CPI3$	$\ln CPI4$	$\ln CPI5$	$\ln CPI6$
Panel 1: 1-Step Ahead Predictions													
RW-D	-1.08	3.09*	2.99*	2.13*	0.26	-0.15	0.49	1.27	1.99*	0.93	1.32	0.21	1.69*
DS	2.03*	0.38	1.87*	1.21	1.52	0.40	0.83	1.78*	2.31*	1.82*	2.21*	1.41	1.59
SUR	1.93*	-0.14	1.32	1.12	-2.16*	-2.67*	-0.48	0.74	2.33*	0.40	2.35*	-1.83*	0.12
PAR	2.06*	1.15	0.92	0.24	-2.16*	-2.07*	-2.66*	1.60	2.35*	1.64	2.29*	-0.75	1.66*
Panel 2: 3-Step Ahead Predictions													
RW-D	-0.68	2.14*	1.65*	1.85*	0.32	-0.10	0.55	1.95*	1.76*	1.58	1.58	0.08	1.45
DS	1.74*	-0.06	1.66*	0.71	0.47	-0.13	0.62	2.05*	2.14*	1.84*	2.55*	0.27	1.55
SUR	1.08	-0.63	1.46	1.45	-1.73*	-1.39	-1.40	1.86*	2.24*	1.43	2.15*	-1.3	1.28
PAR	2.28*	-2.35*	0.91	0.28	-1.63	-0.80	-1.49	2.14*	2.18*	1.98*	2.41*	0.00	1.43
Panel 3: 12-Step Ahead Predictions													
RW-D	-0.56	0.89	0.73	0.97	0.10	0.47	0.19	0.72	0.99	0.88	0.92	0.22	0.69
DS	-0.41	1.36	0.84	0.22	0.30	0.19	-0.10	0.65	1.06	0.85	0.95	0.02	0.71
SUR	-0.80	1.34	0.94	1.01	0.24	-0.02	-1.00	0.75	1.09	0.90	1.14	-0.89	0.69
PAR	-0.65	-0.33	0.82	0.66	-0.54	-1.73*	-0.87	0.74	1.01	0.91	0.97	-1.02	0.70
Panel 4: 60-Step Ahead Predictions													
RW-D	-0.53	0.68	1.62	0.31	0.56	1.11	0.74	0.15	0.82	0.22	0.72	-0.44	0.28
DS	-0.25	-8.07*	0.93	0.38	0.60	-0.60	-2.28*	0.26	1.68*	0.26	0.70	-1.12	0.11
SUR	-0.45	0.61	0.53	0.82	0.72	0.94	1.47	0.23	0.69	0.31	0.55	-0.08	0.14
PAR	-0.64	0.57	-6.4*	-9.78*	-1.16	-0.64	-0.69	0.42	1.69*	0.54	2.2*	-1.30	0.67

$^{(*)}$ See notes to Tables 10a. Diebold and Mariano (DM) test statistics based on MSFE loss are reported. Rejections based upon an assumption that a standard normal limiting distribution is valid, and based upon 10% nominal significance level critical values, are denoted as starred entries. Negative values indicate cases where the MSFE of the RW model is lower than that of the model against which it is being compared. See above for further details.

Table 11b: DM Predictive Accuracy Test Results for First Differecnes Variables and Various Prediction Horizons [(*)]

Benchmark Model: RW

	$\ln IP$	$\ln M1$	$\ln M2$	$\ln M3$	$R1$	$R2$	$R3$	$\ln CPI1$	$\ln CPI2$	$\ln CPI3$	$\ln CPI4$	$\ln CPI5$	$\ln CPI6$
Panel 1: 1-Step Ahead Predictions													
RW-D	-1.76*	-2.22*	-2.06*	-2.35*	-2.31*	-1.79*	-1.84*	-3.15*	-2.06*	-3.13*	-1.91*	-1.79*	-2.62*
DS	2.25*	-0.24	-0.39	-1.63	-1.87*	-1.82*	-1.94*	1.96*	2.57*	2.31*	2.60*	-0.33	-1.16
SUR	2.17*	-0.49	-0.24	-0.91	-1.69*	-2.16*	-2.48*	-0.88	2.04*	-0.87	1.74*	-2.22*	-1.55
PAR	2.23*	1.40	0.45	0.43	-2.63*	-1.94*	-1.47	0.27	1.60	0.51	1.86*	-0.91	-0.47
Panel 2: 3-Step Ahead Predictions													
RW-D	-1.65*	-1.92*	-2.06*	-2.35*	-1.93*	-2.39*	-1.8*	-1.71*	-1.83*	-1.7*	-1.71*	-2.14*	-2.49*
DS	1.08	-1.13	-1.11	-2.29*	-1.84*	-1.52	-2.04*	2.23*	1.52	2.26*	1.44	-1.45	-1.99*
SUR	1.11	-1.05	-0.84	-2.29*	-2.10*	-1.66*	-1.82*	0.86	2.82*	1.24	1.52	-2.59*	-1.96*
PAR	1.08	1.23	-0.11	-1.05	-2.55*	-1.05	-1.24	1.83*	1.54	1.94*	1.40	-0.34	-1.38
Panel 3: 12-Step Ahead Predictions													
RW-D	-2.65*	-1.33	-1.92*	-2.29*	-1.8*	-2.2*	-1.67*	-2.43*	-2.49*	-2.42*	-2.76*	-1.66*	-2.94*
DS	1.28	-0.99	-1.44	-1.98*	-1.8*	-3.06*	-2.02*	-2.06*	0.85	-1.90*	0.20	-2.62*	-2.43*
SUR	-2.65*	-1.46	-1.67*	-2.79*	-2.34*	-2.70*	-2.21*	-2.67*	-2.26*	-3.69*	-1.69*	-3.21*	-3.03*
PAR	-1.28	1.39	-1.87*	-2.69*	-1.18	-0.71	-1.72*	-1.97*	-2.44*	-1.81*	-2.58*	-1.48	-2.2*
Panel 4: 60-Step Ahead Predictions													
RW-D	-1.75*	-3.42*	-2.09*	-1.39	-1.61	-2.42*	-1.83*	-1.54	-2.52*	-1.29	-1.30	-2.26*	-2.00*
DS	-2.41*	-2.63*	-2.86*	-1.84*	-1.77*	-1.31	-2.06*	-1.11	-1.9*	-1.14	-2.07*	-1.53	-1.44
SUR	-1.99*	-2.26*	-2.40*	-2.53*	-2.28*	-2.50*	-2.28*	-1.39	-1.25	-2.12*	-3.62*	-2.02*	-1.35
PAR	-1.29	0.34	-9.72*	-4.14*	-0.54	0.93	-1.51	-2.30*	-2.12*	-1.92*	-1.89*	-1.02	-1.76*

[(*)] See notes to Table 11a.

Table 11c: DM Predictive Accuracy Test Results for Twelvth Difference Variables and Various Prediction Horizons [(*)]

Benchmark Model: RW

	$\ln IP$	$\ln M1$	$\ln M2$	$\ln M3$	$R1$	$R2$	$R3$	$\ln CPI1$	$\ln CPI2$	$\ln CPI3$	$\ln CPI4$	$\ln CPI5$	$\ln CPI6$
Panel 1: 1-Step Ahead Predictions													
RW-D	-1.08	-1.62	-1.27	-1.39	-0.37	-1.16	-0.93	-1.34	-0.13	-1.61	-0.16	-1.15	-1.19
DS	-1.28	-0.95	-0.93	-0.41	0.48	0.04	1.53	-0.96	-1.09	-1.01	-1.37	-0.96	-2.68*
SUR	-2.38*	-1.21	-1.56	-2.62*	-2.72*	-2.44*	-2.54*	-1.92*	-2.42*	-3.3*	-1.69*	-3.13*	-3.30*
PAR	-2.36*	-1.25	-2.07*	-2.79*	-1.71*	-1.41	-1.35	-1.8*	-2.25*	-1.53	-2.54*	-1.94*	-2.64*
Panel 2: 3-Step Ahead Predictions													
RW-D	-2.39*	-1.45	-1.83*	-1.92*	-1.67*	-1.57	-1.46	-1.64	-0.14	-1.60	-0.11	-2.40*	-0.98
DS	-1.18	-1.02	-1.00	-0.28	-0.62	-1.63	1.99*	-2.20*	-0.40	-2.01*	0.56	-2.55*	-1.67*
SUR	-1.49	-2.76*	-1.37	-1.71*	-1.55	-1.46	-1.81*	-1.48	-1.54	-1.42	-1.38	-1.74*	-0.85
PAR	-1.41	-1.21	-1.30	-1.26	-1.51	-1.45	-1.19	-1.62	-1.51	-1.59	-1.37	-1.66*	-1.10
Panel 3: 12-Step Ahead Predictions													
RW-D	-0.79	-1.70*	-1.05	-0.86	-1.05	-1.64	-1.21	-1.29	-0.13	-1.11	-0.03	-1.07	-1.24
DS	-0.98	-1.61	-1.91*	-1.34	-1.28	-2.46*	-1.33	1.41	0.64	1.34	0.67	-1.59	-0.77
SUR	-0.82	-2.05*	-1.06	-1.83*	-1.10	-1.51	-1.10	-1.25	-0.70	-1.27	-0.04	-1.20	-1.12
PAR	-0.75	0.85	-0.82	-1.20	-1.16	-0.94	-0.96	-1.29	-0.51	-1.23	-0.04	-1.40	-1.18
Panel 3: 60-Step Ahead Predictions													
RW-D	-1.41	-2.23*	-0.92	-0.78	-1.67*	-0.21	-1.27	-0.90	0.13	-1.14	0.29	-0.96	-1.48
DS	-0.92	-2.32*	-0.88	-0.78	-1.97*	-0.74	-2.46*	-0.41	-1.17	-0.35	0.31	-1.13	-1.63
SUR	-1.33	-2.28*	-0.85	-1.32	-2.6*	-0.03	-1.73*	-0.94	-0.15	-1.08	-1.74*	-0.98	-1.29
PAR	-1.32	-0.67	-7.45*	-20.19*	1.34	-0.04	-2.07*	-1.07	-0.84	-1.00	-0.87	-1.10	-1.14

[(*)] See notes to Table 11a.

Figure 1: Simulated CPI Levels and 1st Log Differences

Simulation Models are Calibrated Using Monthly U.S. Data for the Period 1991-2004

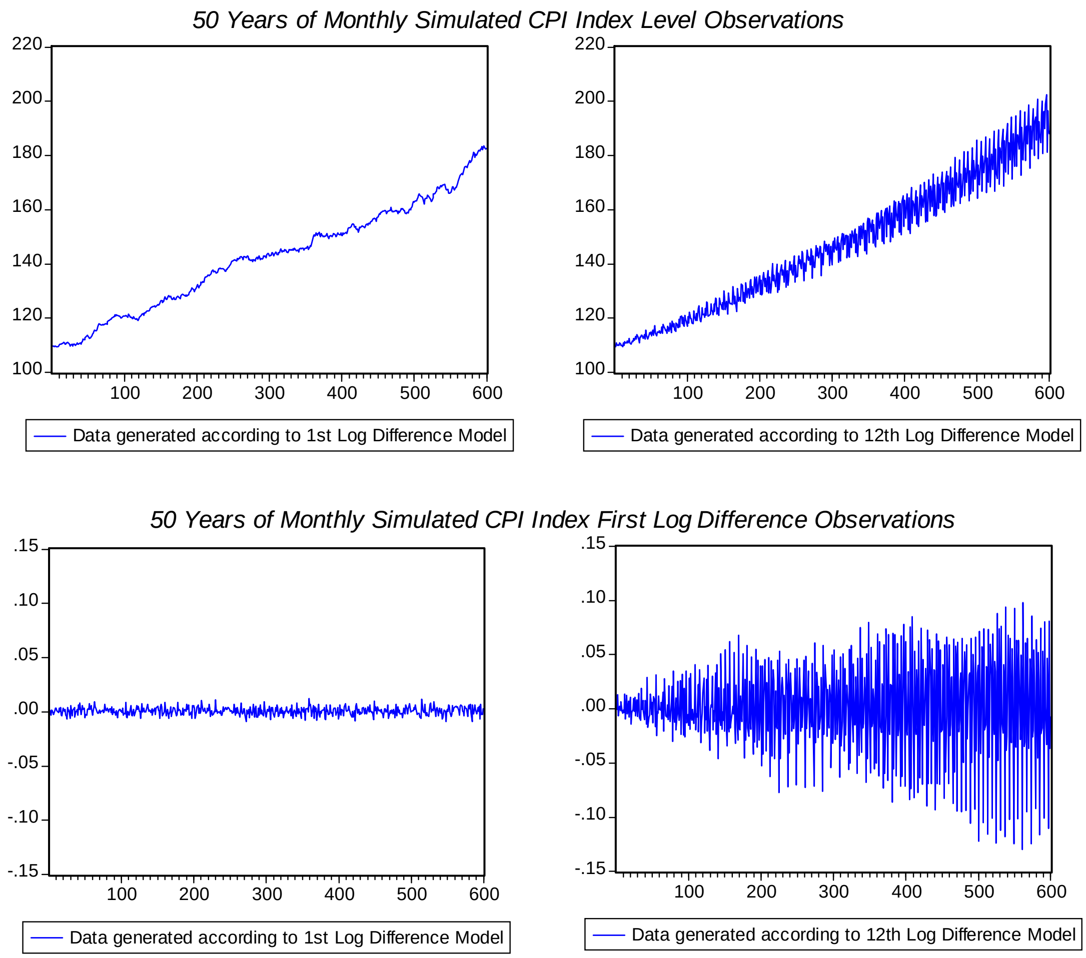
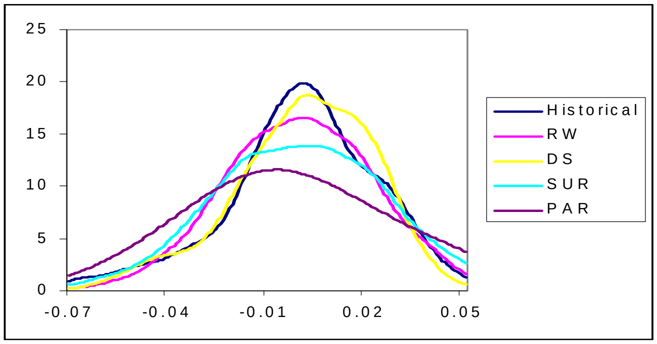
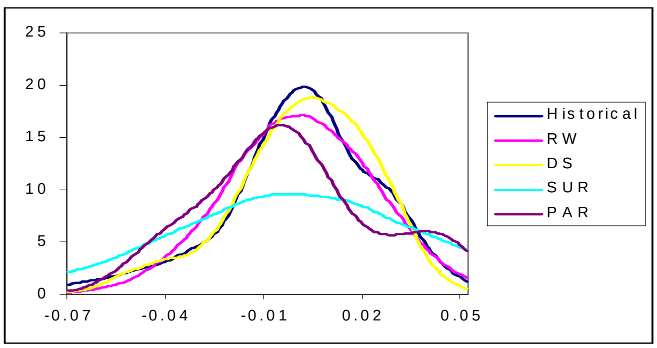


Figure 2a: Historical and Simulated First Log Difference Distributions

Panel 1: IP - 1959:1 - 2004:12 - 60 simulated values using all 100 simulation paths



Panel 2: IP - 1959:1 - 2004:12 - 300 simulated values using all 100 simulation paths



Panel 3: IP - 1959:1 - 2004:12 - 1200 simulated values using all 100 simulation paths

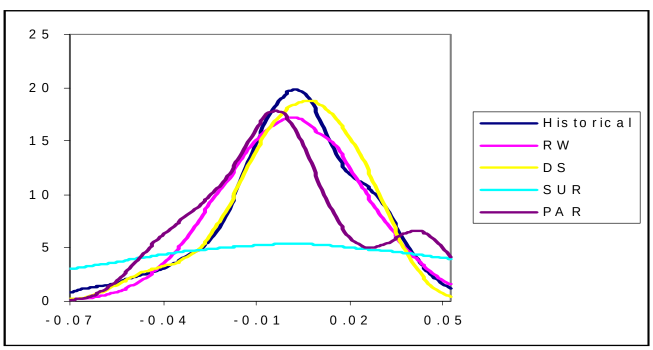
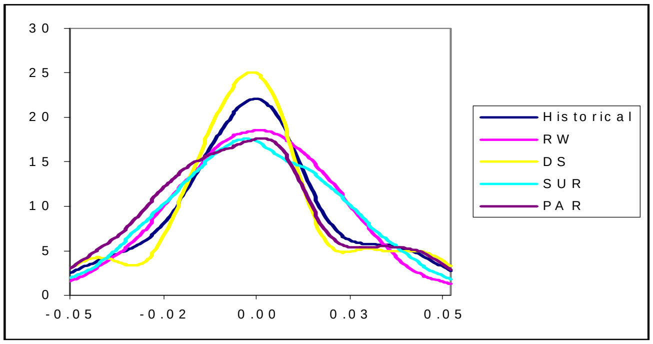
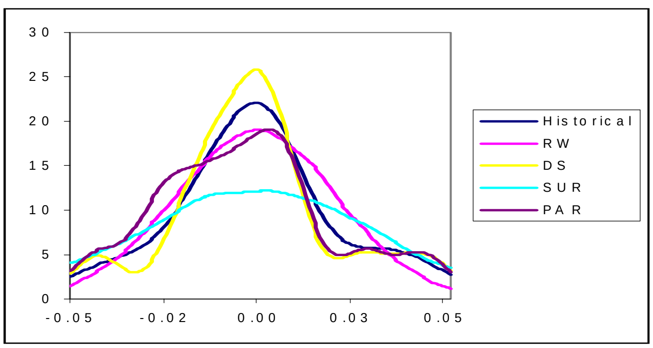


Figure 2b: Historical and Simulated First Log Difference Distributions

Panel 1: IP - 1991:1 - 2004:12 - 60 simulated values using all 100 simulation paths



Panel 2: IP - 1991:1 - 2004:12 - 300 simulated values using all 100 simulation paths



Panel 3: IP - 1991:1 - 2004:12 - 1200 simulated values using all 100 simulation paths

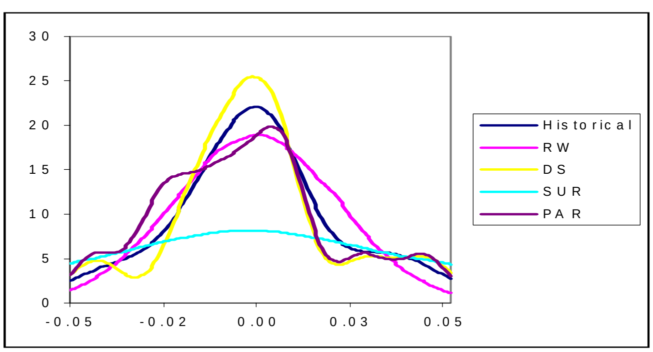
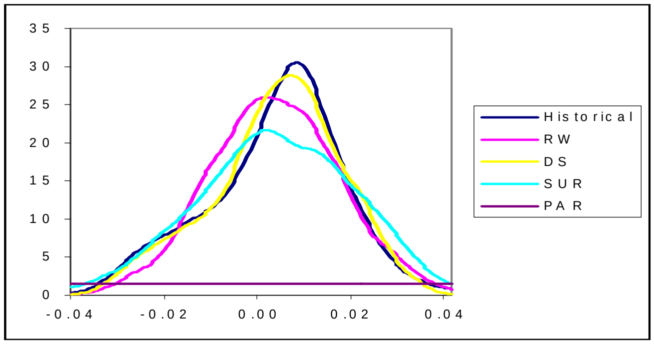
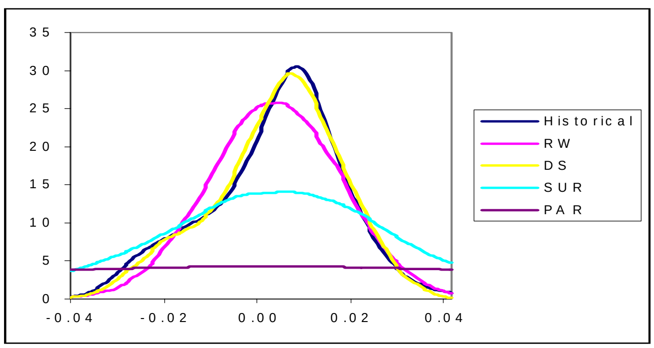


Figure 3a: Historical and Simulated First Log Difference Distributions

Panel 1: M1 - 1959:1 - 2004:12 - 60 simulated values using all 100 simulation paths



Panel 2: M1 - 1959:1 - 2004:12 - 300 simulated values using all 100 simulation paths



Panel 3: M1 - 1959:1 - 2004:12 - 1200 simulated values using all 100 simulation paths

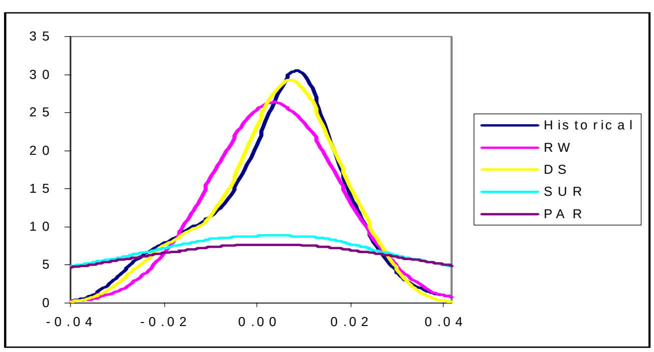
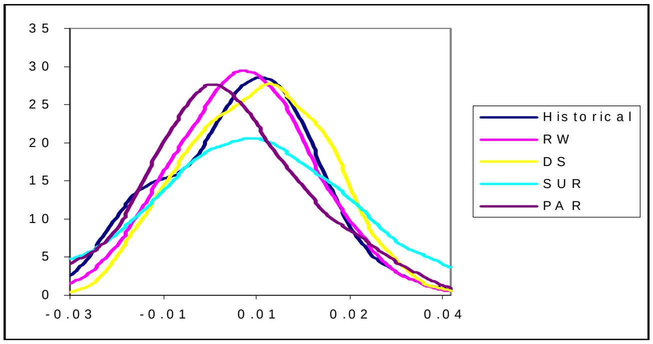
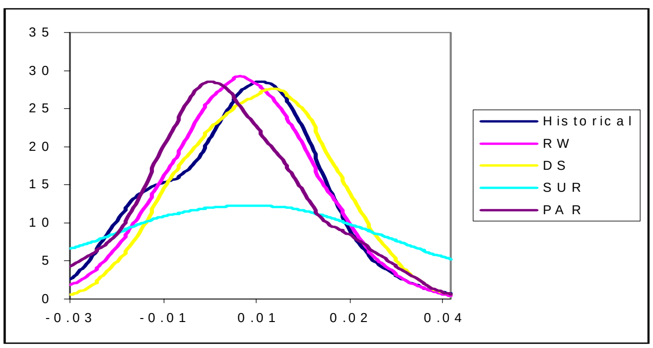


Figure 3b: Historical and Simulated First Log Difference Distributions

Panel 1: M1 - 1991:1 - 2004:12 - 60 simulated values using all 100 simulation paths



Panel 2: M1 - 1991:1 - 2004:12 - 300 simulated values using all 100 simulation paths



Panel 3: M1 - 1991:1 - 2004:12 - 1200 simulated values using all 100 simulation paths

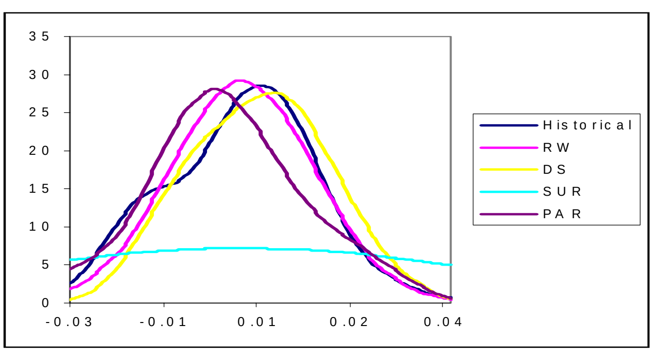
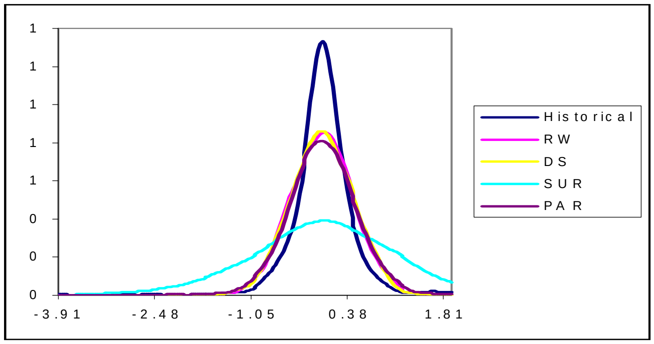
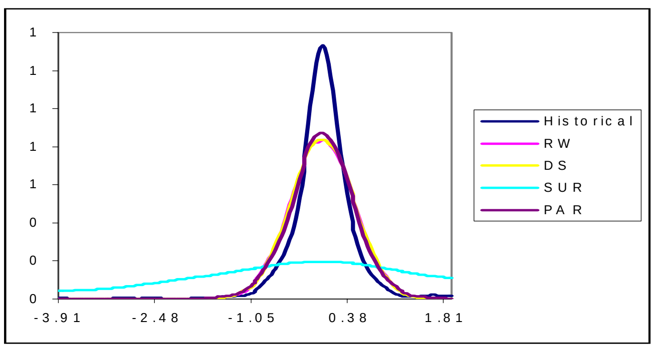


Figure 4a: Historical and Simulated First Difference Distributions

Panel 1: R1 - 1959:1 - 2004:12 - 60 simulated values using all 100 simulation paths



Panel 2: R1 - 1959:1 - 2004:12 - 300 simulated values using all 100 simulation paths



Panel 3: R1 - 1959:1 - 2004:12 - 1200 simulated values using all 100 simulation paths

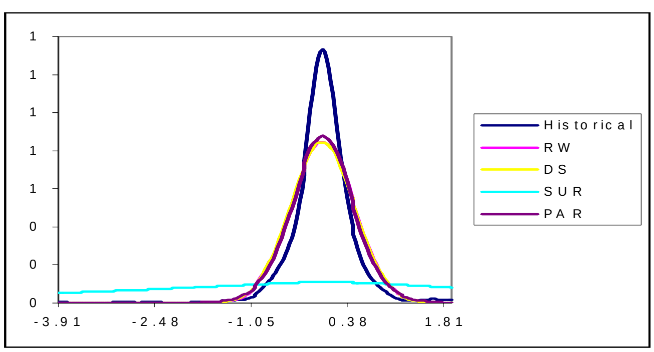
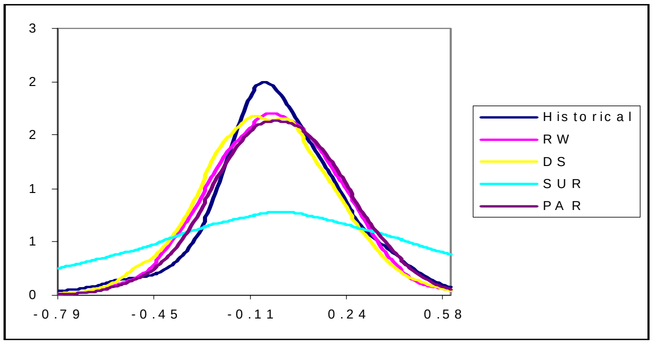
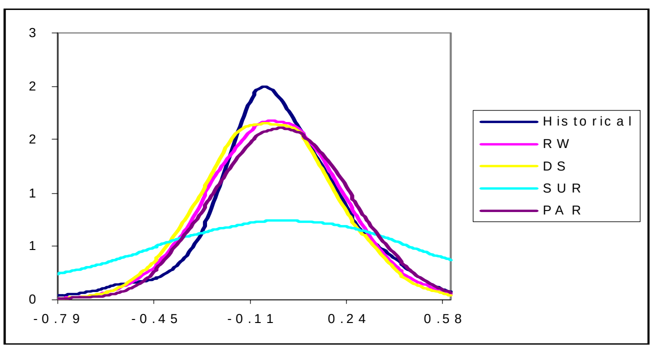


Figure 4b: Historical and Simulated First Difference Distributions

Panel 1: R1 - 1991:1 - 2004:12 - 60 simulated values using all 100 simulation paths



Panel 2: R1 - 1991:1 - 2004:12 - 300 simulated values using all 100 simulation paths



Panel 3: R1 - 1991:1 - 2004:12 - 1200 simulated values using all 100 simulation paths

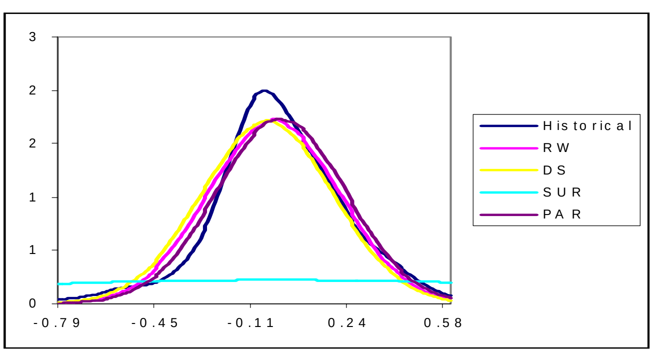
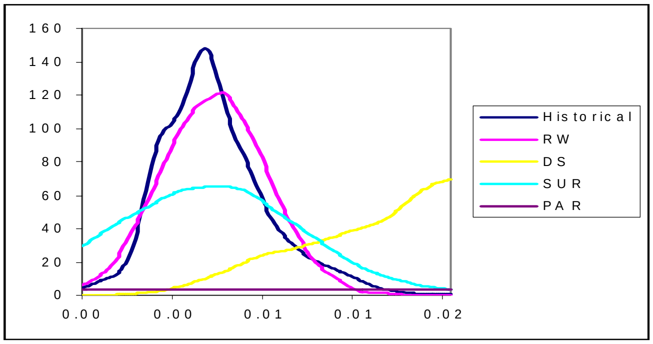
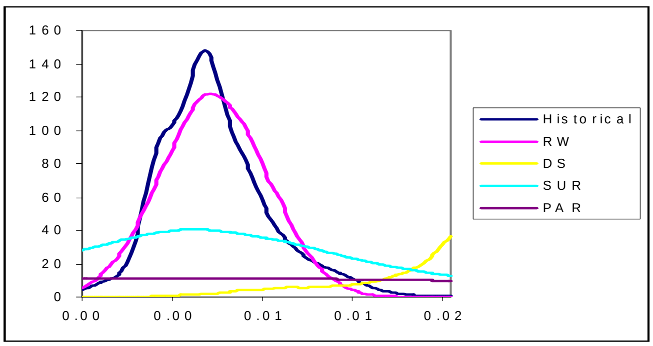


Figure 5a: Historical and Simulated First Log Difference Distributions

Panel 1: CPI1 - 1959:1 - 2004:12 - 60 simulated values using all 100 simulation paths



Panel 2: CPI1 - 1959:1 - 2004:12 - 300 simulated values using all 100 simulation paths



Panel 3: CPI1 - 1959:1 - 2004:12 - 1200 simulated values using all 100 simulation paths

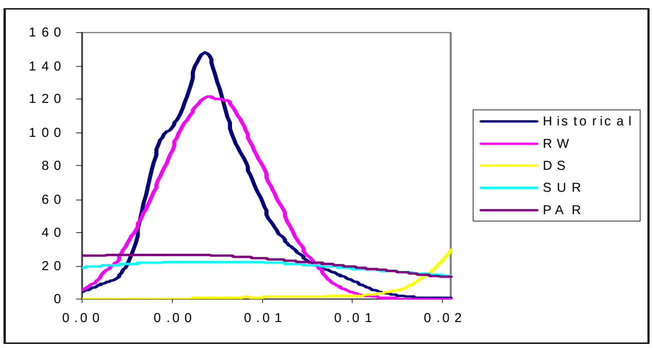
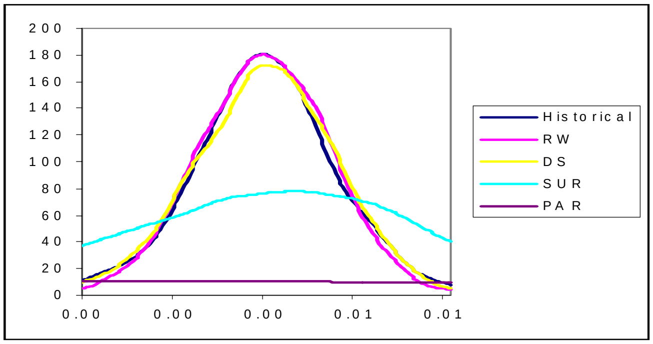
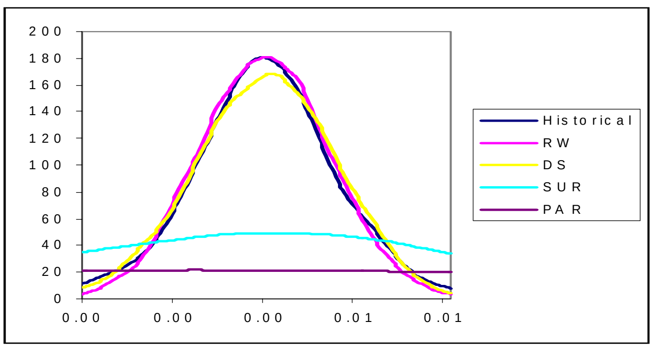


Figure 5b: Historical and Simulated First Log Difference Distributions

Panel 1: CPI1 - 1991:1 - 2004:12 - 60 simulated values using all 100 simulation paths



Panel 2: CPI1 - 1991:1 - 2004:12 - 300 simulated values using all 100 simulation paths



Panel 3: CPI1 - 1991:1 - 2004:12 - 1200 simulated values using all 100 simulation paths