TI 2001-029/4
Tinbergen Institute Discussion Paper

Inflation, Forecast Intervals and Long Memory Regression Models

Charles S. Bos
Philip Hans Franses
Marius Ooms

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Tinbergen Institute Discussion Paper TI-01-029/4

Inflation, forecast intervals and long memory regression models

Charles S. Bos**, Philip Hans Franses** and Marius Ooms*

23 February 2001

Department of Econometrics and Operations Research
vrije Universiteit *amsterdam*
De Boelelaan 1105
1081 HV Amsterdam, The Netherlands
email: ooms@econometriclinks.com

Abstract

We examine recursive out-of-sample forecasting of monthly postwar U.S. core inflation and log price levels. We use the autoregressive fractionally integrated moving average model with explanatory variables (ARFIMAX). Our analysis suggests a significant explanatory power of leading indicators associated with macroeconomic activity and monetary conditions for forecasting horizons up to two years. Even after correcting for the effect of explanatory variables, there is conclusive evidence of both fractional integration and structural breaks in the mean and variance of inflation in the 1970s and 1980s and we incorporate these breaks in the forecasting model for the 1980s and 1990s. We compare the results of the fractionally integrated ARFIMA(0,d,0) model with those for ARIMA(1,d,1) models with fixed order of $d = 0$ and $d = 1$ for inflation. Comparing mean squared forecast errors, we find that the ARMA(1,1) model performs worse than the other models over our evaluation period 1984-1999. The ARIMA(1,1,1) model provides the best forecasts, but its multi-step forecast intervals are too large.

Acknowledgements A first version of this paper was presented at the International Symposium on Forecasting in Lisbon (2000). We thank two referees and the participants at the symposium for their helpful comments.

* Corresponding author, Department of Econometrics and Operations Research, *vrije* Universiteit *amsterdam*, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.
** Econometric Institute, Erasmus University, Rotterdam, and Tinbergen Institute, The Netherlands,

1 Introduction

This paper concerns the usefulness of the ARFIMA-model for U.S. inflation for out-of-sample forecasting. We consider both point and interval forecasts and we also examine the usefulness of explanatory variables for different forecast horizons. Therefore we not only consider time variation in the coefficients for the mean of inflation, but also time variation in the forecast error variance.

The most useful explanatory variables for U.S. consumer price inflation are connected with the Phillips curve, with oil price shocks, and with changes in monetary policy. Gali & Gertler (2000) give references for relevant recent explanations. Unemployment, output gap variables and real unit labour costs correspond with the Phillips curve. Hooker (1999) summarises evidence on the effect of oil price shocks on postwar US inflation. Ball & Mankiw (1995) stress the effect of the sectoral distribution of price changes.

The relevant literature reveals that some variables are important for the explanation of short term inflation dynamics, whereas others may help to explain the longer run dynamics. Moreover, the effects of some variables, like the effect of the oil price on overall inflation, seems to have changed significantly over time. After 1980, oil price shocks did not have the same impact as in the 1970s, and monetary policy seems to have decreased both the mean and the variance of inflation, ceteris paribus. The empirical part of the economic literature also shows the sluggishness of inflation adjustment in adjusting to fundamentals. Indeed, many lags of inflation are statistically significant in reduced form equations of economic models, which is consistent with long memory behaviour found in time series analysis of inflation series, see Bos, Franses & Ooms (1999) and the references cited therein.

We examine the predictive ability of the dynamic regression models for several horizons, extending results of Stock & Watson (1999), who analysed only 12 month ahead forecasting. We confine our analysis to only a few relevant explanatory variables. Therefore we rule out large-scale leading indicator variables and sectoral asymmetry variables. Moreover, we take a statistical time series approach, where we derive multi-step forecasts from the likelihood for the model for one-step ahead predictions. In agreement with Stock & Watson (1999) we use a simulated out-of-sample forecasting framework, but we use fixed specifications for the recursive forecast evaluation period.

The remaining part of our paper is organised as follows. Section 2 starts with a recursive ARFIMAX analysis of monthly U.S. inflation and three leading indicators. We compare specifications for the error process up to an ARFIMAX(1,d,1) model and allow for deterministic regime changes. We use likelihood based time domain estimators based on the algorithm of Sowell (1992), see Doornik & Ooms (1999). This allows us to extend the model with macroeconomic leading indicators from the database developed by Stock & Watson (1999). Our starting point is a model with two level shifts in the period 1960-1999. We investigate the stability of the explanatory effects and we examine how they complement the simple level shift specification. We compare the forecasting ability of the models. We find that forecast intervals are too wide. Section 3 therefore analyses the results of weighted forecasting based on structural shifts in the variance. Section 4 employs statistical tests on the forecasts. Section 5 concludes.

2 Recursive ARFIMAX forecasting

We consider a monthly US consumer price index, as provided by the Bureau of Labor Statistics' website, July 2000. It concerns the influential core consumer price index, that is the U.S. city average items less food and energy, 1982-84=100, BLS code CUUR0000SA0L1E. We use data from 1960:04 to 1999:12 in our statistical analysis. Core inflation has not been affected by many outliers and it is therefore easier to interpret and analyse than other CPI-indexes.

2.1 Basic features of US core inflation

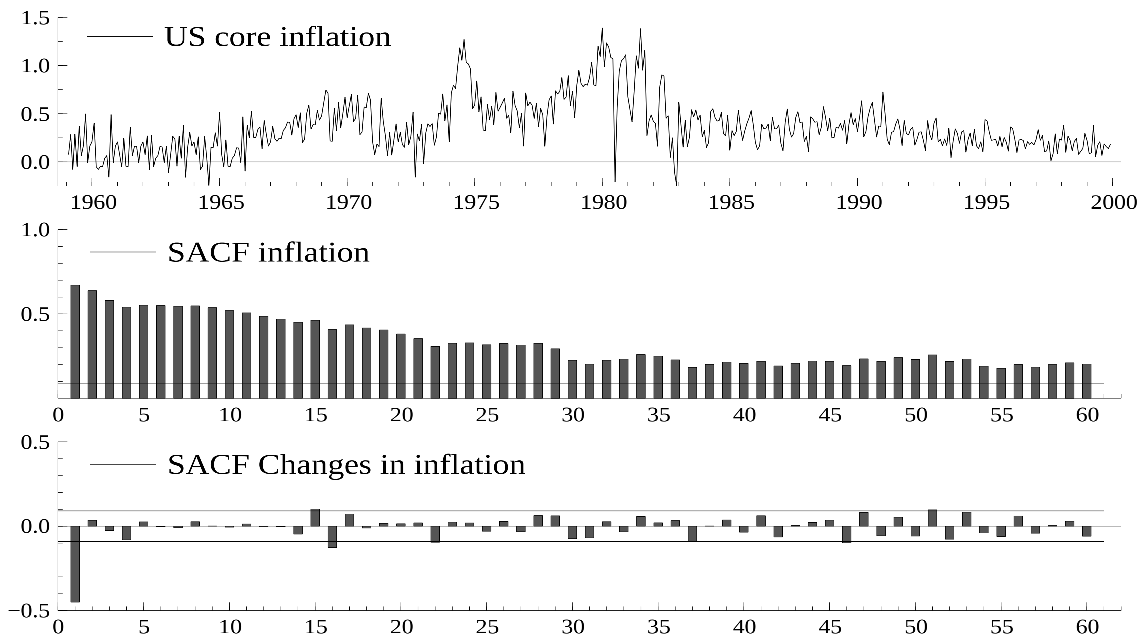

Figure 1: U.S. Monthly Core Inflation, ACF of levels and differences

Figure 1 shows a time series plot of monthly core inflation, measured as 100 times the log differences of the index. We seasonally adjust the series with two sets of centred seasonal dummies allowing for a break in the seasonal pattern in 1984. This break roughly corresponds with a change in the seasonal pattern detected by the official seasonal adjustment procedure used by the BLS. The autocorrelation functions of inflation and changes in inflation establish the long memory property of inflation, that is the inflation series appears "nonstationary", while the differenced inflation series appears to follow an MA(1) process. The combined information of Figure 1 suggests an order of integration larger than zero, but probably smaller than one. The autocorrelations also show that seasonal variation has been removed. In order to concentrate on "nonseasonal" variation only, we condition our analysis on these seasonally adjusted data. All figures were made in a β-version of GiveWin 2, see Doornik & Hendry (2001).

It is clear from Figure 1 that the current mean of inflation is significantly lower than in the 1970s. Different explanations coexist, but the combination of oil price increases and an accommodating interest rate policy by the US authorities is viewed as the most important cause of the exceptional inflation levels in the 1970s. The Volcker-Greenspan regime kept high inflation level at bay after 1983. Since we want to compare realistic forecasting models, we have to allow for at least two structural breaks in the mean for all models, one upward break in 1973.07 and one downward break in 1982.07. Furthermore we allow for a separate mean for inflation in 1980.07, when US prices were fixed in a high inflation period.

2.2 ARFIMAX modelling

The ARFIMA model has been introduced by Granger & Joyeux (1980) and Hosking (1981). Beran (1994) discusses the main asymptotic results for regression models with long memory errors. Hassler & Wolters (1995) and Ooms & Hassler (1997) showed that the ARFIMA(0,d,0) model with deterministic seasonality fits consumer price inflation series of many OECD countries rather well.

It is well known that ARMA models with an AR root close to the unit circle and an approximately cancelling MA root are also able to capture the low frequency characteristics typifying long memory. ARMA models can also be used to forecast long memory processes, see Brodsky & Hurvich (1999). ARIMA models are therefore interesting competitors in a forecasting study. The ARFIMA(1,d,1) model allows us to test ARIMA specifications by imposing restrictions. The last ACF of Figure 1 may also indicate an ARIMA(0,1,1) specification.

The empirical macroeconomic literature suggests that macroeconomic leading indicators can help to forecast yearly inflation rates out of sample, see Stock & Watson (1999). Hence, we extend the ARFIMA model with explanatory variables.

The first explanatory variable, u_t, is the US unemployment rate, using current seasonally adjusted data from the BLS, again to abstract from seasonality. This variable is used as a benchmark leading indicator for one-year-ahead forecasting of yearly inflation in Stock & Watson (1999) and its negative correlation with future inflation rates is associated in empirical macroeconometrics with the (old) "Phillips curve". The second variable is the short term interest rate, r_t, for which we choose the rate on federal three month treasury bills. In the short run, one observes a positive relationship between inflation and r_t. In the longer run, high interest rates are supposed to lead to lower inflation rates. The interrelationship is a complicated one, see Gali & Gertler (2000) for a recent analysis. The third leading indicator is the interest rate spread, $s_t = R_t - r_t$, where R_t, the long term interest rate, is the return on a 10 year U.S. treasury bill with constant maturity. A large spread is associated with higher inflationary expectations.

We specify the ARFIMAX model as follows:

$$(1-\phi_1 L)(1-L)^d(y_t - x_t'\beta - v_{t-H}'\gamma_H) = (1+\theta_1 L)\varepsilon_t. \tag{1}$$

with L the lag operator: $Ly_t = y_{t-1}$, y_t the monthly inflation rate, x_t a vector of deterministic terms, v_t a k-vector of leading indicators, H is the forecast horizon. The zero mean errors ε_t are Gaussian white noise with variance σ_ε^2. We require $0 \leq d \leq 1$, $|\phi_1| < 1$, $|\theta_1| < 1$ and $\phi_1 \neq -\theta_1$. Note that we need to lag the explanatory variables, v_t, at least H times in order to use them in an H-step-ahead forecasting model for inflation.

The column vector x_t consists of a constant, $x_{1,t} = 1$, measuring the "autonomous" mean of inflation, a level-shift for the high inflation period, $x_{2,t}$, and a single dummy for 1980.07, $x_{3,t}$: Observe the huge outlier in Figure 1. The variable $x_{2,t}$ equals one in the period

1973.07-1982.06 and is zero otherwise. A separate level-shift for the Volcker-Greenspan period equalling one in the period after 1982.06 turned out to have approximately zero effect, even in a recursive analysis starting in 1984. Therefore our model incorporates the same inflation regime before 1973 and after 1982.

When $d = 1$, the original constant term, β_1, drops out of the model and we use the $\Delta x_{i,t} = x_{i,t} - x_{i,t-1}$, $i = 2, 3$, $\Delta v_{j,t-H} = v_{j,t-H} - v_{j,t-H-1}$, $j = 1, \ldots, k$, as regressors for Δy_t. Note however, that all forecast results reported below refer to the same observations on y_t and partial sums of y_t, for all values of d.

The "deterministic component" of y_t is $x_t'\beta$. We conduct the time series analysis on y_t, corrected for the "interventions" $x_{2,t}\beta_2$ and $x_{3,t}\beta_3$, conditional on the information in the leading indicators.

Our three main specifications and their respective ranges for the order of integration for the stochastic component of inflation and log prices are summarised in Table 1.

Table 1: *ARFIMA components and their interpretation*

d	ϕ	θ	
0	$< 0, 1 >$	$< -1, 0 >$	: Short memory inflation, log prices are $I(1)$,
1	$< 0, 1 >$	$< -1, 0 >$	: Non-stationary inflation, log prices are $I(2)$,
$< 0, 0.5 >$	0	0	: long memory inflation, log prices are $I(d + 1)$

Table 2 presents estimation results for three models for two samples, 1960:04-1984.01 and 1960:04-1999:11, for the forecast horizon 1. They were obtained by maximum likelihood as implemented by Doornik & Ooms (1999) in Ox, see Doornik (1999). Key statistics for the specifications in Table 2 are compared with other models in Table 3. The results for the ARFIMAX(0,d,0) model in the first two columns of Table 2 provide evidence of fractional integration in the stochastic component of inflation, that is, d is about 0.25, and it differs significantly from 0 and 0.5. This coefficient is markedly lower than the corresponding value in the plain ARFIMA(0,d,0) model, see Table 3. Autonomous "baseline inflation", β_1, in the first and the last part of the sample is estimated between 0.25 and 0.28 percent per month in this specification. The estimate of β_2 indicates that the average autonomous inflation level was more than twice as high in the 1970s compared to the rest of the sample. The estimate for β_3 shows that there was an exceptional one-month drop in inflation in July 1980 of about 1 percent.

The third and fourth column of Table 2 show that the ARMA(1,1) model fits the inflation data about as well as the ARFIMA(0,d,0) specification. The estimates for β_1 and β_2 correspond quite closely to the estimates for the ARFIMA(0,d,0) model. The only marked difference for the initial sample is seen for the standard error of the constant term, β_1, which is smaller for the ARMA model. Apparently, the uncertainty about the mean of inflation is lower if one assumes an ARMA model. This partly reflects the underlying assumption of short memory ARMA models on the speed of convergence of the ML estimate of the mean, $\widehat{\mu}_T$, to the population mean μ as the sample size, T increases. In ARMA models one implicitly assumes $\mathrm{var}(\widehat{\mu}) = cT^{-1}, T \to \infty$. For ARFIMA models with $-0.5 < d < 0.5$ one allows for $\mathrm{var}(\widehat{\mu}) = cT^{2d-1}, T \to \infty$, see Adenstedt (1974) and Beran (1994, Ch. 9). Note that this difference for the standard error of β_1 is smaller for the longer sample, as the AR-parameter of the ARMA-model is closer to unity than in the initial sample. The estimate of the d-parameter of the ARFIMAX model does not change by increasing the sample.

The last columns of Table 2 show the results for the ARIMA(1,1,1) model. This model does not incorporate a fixed mean for inflation, that is, β_1 is not identified. An additional constant measuring the trend of inflation turned out to be insignificant and is therefore omitted. The structural breaks in inflation are not as significant as for the ARFIMAX(0,d,0) and ARMAX(0,0) model. The ARIMA(1,1,1) specification is able to pick up changes in the mean without the introduction of shift-dummies in the model.

The coefficients $\gamma_{1,u}$ and $\gamma_{1,r}$ in Table 2 concern the effect of the leading indicators. The coefficient $\gamma_{1,u}$ of lagged unemployment is negative for this forecast horizon, as expected, but this effect seems to have declined in the 1980s and 1990s. Lagged interest rates have a significant positive and stable effect as reflected in $\gamma_{1,r}$. These conclusions apply to all three models for the stochastic component of inflation.

Finally, diagnostic tests of Table 2 yield satisfactory outcomes for the period 1960-1983. For the period 1960-1999, there is evidence of heteroskedasticity, which is also reflected in the different estimates for σ_ε^2 for the two samples. This may signal a permanent downward shift in the innovation variance in the 1980s and 1990s. This can also be the message of the ARCH-test, see Lamoureux & Lastrapes (1990).

The relationships between y_t and the elements of v_{t-H} need to be relatively stable in order to be useful for forecasting. In theory one may even estimate a complete lag structure on all three variables simultaneously, but here we confine ourselves to one lag per variable at a time. We experimented with adding up to 3 extra lags and differences of the leading indicators, but this did not lead to significant improvements in either the in-sample fit or the out-of-sample forecasts.

Table 3 compares the models of Table 2 with other AR(FI)MA(X) specifications. We present a summary of the main results for the stationary ARMA(1,1), ARFIMA(0,d,0), ARFIMA(1,d,0), ARFIMA(0,d,1), ARFIMA(1,d,1) models and the nonstationary ARIMA(1,1,1) and ARIMA(0,1,1) specifications, both with and without the explanatory variables. An ARFIMA(1,d,1) specification does not provide a significantly better fit than either the ARFIMA(0,d,0) model or the ARMA(1,1) model. More importantly, the parameters of the ARFIMA(1,d,1) are not well identified, since both the FI-part and the ARMA-part can capture the low frequency characteristics of the process. In Table 3 we observe $\widehat{\phi}_1 = 0.98$ combined with $\widehat{d} = -0.63$, so that d cannot really be interpreted as the order of integration of the inflation process in this case.

The effect of the explanatory variables is similar across the different ARFIMAX-models: unemployment has a negative effect, interest rates have a positive effect, the spread does not have an additional effect, and the introduction of the leading indicators lowers the estimate of d in the ARFIMA models and it decreases the estimate of ϕ in the ARMA models. The explanatory variables account partly for the persistence of inflation.

Table 2: *ARFIMA model estimates for core inflation*

Parameter	ARFIMAX(0,d,0)				ARMAX(1,1)				ARIMAX(1,1,1)			
	60:04-84:01		60:04-99:11		60:04-84:01		60:04-99:11		60:04-84:01		60:04-99:11	
ϕ_1					0.757	(.10)	0.839	(.07)	0.153	(.09)	0.167	(.06)
d	0.245	(.05)	0.260	(.04)								
θ_1					-0.519	(.13)	-0.615	(.11)	-0.879	(.07)	-0.897	(.04)
β_1	0.256	(.11)	0.188	(.09)	0.251	(.08)	0.174	(.08)				
β_2	0.296	(.06)	0.298	(.05)	0.322	(.05)	0.303	(.05)	0.223	(.08)	0.255	(.07)
β_3	-0.919	(.19)	-0.941	(.16)	-0.934	(.19)	-0.961	(.16)	-0.930	(.19)	-0.938	(.16)
$\gamma_{1,u}$	-0.038	(.01)	-0.023	(.01)	-0.040	(.01)	-0.022	(.01)	-0.044	(.02)	-0.029	(.02)
$\gamma_{1,r}$	0.045	(.01)	0.041	(.01)	0.047	(.01)	0.043	(.01)	0.039	(.01)	0.036	(.01)
σ_ε	0.191		0.166		0.191		0.166		0.195		0.167	
LL	66.680		180.136		68.085		180.362		61.777		175.549	
AIC	-119.360		-346.273		-120.170		-344.725		-109.554		-337.098	
Normality	6.616	0.04	18.165	0.00	6.752	0.03	17.998	0.00	7.472	0.02	25.264	0.00
ARCH	2.436	0.12	9.358	0.00	1.958	0.16	11.147	0.00	3.574	0.06	12.518	0.00
Box-Pierce	20.574	0.90	27.922	0.57	20.010	0.89	31.972	0.32	25.339	0.71	33.327	0.31

Estimates of (1) for 3 specifications and two sample sizes, standard errors in parentheses. Diagnostic tests on residuals: Normality test on third and fourth moment, ARCH test using 1 lag, Box-Pierce test using 36 lags and p-values of the asymptotic distributions under the assumption of correct specification, see Doornik & Ooms (1999).

Table 3: *ARFIMAX-model estimates*

ϕ	d	θ	$\gamma_{1,u}$	$\gamma_{1,r}$	$\gamma_{1,s}$	AIC
.	.26	.	-.023	.041	.	-346.273
.06	.30	.	-.024	.041	.	-344.920
.	.28	.30	-.024	.041	.	-344.826
.84	.	-.62	-.022	.043	.	-344.725
.98	-.64	-.11	-.020	.043	.	-346.367
.	.26	.	-.027	.043	-.007	-344.442
-.06	.29	.	-.026	.042	-.005	-343.014
.	.29	-.05	-.026	.042	-.005	-342.925
.83	.	-.60	-.031	.047	-.014	-343.404
.97	-.63	-.11	-.028	.047	-.013	-344.939
.	.35	.	.	.	.	-320.429
-.11	.42	.	.	.	.	-321.180
.	.41	-.11	.	.	.	-320.796
.92	.	-.65	.	.	.	-319.111
-.29	.43	.18	.	.	.	-319.459
.17	1	-.90	-.029	.036	.	-337.098
.	1	-0.76	-.00	-.00	.	-308.251
.17	1	-.90	-.026	.030	.012	-335.599
.	1	-.76	.001	-.001	.004	-307.998
.17	1	-.85	.	.	.	-317.565
.	1	-.76	.	.	.	-312.148

Sample period: 1960:4-1999:11, a . indicates that parameter is restricted to zero

Table 4: *Asymptotic rates of growth of the variances of long-run forecast intervals for univariate models for log prices, given the order of integration d for the inflation process*

$d = 0$:	$\text{var}(\log(P_{t+H}/P_t)\|t)$ is $c \cdot H$,
$d = 1$:	$\text{var}(\log(P_{t+H}/P_t)\|t)$ is $c \cdot H^3$
$0 < d < 1$:	$\text{var}(\log(P_{t+H}/P_t)\|t)$ is $c \cdot H^{2d+1}$

$H \to \infty, H/T \to 0$, see Beran (1994, §8.6)

2.3 Recursive estimation and forecasting

Next we examine the point forecasts of the different models in a simulated out-of-sample experiment. We estimate parameters of a range of models recursively. Again, we compare univariate models with specifications with both single regressors and multiple regressors, that is: we vary the orders p, q and k in (1) and we estimate d or put it equal to zero or unity.

We start with sample 1960.04-1984.1 and end with sample 1960.04-1999:11. We make point predictions of monthly inflation, $\widehat{y}_{t+H}$, for multiple horizons, $H = 1, 3, 6, 12, 24$. Moreover, we compute cumulative ("July 1998-June 1999") $\sum_{i=1}^{H} \widehat{y}_{t+i}$ predictions for quarterly, half-yearly, yearly and two-yearly inflation. Cumulative inflation forecasts can be interpreted in other ways. They correspond with forecasts of the log price level H periods ahead, $\log(P_{t+H})$, minus todays log price level, $\log(P_t)$. For the univariate models, they also equal the forecasts of H times the future mean of inflation over the forecast horizon.

For the univariate models, $k = 0$ in (1), we expect the longer run forecasts of the ARMA(1,1) specification to change the least as new inflation shocks enter the information set. The ARFIMA(0,d,0) forecasts allow for longer lasting deviations from the long run mean and are therefore expected to be more variable. The ARIMA(1,1,1) forecasts should show the largest variation. The differences between the three models in the variability of the long run forecasts should be reflected in their estimated forecast intervals. We summarise the asymptotic characteristics of the forecast intervals of the univariate models in Table 4, where it is assumed that $H \to \infty, H/T \to 0$, where T is the sample size, so that the mean can be treated as known, see Beran (1994, §8.6). Table 4 makes clear that d is an influential parameter for long run interval forecasting of the log price level, at least theoretically. As the explanatory power of the leading indicators decreases for longer forecast horizons we can expect a similar behaviour for the long run forecast intervals of the models with regressors. It remains an empirical question whether the asymptotic results of Table 4 provide a good indication for the sample sizes and forecast horizons in our study.

2.4 Recursive estimates

Before analysing the recursive out-of-sample forecasting performance of the models, we examine their recursive parameters estimates. We first address the stability of the dynamic parameters d, ϕ, θ and the mean parameter β_1 over our recursive estimation period. Figure 2 shows the recursive estimates of the ARFIMAX(0,d,0) model with u_{t-1} and r_{t-1} as leading indicators. Figure 3 shows the recursive estimates of the corresponding ARMA(1,1) model. The beginpoints and endpoints of these figures correspond to the results in Table 2. Both figures indicate that the effect of interest rates and unemployment has become less significant over the last decades, although the effect of interest rates has stabilised after 1993. The estimated d for the ARFIMAX(0,d,0) model is more stable than the corresponding ϕ and θ of the ARFIMAX(1,0,1) model. For comparison we present the recursive parameter estimate of the univariate ARFIMA(0,d,0) model in Figure 4. The parameters of this simple model hardly change over time. This may be an advantage in forecasting. Note again that d is much higher in this univariate model.

The difference between the univariate specification and the model with leading indicators is smaller for multi-step forecasting at longer horizons. This is illustrated in Figure 5 which shows the recursive estimates of an ARFIMAX(0,d,0) model with r_{t-24} and u_{t-24} as the leading indicators. These estimates for d are closer to the values obtained for the univariate model, as the explanatory power of r_{t-24} and u_{t-24} is considerably smaller than in the one-step-ahead forecasting model. Note also the negative sign of $\gamma_{24,r}$ as high short term interest rates are related with lower monthly inflation figures after two years. This also leads to a higher estimate for β_1 compared with Figure 5 as $\gamma_{24,r}r_{t-24}$ has a negative mean over the sample whereas $\gamma_{1,r}r_{t-1}$ was clearly positive.

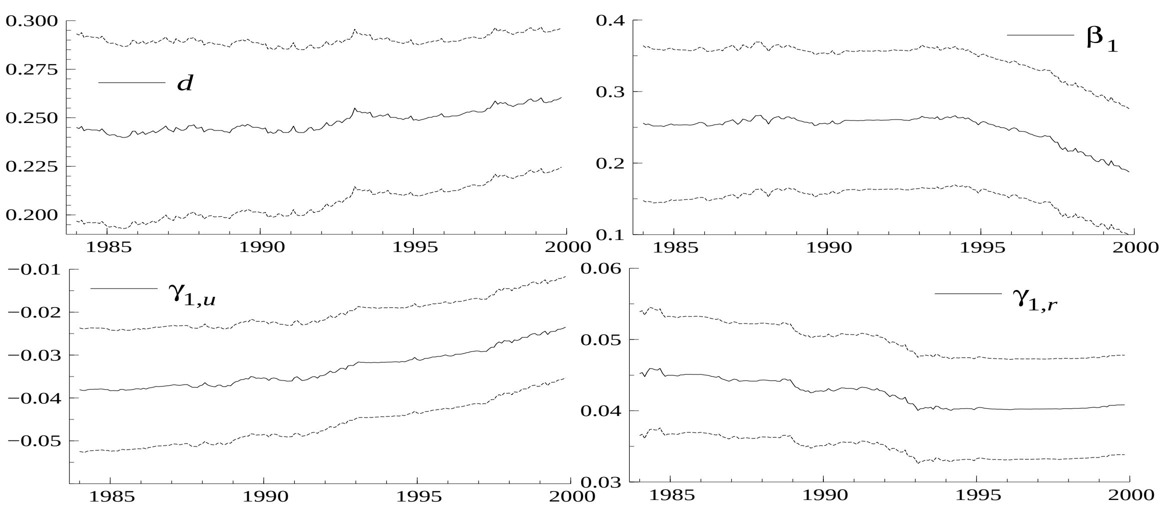

Figure 2: Recursive ARFIMAX(0,d,0) estimates ± one s.e., d, β_1, $\gamma_{1,u}$, $\gamma_{1,r}$

2.5 Recursive forecasting

Computing forecasts for an ARFIMA(p, d, q) model is not trivial. An ARFIMA(p, d, q) model cannot be written as a finite order ARMA model or as a finite dimensional state space model, so that standard methods do not apply. Here we use the optimal linear forecast of $y_{t+H} - x'_{t+H}\beta - v'_t\gamma_H$ given $y_1 - x'_1\beta - v'_{1-H}\gamma_H, \dots, y_t - x'_t - v'_{t-H}\gamma_H$, see Beran (1994, §8.7) and Doornik & Ooms (1999) for the exact implementation. This forecast function explicitly assumes that forecasts are generated out of a finite sample. It takes the estimated covariance function of the stochastic part of the model as an input and results in a time dependent forecast function depending on all available observations. Forecast standard error estimates are computed accordingly. We use RMSEmod, the RMSE as derived from the model, to denote this forecast standard error estimate below. For the ARIMA model, $d = 1$, we cumulate the ARMA forecasts for the changes in inflation back to inflation forecasts. Forecast standard errors are adjusted for this cumulation as well.

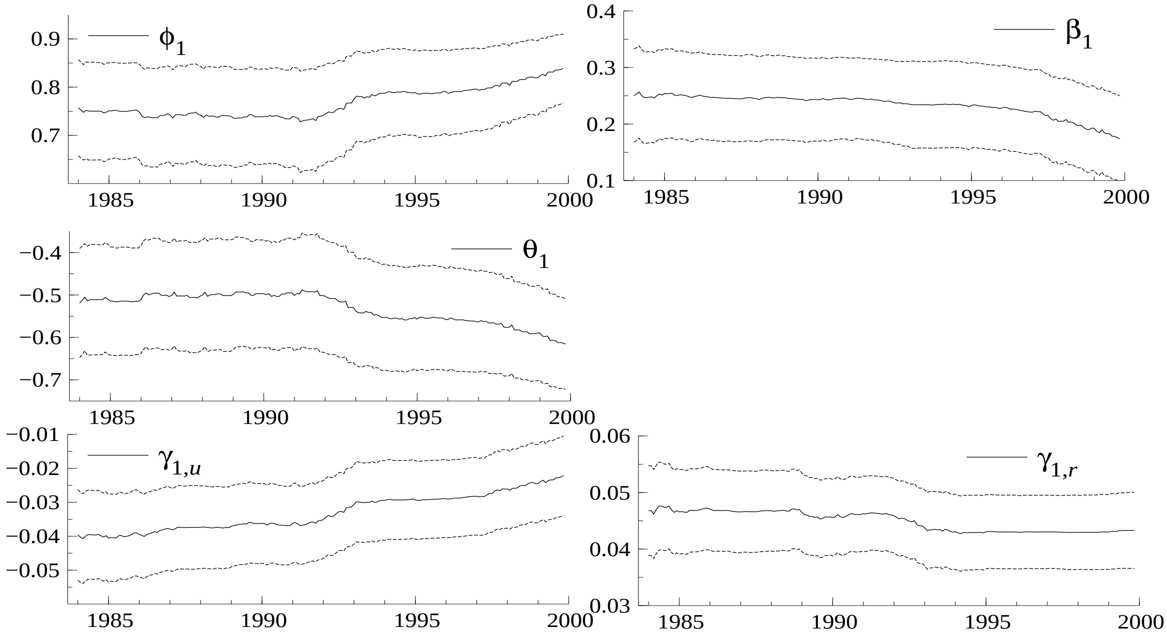

Figure 3: Recursive ARMAX(1,1) estimates ± one s.e., ϕ_1, θ_1, β_1, $\gamma_{1,u}$, $\gamma_{1,r}$

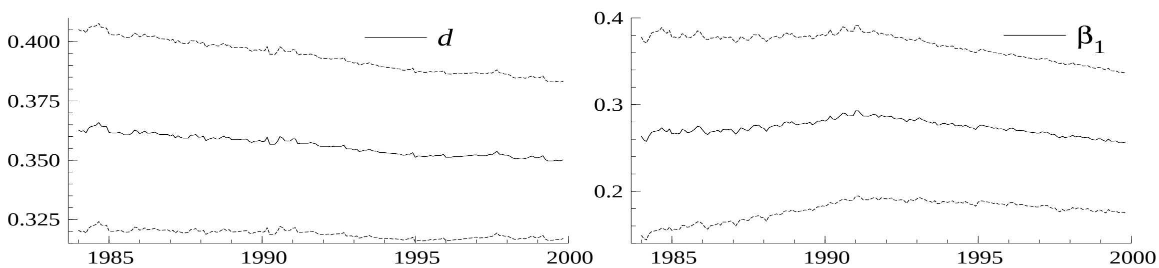

Figure 4: Recursive ARFIMA(0,d,0) estimates ± one s.e., d, β_1

We recursively compute the forecasts for the different horizons and compare them with actual values. Figure 6 illustrates 2 year ahead forecasting. It displays time series plots of the forecasts $\widehat{y}_{t+24|t}$, and the "predetermined" part of the forecast, $x'_{t+H}\widehat{\beta}_t$, together with the actual values y_{t+24}. We present only the forecasts of the univariate ARFIMA(0,d,0) model in graphical form. We give numerical evidence on all relevant models and all forecast horizons below. For the ARFIMA(0,d,0) model of Figure 6 we observe a persistent deviation between $\widehat{y}_{t+24|t}$ and $x'_{t+24}\widehat{\beta}_t$, for t between 1984 and 1994. This period corresponds to a period of persistently higher inflation, captured by the stochastic part of the ARFIMAX(0,d,0) model. The forecasts of the ARIMAX(1,1,1) model, not shown here, are not as smooth, but they follow the trend in inflation more closely.

The bottom graph of Figure 6 shows the cumulative forecast intervals for 2-year inflation.

The ARFIMA(0,d,0) model tracks the downward swing in average inflation in the beginning of the 1990s more closely than the ARMA(1,1) model. The forecast intervals of the two

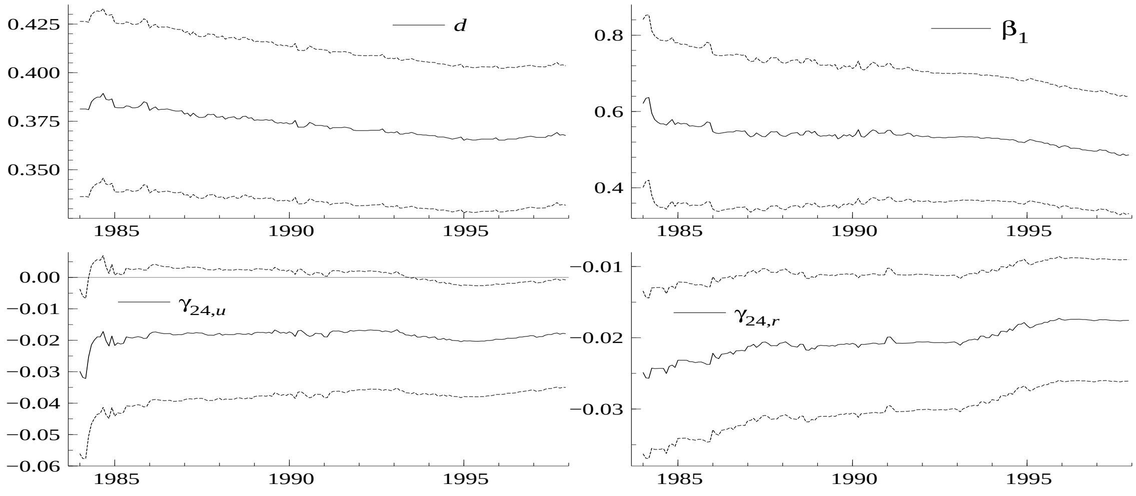

Figure 5: Recursive ARFIMAX(0,d,0) estimates $\pm$ one s.e., d, β_1, $\gamma_{24,u}$, $\gamma_{24,r}$

models for this horizon are equally wide. This suggests that the differences between these two models are not as large in practice as asymptotic theory suggests. The cumulative forecasts of the ARIMA(1,1,1) model are again more volatile, but they naturally follow the persistent decline in inflation in the 1990s quite well. The forecast intervals are clearly much wider than for the ARFIMA(0,d,0) model. The leading indicator models show more volatility than the univariate models in their 2 year ahead forecasts, especially in the beginning of the evaluation period in 1984, when unemployment and interest rates were volatile and values for 1985 and 1986 were forecast.

Table 5 summarises the cumulative forecasting results for four univariate models, the ARMA(1,1), the ARFIMA(0,d,0), the ARFIMA(1,d,0) and the ARIMA(1,1,1) model. Results for the leading indicator models are presented below.

We present three measures of forecasting performance. First, we report the mean forecast error, MFE $= (t_2 - t_1 + 1)^{-1} \sum_{t=t_1}^{t_2} e_{t+H}$, with $e_{t+H} = \widehat{z}_{t+H} - z_{t+H}$, where $\widehat{z}_{t+H}$ is the cumulative inflation estimate for the period t to $t+H$, using the data and parameter estimates up to time t, $t+H = (1984.1)+H \ldots 1999.12$. Second, Table 5 reports the root mean squared error, RMSE $= \left((t_2 - t_1 + 1)^{-1} \sum e_{t+H}^2\right)^{0.5}$, Third, we compute the mean absolute prediction error, MAPE $= (t_2 - t_1 + 1)^{-1} \sum |e_{t+H}|$. We also report the root of the H-step-ahead forecast error variance as predicted by the model, RMSEmod. The RMSEmod-values in the table are the mean values of the recursively computed model estimates over the evaluation sample. Finally we present the ratio of RMSEmod and RMSE to give a first indication of the coverage probability of the forecast intervals. We note that forecast intervals for the ARFIMAX(0,d,0) model and the ARMAX(1,1) specification are approximately equally wide, whereas the actual forecast RMSE for the ARFIMAX(0,d,0) model is considerably smaller for larger horizons.

The results for the best model for each forecast criterion are printed in boldface. It appears that the ARIMA(1,1,1) model provides the best forecasts overall at all horizons. The ARMA(1,1) model performs the worst. The relative differences in forecasting performance between the models increase with the forecast horizon. The one-step-ahead forecasts of the models are very similar. The estimated model forecast error variances (RMSEmods) agree

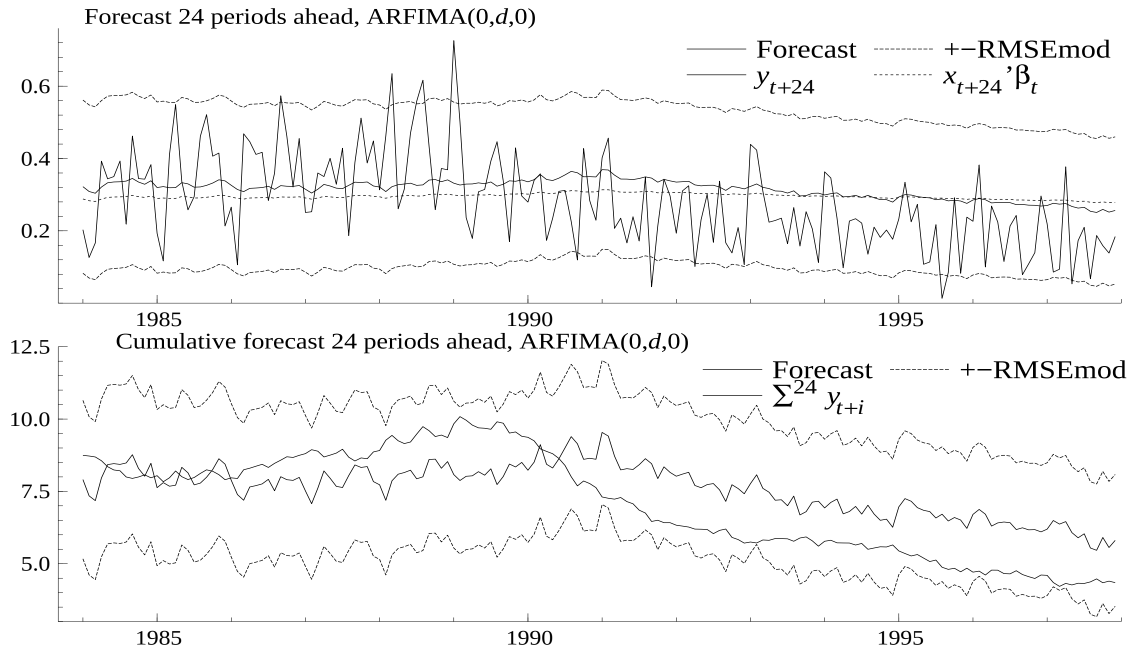

Figure 6: Recursive two-year-ahead forecasts ARFIMA(0,d,0) model $\pm$ one s.e., exogenous part of the forecast and actual inflation corresponding period, Bottom graph: cumulative forecasts over two year period

very closely for horizons 1 to 6. Under correct specification these RMSEmods are expected to slightly underestimate the true forecast error variance in finite samples, because they neglect the effect of parameter uncertainty, but here we observe for all models that the RMSEmods are significantly higher than the actual RMSEs. In particular for the ARIMA(1,1,1) model, the estimated forecast error variances for the 12 and 24 period ahead cumulative forecasts are much too large. Refer also back to Figure 6, which shows that all 2-year inflation outcomes lie in a one-σ forecast interval.

Table 6 contains a selection of the corresponding simulated out-of-sample forecasting results for a range of leading indicator models. We examined the forecasting performance of 6 specifications for the stochastic part and 5 specifications for the leading indicator part. We allowed for ARMA(1,1), ARFIMA(0,d,0), ARFIMA(1,d,0), ARFIMA(0,d,1), ARIMA(1,1,1), ARIMA(0,1,1) errors. We use three single leading indicator models with u, r, s, one double leading indicator model with u and r (in Table 2) and finally a model with both u, r and s. We present the results for the best and for the worst of these 30 models for three forecasting criteria, MFE, RMSE and MAPE, and for 5 horizons: 1, 3, 6, 12, and 24 months. However, for the single leading indicator r_{t-H} we show the outcomes for all stochastic specifications, as the interest rates proves to be the most interesting explanatory variable.

Table 6 shows that r_{t-1} is the best leading indicator for one-step-ahead forecasting. Using only r_{t-1} brings the mean forecast error in inflation down to .001 percent for the ARFIMAX(0,d,0), ARFIMAX(1,d,0), ARFIMAX(0,d,1) and ARIMAX(1,1,1) models. The RMSE is brought down to 0.113 percent per month. Using u_{t-1} or s_{t-1} does not lead to predictions that outperform the univariate models in short run forecasting. Using a combi-

Table 5: *Simulated cumulative out-of-sample forecasts 1984-1999 for univariate models for core inflation*

	Model			*H*				
	p	d	q	1	3	6	12	24
MFE	1	0	1	**0.001**	**0.001**	**0.015**	**0.059**	**0.280**
MFE	1	1	1	*0.006*	0.019	0.051	0.153	0.453
MFE	0	d	0	0.005	0.020	0.058	0.164	0.471
MFE	1	d	0	*0.006*	0.023	*0.064*	*0.179*	*0.501*
MFE	0	d	1	*0.006*	0.022	0.063	0.175	0.493
RMSE	1	0	1	*0.118*	*0.251*	*0.386*	*0.697*	*1.550*
RMSE	1	1	1	**0.115**	**0.233**	**0.339**	**0.552**	**1.228**
RMSE	0	d	0	0.116	0.240	0.358	0.616	1.344
RMSE	1	d	0	0.117	0.241	0.355	0.600	1.317
RMSE	0	d	1	0.116	0.240	0.354	0.601	1.320
MAPE	1	0	1	*0.094*	*0.202*	*0.315*	*0.589*	*1.398*
MAPE	1	1	1	**0.091**	**0.182**	**0.266**	**0.439**	**0.969**
MAPE	0	d	0	0.092	0.194	0.296	0.528	1.211
MAPE	1	d	0	0.092	0.194	0.289	0.510	1.182
RMSE	0	d	1	0.092	0.193	0.289	0.511	1.185
RMSEmod	1	0	1	0.183	0.415	0.761	1.455	2.530
RMSEmod	1	1	1	0.185	0.424	0.761	1.519	3.439
RMSEmod	0	d	0	0.184	0.427	0.751	1.354	2.484
RMSEmod	1	d	0	0.183	0.417	0.748	1.389	2.611
RMSEmod	0	d	1	0.183	0.418	0.747	1.381	2.582
RMSEmod/RMSE	1	0	1	1.554	1.657	1.972	2.088	1.633
RMSEmod/RMSE	1	1	1	1.612	1.820	2.247	2.750	2.800
RMSEmod/RMSE	0	d	0	1.587	1.779	2.096	2.197	1.848
RMSEmod/RMSE	1	d	0	1.569	1.729	2.108	2.315	1.983
RMSEmod/RMSE	0	d	1	1.577	1.743	2.111	2.299	1.956

Best ranking results in **boldface**, worst ranking results in *italics*

nation of u_{t-1} and r_{t-1} does not improve upon the univariate models either. We observe similar results for forecast horizons of three and six months. Using only r_{t-3} and r_{t-6} respectively, leads to the most accurate forecasts. Forecast efficiency as measured by the RMSE is improved by two percent compared with univariate specifications.

For the one year and two year horizon we find r_t is no longer the dominating leading indicator. At the one year horizon, the ARIMA(1,1,1) model with s_{t-12} as a single leading indicator provides the most accurate forecasts, although the other models with s_{t-12}, not shown in the table, are still dominated by their counterparts with r_{t-12}. At the two year horizon we see that u_{t-24} is the best leading indicator, despite the slowly declining effect we observed in the recursive parameter estimates. However, the univariate models provide forecasts with better RMSEs.

Apparently, a substantial part of the long swings in inflation the 1980s and 1990s is successfully accounted for by the positive effect of lagged short term interest rates. On the other hand, an increase in short term interest rates can also be associated with a lower inflation in the longer run. This was illustrated by the recursive estimates of the negative coefficient γ_{24}, of r_{t-24} in Figure 5. This coefficient is less pronounced than the coefficient of r_{t-1}.

Overall, the ARIMA(1,1,1) model dominates the forecast performance of the other spec-

Table 6: *Simulated cumulative out-of-sample forecast performance 1984-1999 for leading indicator models for US core inflation*

	Model				*H*				
		p	d	q	1	3	6	12	24
MFE	r	1	0	1	0.003	0.004	**0.009**	**0.048**	0.486
MFE	r	1	1	1	**0.001**	0.010	0.048	0.157	1.042
MFE	r	1	d	0	**0.001**	0.007	0.044	0.143	0.955
MFE	r	0	d	0	**0.001**	0.006	0.035	0.120	0.815
MFE	r	0	d	1	**0.001**	0.007	0.041	0.137	0.938
MFE	u	1	0	1	0.002	**0.004**	0.013	**0.044**	**0.203**
MFE	u	1	1	1	*0.015*	0.066	0.175	0.368	0.567
MFE	urs	1	1	1	0.010	*0.070*	*0.215*	*0.512*	1.636
MFE	urs	1	d	0	0.006	0.025	0.088	0.246	*1.187*
RMSE	r	1	0	1	0.115	0.245	0.378	0.658	1.732
RMSE	r	1	1	1	**0.113**	**0.229**	**0.334**	0.559	2.066
RMSE	r	1	d	0	0.114	0.235	0.345	0.571	1.679
RMSE	r	0	d	0	0.114	0.234	0.344	0.575	1.632
RMSE	r	0	d	1	0.114	0.235	0.344	0.571	1.669
RMSE	s	1	1	1	0.116	0.247	0.359	**0.514**	1.701
RMSE	u	1	1	1	0.116	0.233	0.360	0.616	**1.245**
RMSE	ur	1	0	1	*0.119*	*0.268*	*0.432*	0.748	1.724
RMSE	urs	1	0	1	*0.119*	*0.268*	0.430	*0.753*	1.896
RMSE	urs	1	1	1	0.114	0.247	*0.445*	0.724	*2.407*
MAPE	r	1	0	1	0.092	0.199	0.307	0.546	1.538
MAPE	r	1	1	1	**0.090**	**0.183**	**0.263**	0.444	1.583
MAPE	r	1	d	0	0.091	0.191	0.277	0.483	1.404
MAPE	r	0	d	0	0.091	0.191	0.278	0.488	1.383
MAPE	r	0	d	1	0.091	0.191	0.277	0.483	1.396
MAPE	s	1	1	1	0.092	0.195	0.294	**0.415**	1.356
MAPE	u	1	1	1	0.092	0.187	0.281	0.499	**0.997**
MAPE	ur	1	0	1	0.096	*0.225*	*0.362*	0.616	1.466
MAPE	ur	1	1	1	0.092	0.195	0.284	0.536	*1.982*
MAPE	urs	1	0	1	*0.097*	*0.225*	*0.362*	*0.617*	1.623

Best ranking results in **boldface**. Worst ranking results in *italics*

ifications for the error term. Although the persistent shifts in inflation can be modelled by explanatory variables, this does not entail a better forecasting performance for the ARFIMAX(0,d,0) model, compared with ARIMAX(1,1,1). Comparing the results of Tables 5 and 6 overall, we do not observe large increases in forecasting precision by the addition of explanatory variables, but the regression variables do help for short run forecasting. For $H = 24$ the univariate models outperform the regression models on all criteria. Adding more regressors worsens the forecasting performance at all horizons. For short horizons, the precision loss is a few percent, but for two-year-ahead forecasting the differences are dramatic, resulting in RMSEs which are 20 percent larger than for univariate models.

Table 7 presents the predicted forecast root mean squared error, RMSEmod, of the leading indicator models and compares them with their actual forecast RMSEs. Comparing the RMSEmods for the ARFIMAX(0,d,0) model using r_{t-H} in Table 7 with the RMSEmods for the univariate ARFIMA(0,d,0) model in Table 5 we see lower values for shorter horizons and higher values for $H = 24$. These higher RMSEmods are connected with two effects, first the lower explanatory power of r_{t-24} compared with r_{t-1} and second the higher estimated d in the model with $H = 24$. Overall, the predicted forecast error variances as measured by

RMSEmod are much larger than actual RMSEs, especially for the ARIMAX(1,1,1) models. For horizon 1, RMSEmods are a factor 1.6 too large on average. For $H = 24$ this factor is even 2.7 for the ARIMAX(1,1,1) model with leading indicator u_{t-24}. The most likely reason for this overestimation of the scale of forecast intervals is the persistently low innovation variance over the period of the forecasting exercise, compared to the variance in the earlier estimation period.

Table 7: *Predicted and actual root mean squared error of leading indicator models for US core inflation*

	Model				H				
		p	d	q	1	3	6	12	24
RMSEmod	r	1	0	1	0.179	0.406	0.741	1.381	3.117
RMSEmod	r	1	1	1	**0.181**	**0.414**	**0.754**	1.522	3.342
RMSEmod	r	1	d	0	0.179	0.402	0.728	1.367	2.737
RMSEmod	r	0	d	0	0.179	0.405	0.722	1.328	2.561
RMSEmod	r	0	d	1	0.179	0.402	0.725	1.357	2.713
RMSEmod	s	1	1	1	0.182	0.404	0.698	**1.464**	3.382
RMSEmod	u	1	1	1	0.183	0.408	0.697	1.316	**3.367**
RMSEmod	ur	1	0	1	0.177	0.397	0.700	1.282	3.233
RMSEmod	urs	1	0	1	0.177	0.396	0.694	1.222	2.251
RMSEmod	urs	1	1	1	0.180	0.398	0.686	1.314	2.885
RMSEmod/RMSE	r	1	0	1	1.553	1.661	1.962	2.098	1.800
RMSEmod/RMSE	r	1	1	1	**1.600**	**1.808**	**2.257**	2.720	1.618
RMSEmod/RMSE	r	1	d	0	1.563	1.706	2.111	2.392	1.630
RMSEmod/RMSE	r	0	d	0	1.575	1.732	2.098	2.308	1.570
RMSEmod/RMSE	r	0	d	1	1.567	1.716	2.110	2.376	1.625
RMSEmod/RMSE	s	1	1	1	1.569	1.636	1.946	**2.851**	1.988
RMSEmod/RMSE	u	1	1	1	1.577	1.755	1.938	2.137	**2.705**
RMSEmod/RMSE	ur	1	0	1	1.493	1.481	1.622	1.713	1.875
RMSEmod/RMSE	ur	1	1	1	1.570	1.694	1.907	1.991	1.260
RMSEmod/RMSE	urs	1	0	1	1.482	1.478	1.612	1.622	1.187
RMSEmod/RMSE	urs	1	1	1	1.570	1.608	1.542	1.814	1.198

Results for best ranking models in RMSE terms, see Table 6, in **boldface**

In the next section we apply weighted estimation to account for the level shifts in the innovation variance, which seem to have accompanied the level shifts in the mean of inflation. We shall see that this weighted estimation provides an adequate remedy for the overestimation of inflation forecast uncertainty in the 1980s and 1990s.

3 Recursive weighted ARFIMAX forecasting

So far we have not considered changes in the innovation variance. It is not unreasonable to assume that the Volcker-Greenspan regime in the 1980s and 1990s not only reduced mean, but also the (innovation) variance of inflation. This was already indicated by the ARCH-tests for the full-sample models in Table 2. In our case it is more natural to apply a heteroskedasticity test where the null of homoskedasticity is tested against an innovation variance depending on the regime since we are considering a level shift in variance. We apply a Breusch-Pagan test, which has a chi-squared distribution with 2 degrees of freedom under the null. The resulting test statistic of around 24 strongly rejects the null in all the models examined so far. Figure 7 illustrates the weights for the 3 regimes that we used to model the level shifts in variance. These weights are based on subsample estimates of the innovation variance. The observations in the first regime receive a weight 1/2, the second regime a weight 1/3 and the last regime gets weight 1. Our key assumption is that the inflation rates after 1984 are more important than previous observations to estimate the current innovation variance. In the weighted estimation we use information from the full sample and the results can therefore no longer be interpreted as the analysis of "true" out-of-sample forecasts for the whole period 1984-1999. However, the analysis indicates that for the latter part of the forecast sample, weighted estimation could have significantly improved the interval forecasts.

A more extreme solution would have been to use only the observations of the last regime, but this also would have made the recursive out-of-sample forecasting analysis of the long memory model practically impossible. In that case, recursive estimates of the RMSEmods would only be relevant for the last few years of the forecast sample. However, some eight years of observations from the beginning of the estimation sample can be deleted. We experimented with recursively deleting the observations from the first regime. The key parameters remained relatively stable, even when we used only data from 1970 onward.

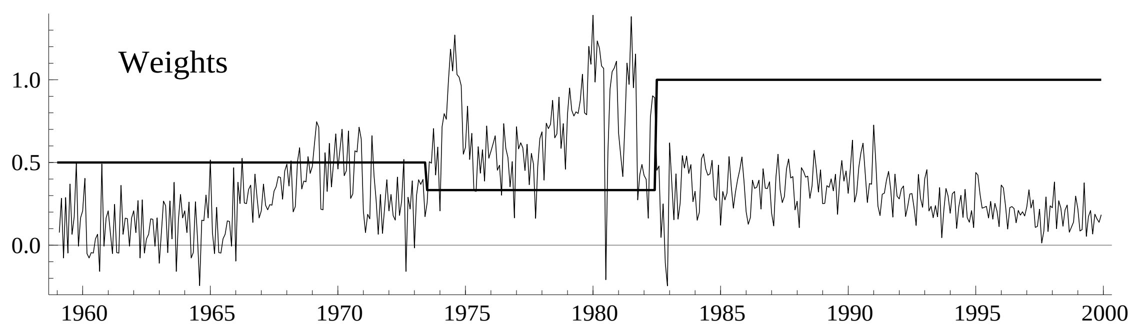

Figure 7: Weights used in weighted estimation

The introduction of the different weights has a beneficial influence on the estimated variance of the parameter estimates and on the (scaled) estimate of the innovation variance. However, the recursive estimates of the dynamic parameters, not shown here, do not change significantly by the weighting, although they naturally become more variable as the later observations have the highest weights. The weighted estimation leads to similar or lower forecast errors. Estimated standard errors are significantly lower and forecast intervals prove to be more realistic over our forecasting period. Table 8 presents the same statistics as provided in Tables 5 and 6, but allowing for weighting. The main purpose of the weighting, reducing the overall difference between RMSEmods and RMSEs, is clearly achieved, in particular for the ARFIMAX(0,d,0) model and the ARMAX(1,1) model. The RMSEmods of the ARIMAX(1,1,1) model for $H = 24$ still seem to large.

4 Recursive ARFIMAX forecast tests

In this section we evaluate the statistical performance of the recursive forecast intervals. We test for the adequacy of unconditional coverage, following Christoffersen (1998). We expect the test to reject models where the predicted forecast error variance deviates too much from the actual forecast error variance.

Table 8: *Simulated cumulative out-of-sample forecasts 1984-1999 for univariate models and regression models for core inflation, using weighted estimation*

	Model				H				
		p	d	q	1	3	6	12	24
MFE		1	0	1	0.008	0.025	0.068	0.196	0.582
MFE		1	1	1	0.007	0.023	0.058	0.167	0.482
MFE		0	d	0	0.008	0.030	0.078	0.196	0.478
MFE	r	1	0	1	0.002	**-0.001**	**0.010**	**0.029**	0.442
MFE	r	1	1	1	**0.000**	0.014	0.060	0.146	0.924
MFE	r	1	d	0	**-0.000**	0.005	0.041	0.086	0.775
MFE	r	0	d	0	**-0.000**	0.004	0.033	0.074	0.681
MFE	r	0	d	1	-0.000	0.005	0.040	0.084	0.770
MFE	s	1	0	1	0.002	0.004	**0.010**	**0.029**	0.705
MFE	u	1	0	1	0.007	0.019	0.032	0.065	**0.221**
MFE	u	1	1	1	*0.017*	0.068	0.162	0.274	0.563
MFE	urs	1	0	1	**0.000**	-0.010	0.122	0.041	0.871
MFE	urs	1	1	1	0.008	*0.071*	*0.210*	*0.397*	*1.178*
RMSE		1	0	1	0.117	0.242	0.368	0.653	1.525
RMSE		1	1	1	0.114	0.228	0.330	0.541	**1.214**
RMSE		0	d	0	0.116	0.241	0.368	0.648	1.417
RMSE	r	1	0	1	0.114	0.238	0.366	0.618	1.519
RMSE	r	1	1	1	**0.113**	**0.225**	**0.328**	0.519	1.716
RMSE	r	1	d	0	0.115	0.234	0.349	0.577	1.525
RMSE	r	0	d	0	0.114	0.233	0.350	0.582	1.525
RMSE	r	0	d	1	0.114	0.234	0.348	0.577	1.523
RMSE	s	1	1	1	0.115	0.246	0.344	**0.500**	1.573
RMSE	s	1	d	0	*0.118*	0.256	0.377	0.623	1.616
RMSE	u	1	1	1	0.115	0.229	0.345	0.552	1.246
RMSE	urs	1	0	1	0.117	*0.260*	0.388	*0.681*	1.723
RMSE	urs	1	1	1	0.114	0.255	*0.418*	0.627	*2.049*
MAPE		1	0	1	0.093	0.196	0.299	0.560	1.363
MAPE		1	1	1	0.091	0.180	0.263	0.432	**0.970**
MAPE		0	d	0	0.092	0.196	0.305	0.564	1.285
MAPE	r	1	0	1	0.092	0.194	0.296	0.511	1.373
MAPE	r	1	1	1	**0.090**	**0.179**	**0.261**	0.418	1.286
MAPE	r	1	d	0	0.092	0.192	0.283	0.487	1.320
MAPE	r	0	d	0	0.091	0.191	0.286	0.491	1.333
MAPE	r	0	d	1	0.091	0.191	0.283	0.487	1.317
MAPE	s	1	1	1	0.092	0.194	0.279	**0.405**	1.256
MAPE	u	1	1	1	0.091	0.184	0.272	0.454	1.012
MAPE	ur	1	0	1	*0.095*	0.213	*0.320*	0.551	1.402
MAPE	urs	1	0	1	*0.095*	*0.213*	0.310	*0.561*	1.485
MAPE	urs	1	d	0	*0.095*	0.208	0.306	0.504	1.504
MAPE	urs	0	d	0	*0.095*	0.207	0.310	0.512	*1.524*
MAPE	urs	0	d	1	*0.095*	0.208	0.306	0.505	1.506

Best ranking results in **boldface**. Worst ranking results in *italics*.

Table 10 reports the empirical unconditional coverage probabilities for the best forecasting models. After including the weights in the estimation, the 60 % unconditional coverage is approximately correct for the one-step forecasts of all reported models. However, the ARIMA(1,1,1) and ARIMAX(1,1,1) models lead to forecast intervals which are often too wide for multi-step forecasts. Note there are only 7 non-overlapping 2-year inflation periods in the evaluation period, so the power of tests based on the coverage probabilities for $H = 24$ is not high.

Subsequently we investigate whether there is serial correlation in the scale of the actual

Table 9: *Predicted and actual forecast root mean squared error, 1984-1999, of univariate models and regression models for core inflation, using weighted estimation*

		Model			H				
		p	d	q	1	3	6	12	24
RMSEmod		1	0	1	0.119	0.252	0.445	0.839	1.482
RMSEmod		1	1	1	0.119	0.263	0.458	0.881	**1.931**
RMSEmod		0	d	0	0.119	0.268	0.460	0.809	1.444
RMSEmod	r	1	0	1	0.115	0.246	0.443	0.823	1.757
RMSEmod	r	1	1	1	**0.116**	**0.258**	**0.456**	0.877	1.904
RMSEmod	r	1	d	0	0.115	0.254	0.449	0.804	1.576
RMSEmod	r	0	d	0	0.115	0.255	0.448	0.793	1.501
RMSEmod	r	0	d	1	0.115	0.254	0.449	0.802	1.573
RMSEmod	s	1	1	1	0.117	0.252	0.435	**0.845**	1.910
RMSEmod	u	1	1	1	0.118	0.255	0.429	0.829	1.893
RMSEmod/RMSE		1	0	1	1.017	1.041	1.244	1.285	1.029
RMSEmod/RMSE		1	1	1	1.044	1.154	1.387	1.628	**1.591**
RMSEmod/RMSE		0	d	0	1.026	1.112	1.250	1.248	1.019
RMSEmod/RMSE	r	1	0	1	**1.006**	**1.036**	**1.212**	1.331	1.156
RMSEmod/RMSE	r	1	1	1	1.024	1.147	1.388	1.688	1.110
RMSEmod/RMSE	r	1	d	0	1.003	1.084	1.288	1.393	1.033
RMSEmod/RMSE	r	0	d	0	1.011	1.095	1.281	1.363	0.984
RMSEmod/RMSE	r	0	d	1	1.004	1.086	1.288	1.390	1.033
RMSEmod/RMSE	r	1	1	1	1.018	1.065	1.306	1.675	1.247
RMSEmod/RMSE	s	1	1	1	1.012	1.026	1.267	**1.691**	1.214
RMSEmod/RMSE	u	1	1	1	1.024	1.117	1.242	1.500	1.519

Results for best ranking models in RMSE terms, see Table 8, in **boldface**

Table 10: *Unconditional coverage test of 60% forecast intervals, core inflation, using weighted estimation*

	Model			H				
	p	d	q	1	3	6	12	24
	1	0	1	0.58	0.75*	0.71	0.60	0.29
	1	1	1	0.63	0.81*	0.90*	0.80	0.86
	1	d	0	0.58	0.76*	0.77*	0.60	0.57
	0	d	0	0.59	0.76*	0.74	0.60	0.43
r	1	0	1	0.56	0.71	0.71	0.60	0.43
r	1	1	1	0.57	0.78*	0.94*	0.87*	0.71
r	1	d	0	0.57	0.71	0.77*	0.60	0.43
r	0	d	0	0.57	0.71	0.71	0.60	0.43
s	1	1	1	0.57	0.73*	0.84*	0.87*	0.71
u	1	1	1	0.59	0.73*	0.84*	0.87*	0.86
L_c				0.530	0.478	0.425	0.349	0.236
U_c				0.669	0.718	0.766	0.834	0.925
N				191	63	31	15	7

Empirical coverage probabilities which differ significantly (at 95% confidence level) from the 60% theoretical coverage indicate by an asterisk. Bottom rows indicate the bounds of the acceptance region. N is the number of nonoverlapping forecast intervals, 1984-1999

forecast error distribution using the independence test as suggested by Christoffersen (1998). We expect power for this test against persistent changes in the forecast error variance over the evaluation period.

The results of the independence test are given in Table 11. Reported are the likelihood ratio test statistics, which have $\chi^2(1)$ limit distribution. Since the forecasts are correlated for horizons $H > 1$ we use only one out of every H forecast errors in the independence test. None of the independence tests rejects.

Table 11: *Independence test of coverage of 60% forecast intervals, core inflation, using weights*

	Model			H				
	p	d	q	1	3	6	12	24
	1	0	1	0.40	0.05	.	0.47	.
	1	1	1	0.01	0.35	.	0.50	.
	1	d	0	0.40	0.23	.	0.47	.
	0	d	0	1.07	0.23	0.10	0.47	.
r	1	0	1	0.95	0.39	0.20	0.30	.
r	1	1	1	0.30	0.07	.	2.14	.
r	1	d	0	0.13	0.39	.	0.30	.
r	0	d	0	0.01	0.39	0.20	0.30	.
s	1	1	1	0.45	0.75	0.39	.	0.59
u	1	1	1	1.07	0.20	.	2.14	.

Likelihood ratio test on independence. which asymptotically is $\chi^2(1)$ distributed, see Christoffersen (1998). Included in the test are only independent forecasts at a distance of H months. A . indicates there are not enough relevant observations: $n_{01} = 0$ or $n_{10} = 0$.

Finally we examine the possibility of improving the forecasts by using combinations of models in a so-called forecast encompassing framework, see e.g. Harvey, Leybourne & Newbold (1998) for a recent review and Harvey & Newbold (2000) for the extension to multivariate forecast encompassing. West (2001) discusses forecast encompassing in the context of forecasts from recursive regressions. The basic idea is to construct a combined forecast f_{ct} as a weighted average of the forecasts of a baseline Model 1, f_{1t} and the forecasts of other non-nested Models 2 and 3, say f_{2t} and f_{3t}, that is

$$f_{ct} = (1 - \lambda_2 - \lambda_3)f_{1t} + \lambda_2 f_{2t} + \lambda_3 f_{3t}, \quad 0 \leq \lambda_i \leq 1 \tag{2}$$

For testing purposes one rewrites (2) as follows:

$$e_{1t} = c + \lambda_2(e_{1t} - e_{2t}) + \lambda_3(e_{1t} - e_{3t}) + u_t \tag{3}$$

with $e_{it} = y_t - f_{it}$ and $u_t = y_t - f_{ct}$. The constant c equals zero if all forecasts are unbiased. This assumption does not always hold, so it is good practice to add a constant if one employs (3) as a test regression.

When λ_1 and λ_2 nearly add up to unity, this indicates that the alternative forecasts perform better. When all forecast errors are zero mean Gaussian without serial correlation, standard statistical regression theory applies, at least asymptotically. This analysis provides an easy-to-compute statistical measure of the relative forecasting performance of the models under scrutiny. The procedure provides an extra measure of the usefulness of the different explanatory variables, where not only the in-sample fit, but also the stability of the explanatory power over the period 1984-1999 plays a role.

The first rows of Table 12 provide bilateral model comparisons, where models with the best leading indicator, r_{t-1}, are chosen as benchmark Model 1. Each row in the table corresponds to one (multiple) forecast encompassing test. It appears from test no. 6 that the ARIMAX(1,1,1) model with u_{t-1} can help to improve the forecasts of the ARFIMAX(0,d,0) with r_{t-1}, but the p-value of this test is not very low. The last rows of Table 12 show that the forecasts of ARFIMAX models with r_{t-1} cannot be improved by combining them with corresponding ARFIMAX models with other explanatory variables: none of the estimated λ_is differs significantly from zero.

Table 12: *Multivariate one-step-ahead forecast encompassing tests*

no.	Model 1		Model 2		Model 3		F	p-F
1	r	0,d,0	r 1,0,1					
			0.17	[0.57]			0.32	0.57
2	r	0,d,0	s 1,0,1					
			0.16	[0.61]			0.37	0.54
3	r	0,d,0	u 1,0,1					
			0.17	[0.59]			0.35	0.55
4	r	0,d,0	r 1,1,1					
			0.76	[1.74]			3.01	0.08
5	r	0,d,0	s 1,1,1					
			0.40	[1.78]			3.16	0.08
6	r	0,d,0	u 1,1,1					
			0.52	[1.93]			3.74	0.05
7	r	0,d,0	s 0,d,0					
			0.08	[0.30]			0.09	0.76
8	r	0,d,0	u 0,d,0					
			0.15	[0.49]			0.24	0.62
9	r	1,0,1	u 1,0,1		s 1,0,1			
			0.08	[0.20]	0.13	[0.38]	0.22	0.80
10	r	1,1,1	u 1,1,1		s 1,1,1			
			0.33	[1.07]	0.18	[0.66]	1.31	0.27
11	r	1,d,1	u 1,d,1		s 1,d,1			
			0.12	[0.28]	0.01	[0.03]	0.07	0.93
12	r	0,d,0	u 0,d,0		s 0,d,0			
			0.17	[0.39]	−0.02	[−0.06]	0.12	0.88

The columns under Model 2 and Model 3 give the parameter estimates of λ_2 and λ_3 in (3), with corresponding t-values in brackets. Models are indicated by their single leading indicator and the orders p,d,q. Results compare forecasts of Model 2 and Model 3 with the baseline Model 1. F-statistics for $\lambda_2 (= \lambda_3) = 0$ and corresponding p-values presented under F and p-F .

5 Conclusion

Statistical analysis shows that US postwar inflation is long memory, with an order of integration of around 0.3, even after allowing for a structural shift in the mean and variance to capture the high inflation period in the 1970s. Recursive estimation shows that the order of integration has remained quite stable. Statistical analysis of dynamic regression models for inflation conditioning on lags of unemployment and interest rates shows a stable effect of short lags of short term interest rates. The errors of the regression models are still long memory. An ARMA(1,1) model and an ARIMA(1,1,1) model provide a similar in-sample fit as the ARFIMA(0,d,0) model.

We performed a recursive out-of-sample forecasting exercise for the period 1984-1999 for cumulative inflation forecasting up to a two year horizon, using univariate models and long memory regression models. The ARIMA(1,1,1) model performs better than the ARFIMA(0,d,0) model regarding the precision of point forecasts. The introduction of conditioning variables improves forecasting precision at short horizons. With regard to forecast interval estimation, downweighting the observations in the 1970s turns out to be essential in order to get realistic intervals for inflation in the 1990s. The empirical multi-step prediction intervals for the ARIMA(1,1,1) model with or without explanatory variables are too wide, also when weighting is applied.

References

Adenstedt, R. K. (1974), 'On large-sample estimation for the mean of a stationary random sequence', *The Annals of Statistics* **2**, 1095–1107.

Ball, L. & Mankiw, N. G. (1995), 'Relative-price changes as aggregate supply shocks', *The Quarterly Journal of Economics* **110**(1), 161–194.

Beran, J. (1994), 'On a class of M-estimators for Gaussian long-memory models', *Biometrika* **81**(4), 755–766.

Bos, C. S., Franses, P. H. & Ooms, M. (1999), 'Long memory and level shifts: Re-analyzing inflation rates', *Empirical Economics* **24**, 427–449.

Brodsky, J. & Hurvich, C. M. (1999), 'Multi-step forecasting for long-memory processes', *Journal of Forecasting* **18**, 59–75.

Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International Economic Review* **39**(4), 841–862.

Doornik, J. A. (1999), *Object-Oriented Matrix Programming using Ox*, 3rd edn, London: Timberlake Consultants Ltd. See `http://www.nuff.ox.ac.uk/Users/Doornik`.

Doornik, J. A. & Hendry, D. F. (2001), *GiveWin: An Interface to Empirical Modelling*, 3rd edn, Timberlake Consultants Press, London.

Doornik, J. A. & Ooms, M. (1999), A package for estimating, forecasting and simulating Arfima models: Arfima package 1.0 for Ox, Technical report, Nuffield College, Oxford, UK.

Gali, J. & Gertler, M. (2000), Inflation dynamics: A structural econometric analysis, Technical Report 7551, National Bureau of Economic Research.

Granger, C. W. J. & Joyeux, R. (1980), 'An introduction to long-memory time series models and fractional differencing', *Journal of Time Series Analysis* **1**, 15–29.

Harvey, D. I. & Newbold, P. (2000), 'Tests for multiple forecast encompassing', *Journal of Applied Econometrics* **15**(5), 471–482.

Harvey, D. I., Leybourne, S. J. & Newbold, P. (1998), 'Tests for forecast encompassing', *Journal of Business & Economic Statistics* **16**, 254–259.

Hassler, U. & Wolters, J. (1995), 'Long memory in inflation rates: International evidence', *Journal of Business and Economic Statistics* **13**, 37–46.

Hooker, M. A. (1999), Are oil shocks inflationary? asymmetric and nonlinear specifications versus changes in regime, Technical Report 1999-65, Board of Governors of the Federal Reserve System (U.S.).

Hosking, J. R. M. (1981), ‘Fractional differencing’, *Biometrika* **68**(1), 165–176.

Lamoureux, C. & Lastrapes, W. (1990), ‘Persistence-in-variance, structural change and the GARCH model’, *Journal of Business and Economic Statistics* **8**, 225–234.

Ooms, M. & Hassler, U. (1997), ‘On the effect of seasonal adjustment on the log-periodogram regression’, *Economics Letters* **56**, 135–141.

Sowell, F. (1992), ‘Maximum likelihood estimation of stationary univariate fractionally integrated time series models’, *Journal of Econometrics* **53**, 165–188.

Stock, J. H. & Watson, M. W. (1999), ‘Forecasting inflation’, *Journal of Monetary Economics* **44**, 293–335.

West, K. D. (2001), ‘Tests for forecast encompassing when forecasts depend on estimated regression parameters’, *Journal of Business and Economic Statistics* **19**, 29–33.