EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1363 / JULY 2011

FORECASTING INFLATION WITH GRADUAL REGIME SHIFTS AND EXOGENOUS INFORMATION

by Andrés González,
Kirstin Hubrich and
Timo Teräsvirta

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1363 / JULY 2011

FORECASTING INFLATION WITH GRADUAL REGIME SHIFTS AND EXOGENOUS INFORMATION

by Andrés González[1], Kirstin Hubrich[2] and Timo Teräsvirta[3]

In 2011 all ECB publications feature a motif taken from the €100 banknote.

1 Banco de la República, Bogotá, Colombia.

2 European Central Bank, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany; e-mail: kirstin.hubrich@ecb.europa.eu

3 CREATES, Aarhus University, Denmark.

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

ISSN 1725-2806 (online)

CONTENTS

Acknowledgements. This research has been supported by the Banco de la República, European Central Bank and the Danish National Research Council. Material from this paper has been presented at the conference 'Forecasting in Rio', July 2008, the Applied Economics Ruby Conference, Cambridge, September 2008, the ETSERN Workshop, Copenhagen, and the CIMF Workshop 'Forecasting under Model Instability', Trinity College, Cambridge, both in November 2008, the Deutsche Bundesbank conference 'Forecasting and Monetary Policy', Berlin, March 2009, the conference on 'Recent Developments in Macroeconomics', Mannheim, July, 2009, the Econometric Society European Meeting, Barcelona, August 2009, the Canadian Econometric Society Study Group Meeting, October 2009, the NBER-NSF Time Series Conference, October 2010, and in seminars at University of Michigan, University of New South Wales, Federal Reserve Bank of Atlanta, Beijing International Studies University and University of Adelaide. Comments from participants, Lutz Kilian and Christian Schumacher in particular, are gratefully acknowledged. We would also like to thank Ron Smith for useful discussions, and Marco Buchmann, Kerstin Holzheu and Norberto Rodriguez for valuable research assistance. The responsibility for any errors and shortcomings in this work remains ours. The views expressed in this paper are the authors' own and do not necessarily reflect those of the European Central Bank or Banco de la República.

Abstract

We propose a new method for medium-term forecasting using exogenous information. We first show how a shifting-mean autoregressive model can be used to describe characteristic features in inflation series. This implies that we decompose the inflation process into a slowly moving nonstationary component and dynamic short-run fluctuations around it. An important feature of our model is that it provides a way of combining the information in the sample and exogenous information about the quantity to be forecast. This makes it possible to form a single model-based inflation forecast that also incorporates the exogenous information. We demonstrate, both theoretically and by simulations, how this is done by using the penalised likelihood for estimating the model parameters. In forecasting inflation, the central bank inflation target, if it exists, is a natural example of such exogenous information. We illustrate the application of our method by an out-of-sample forecasting experiment for euro area and UK inflation. We find that for euro area inflation taking the exogenous information into account improves the forecasting accuracy compared to that of a number of relevant benchmark models but this is not so for the UK. Explanations to these outcomes are discussed.

Keywords: Nonlinear forecast; nonlinear model; nonlinear trend; penalised likelihood; structural shift; time-varying parameter

JEL Classification Codes: C22; C52; C53; E31; E47

Non-technical Summary

There has been increased interest in both characterizing and forecasting inflation using models that capture structural change, in particular in the light of the recent crisis. In this study we propose a new method for generating medium-term forecasts incorporating quantitative exogenous information available about the future of the variable of interest. We show how a flexible nonstationary parametric model can be used to both characterise inflation and generate medium-term forecasts making use of quantitative exogenous information about future inflation.

Parameters of a time series model for inflation may change over time for a variety of reasons. For example, changes in monetary policy regimes may affect the parameters of the model. A typical assumption in this context has been that the regime changes are abrupt. This implies that the effects of monetary policy changes are immediate and that the new regime is stable until there is another break in the model. However, it is often useful to think of these parameter changes as occurring gradually over a period of time. The shift in preferences towards strong price stability, reflected in the downward trend of euro area inflation during the 1980s, may be viewed as an example of this type. The downward shift in US inflation under Volcker constitutes another prominent example.

There are many ways of incorporating the idea of smooth continuous changes in parameters into an inflation model. In this paper, we assume that the inflation process has a gradually shifting mean, and the fluctuation of the process around this mean is described by an autoregressive process. This leads to the Shifting Mean Autoregressive (SM-AR) model, in which the inflation process is assumed to contain two components: a deterministically time-varying mean and an autoregressive component that is stationary around the mean. We show how this model can be used to analyse inflation and we propose a new method for medium-term forecasting using exogenous information based on this model.

Our model of time-varying mean inflation is well suited for tracking the developments in headline inflation that will persist in the medium term. Transient features due to temporary shocks to the economy are explained by the autoregressive structure of the model. We provide a timely measure of medium-term inflation based on a single time series. We show this by fitting the SM-AR model to euro area, UK and US inflation. This measure can also be useful if one wants to compare medium-term shifts in inflation between countries.

Another new feature of our model is that it allows incorporating exogenous information into inflation forecasts from this model within a classical framework. We propose a framework for medium-term forecasting and show, both theoretically and by simulation, how exogenous information, when available, can be included in medium-term forecasts in our framework. This is done by penalizing the log-likelihood function in the estimation of model parameters. In forecasting inflation, the central bank inflation target, or any other quantitative definition of price stability, is a natural example of such information. Since central banks that have committed to a quantitative aim of price stability, for instance in the form of an inflation target, aim at keeping inflation close to the target value, at least over the medium-term the target contains information that should be incorporated in the forecasts. In our model, a

penalty term determines the weight of the exogenous information. It reflects the forecaster's subjective assessment of the commitment of the Central Bank to the target and chances of success of its monetary policy.

We apply our procedure to forecasting the euro area as well as the UK inflation rate. In the former case the exogenous information comprises the definition of price stability of the European Central Bank (ECB), whereas the inflation target of the Bank of England plays the same role in the latter. We find that in forecasting euro area inflation taking this exogenous information into account does improve the medium-term forecast accuracy over that of a number of relevant benchmark models.

This finding is particularly interesting given that our sample includes the period of high volatility of inflation during the recent financial crisis. The usefulness of our model for medium-term forecasting is not limited to improvements in the accuracy of point forecasts. The forecasting process provides a whole density forecast whose shape is a function of the weight the forecaster allocates to the exogenous information.

1 Introduction

There has been increased interest recently in both characterising and forecasting inflation using models that involve structural change. In this study we propose a new method for generating medium-term forecasts incorporating quantitative exogenous information available about the future of the variable of interest. We show how a flexible nonstationary parametric model can be used to both characterise inflation and generate medium-term forecasts making use of quantitative exogenous information about future inflation.

Parameters of a time series model for inflation may change over time for a variety of reasons. For example, changes in monetary policy regimes may affect the parameters of the model.[1] A typical assumption in this context has been that the regime changes are abrupt. This implies that the effects of monetary policy changes are immediate and that the new regime is stable until there is another break in the model. However, it is often useful to think of these parameter changes as occurring gradually over a period of time. The shift in preferences towards strong price stability, reflected in the downward trend of euro area inflation during the 1980s, may be viewed as an example of this type. The downward shift in US inflation under Volcker constitutes another prominent example.

There are many ways of incorporating the idea of smooth continuous changes in parameters into an inflation model. In this paper, we assume that the inflation process has a gradually shifting mean, and the fluctuation of the process around this mean is described by an autoregressive process. This leads to the Shifting Mean Autoregressive (SM-AR) model, in which the inflation process is assumed to contain two components: a deterministically time-varying mean and an autoregressive component that is stationary around the mean. The shifting mean may then be interpreted as a measure of the implicit inflation target of the central bank.[2] It can also be viewed as a proxy for unobservable variables or other driving forces that are difficult or even impossible to quantify in a satisfactory manner. Examples include the decline in inflation due to increasing international consensus in monetary policy aiming at price stability after high and volatile inflation during the 1970s, or increasing globalisation that has led to intensified competition. The time-varying mean may also be considered a measure of the underlying trend in inflation that is often referred to as 'core inflation'.[3]

[1] See e.g. Stock and Watson (2007), Sims and Zha (2006), Schorfheide (2005), Lendvai (2006) and Pesaran, Pettenuzzo and Timmermann (2006).

[2] See e.g. Orphanides and Williams (2005) and Kozicki and Tinsley (2005).

[3] See e.g. Cogley (2002), Clark (2001) and Cristadoro, Forni, Reichlin and Veronese (2005) for recent suggestions and/or discussions of core inflation measures.

Our model of time-varying mean inflation is well suited for tracking the developments in headline inflation that will persist in the medium term. Transient features due to temporary shocks to the economy are explained by the autoregressive structure of the model. We provide a timely measure of medium-term inflation based on a single time series. This measure can also be useful if one wants to compare medium-term shifts in inflation between countries.

Another new feature of our model is that it allows incorporating exogenous information into inflation forecasts from this model within a classical framework. Manganelli (2009) recently suggested another way of doing that. We propose a framework for medium-term forecasting and show, both theoretically and by simulation, how exogenous information, when available, can be included in medium-term forecasts in our framework. This is done by penalising the log-likelihood function in the estimation of model parameters. In forecasting inflation, the central bank inflation target, or any other quantitative definition of price stability, is a natural example of such information. Beechey and Österholm (2010) considered this idea in a Bayesian framework. We apply our procedure to forecasting the euro area as well as the UK inflation rate. In the former case the exogenous information comprises the definition of price stability of the European Central Bank (ECB), whereas the inflation target of the Bank of England plays the same role in the latter. We find that in forecasting euro area inflation taking this exogenous information into account does improve the medium-term forecast accuracy over that of a number of relevant benchmark models.

This finding is particularly interesting given that our sample includes the period of high volatility of inflation during the recent financial crisis. The usefulness of our model for medium-term forecasting is not limited to improvements in the accuracy of point forecasts. The forecasting process gives us a whole density forecast whose shape is a function of the weight the forecaster allocates to the exogenous information.

The plan of the paper is the following: The SM-AR model and outlines of modelling are presented in Section 2. Empirical results for modelling the euro area, UK and US inflation appear in Section 3. In Section 4 it is shown how sample information and exogenous information can be combined into a single (density) forecast using the SM-AR model. Section 5 contains results from a pseudo out-of-sample forecasting experiment in which medium-term forecasts from our SM-AR model are compared with forecasts from a number of benchmark models. Our conclusions can be found in Section 6.

2 A framework for modelling inflation

2.1 An autoregressive model with a shifting mean

The modelling and forecasting tool in this work is the autoregressive model with a shifting mean, the SM-AR model. The shift is a smooth deterministic function of time, which implies assuming inflation to be a nonstationary process. The SM-AR model of order p has the following definition, see González and Teräsvirta (2008):

$$y_t = \delta(t) + \sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t \tag{1}$$

where the roots of the lag polynomial $1 - \sum_{j=1}^{p} \phi_j \mathsf{L}^j$ lie outside the unit circle, L is the lag operator: $\mathsf{L}x_t = x_{t-1}$. As all roots of the lag polynomial lie outside the unit circle, $\{y_t\}$ is stationary around the shifting mean. The errors ε_t form a sequence of independent, identically $(0, \sigma^2)$ distributed random variables, and $\delta(t)$ is a bounded deterministic nonlinear shift function or shifting intercept. In parameter estimation and statistical inference it is assumed that the error distribution is normal.

In empirical work, $\delta(t)$ is often a linear function of t, in which case y_t in (1) is called 'trend-stationary'. Contrary to this, González and Teräsvirta (2008) define $\delta(t)$ as a bounded function of time:

$$\delta(t) = \delta_0 + \sum_{i=1}^{q} \delta_i g(\gamma_i, c_i, t/T) \tag{2}$$

where δ_i, $i = 1, \ldots, q$, are parameters, T is the number of observations, and $g(\gamma_i, c_i, t/T)$, $i = 1, \ldots, q$, are logistic transition functions or sigmoids:

$$g(\gamma_i, c_i, t/T) = \left(1 + \exp\{-\gamma_i(t/T - c_i)\}\right)^{-1} \tag{3}$$

with $\gamma_i > 0$, $i = 1, \ldots, q$. The components in the shift function (2) are exchangeable, and identification is achieved for example by assuming $c_1 < \cdots < c_q$.

The parametric form of (2) is very flexible and contains as special cases well known examples of nonlinear functions. For instance, when $\delta_1 = \cdots = \delta_q = 0$, (2) becomes constant, and when $q = 1$, $\delta(t)$ changes smoothly from δ_0 to $\delta_0 + \delta_1$ as a function of t, with the centre of the change at $t = c_1 T$. The smoothness of the change is controlled by γ_1: the larger γ_1, the faster the transition. When $\gamma_1 \to \infty$, $\delta(t)$ collapses into a step function, so there is a single break in the intercept. On the contrary, when γ_1 is close to zero,

$\delta(t)$ represents a slow monotonic shift that is approximately linear around c. Values $q > 1$ add flexibility to $\delta(t)$ by making nonmonotonic shifts possible.

More generally, $\delta(t)$ is a so-called universal approximator. Suppose $y_t = f(t)$, that is, there exists a functional relationship between y and t. Then, under mild regularity conditions for f, the relationship is arbitrarily accurately approximated by replacing $f(t)$ by (2) where $q \leq q_0 < \infty$, see, for example, Cybenko (1989), Funahashi (1989) or Hornik, Stinchcombe and White (1989). One could also use a completely nonparametric function as in Priestley and Chao (1972) and Benedetti (1977), but the linear combination of sigmoids (2) as in neural network models appears more suitable for our forecasting problem. From (1) it follows that the time-varying mean of the process equals

$$\mathsf{E}_t y_t = (1 - \sum_{j=1}^{p} \phi_j \mathsf{L}^j)^{-1} \delta(t).$$

2.2 Model specification and estimation

The specific form of the SM-AR model has to be determined from the data. This implies selecting p and q, which will be done by using statistical inference. There is no natural order in which the choice is made. Priority may be given to selecting q first if the emphasis lies on specifying a model with a shifting mean. For example, if one is modelling the developments in the 1980's and wants to proxy the unobservable tendencies by time instead of including them in the autoregressive component of the model, one may want to select q first. Some techniques of modelling structural change by breaks use an analogous order: the break-points are determined first, and the dynamic structure of the regimes thereafter. The decision is left to the model builder. Nevertheless, when q is selected first, one may use a heteroskedasticity-autocorrelation consistent (HAC) estimator for the covariance matrix of the estimators throughout the selection process and thus account for the fact that there is autoregressive structure around the mean. This is the case in the applications of Section 3.

In this work we apply a procedure for selecting q that González and Teräsvirta (2008) call QuickShift. It has two useful properties. First, it transforms the model selection problem into a problem of selecting variables, which simplifies the computations. Second, overfitting is avoided. QuickShift is a modification of QuickNet, a recent method White (2006) developed for building and estimating artificial neural network models. The functioning of QuickShift is described in Appendix A. One could also apply Autometrics (Doornik 2008, 2009) or the Marginal Bridge Estimator (Huang, Horowitz and Ma, 2008) to this specification problem, see Kock and Teräsvirta (2011) for a related example, but it has not been done here.

Full maximum likelihood estimation of parameters of the SM-AR model including γ_i and c_i, $i = 1, \ldots, q$, may not be necessary, because QuickShift in general provides good approximations to maximum likelihood estimates when the grid is sufficiently dense. Nevertheless, if one wants to continue, a derivative-based algorithm with a short initial step-length should thus be sufficient to maximize the log-likelihood. Should there be numerical problems, however, they may be solved by applying a global optimization algorithm such as simulated annealing (with a rather low initial temperature) or a genetic algorithm and using the vector of parameters $(\boldsymbol{\gamma}', \mathbf{c}')'$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$ and $\mathbf{c} = (c_1, \ldots, c_q)'$ are selected by QuickShift, as initial values. The maximum likelihood estimators of the parameters of the SM-AR model are consistent and asymptotically normal. The proofs require time to be rescaled to the interval $(0, 1]$. They can be found In Appendix B.

This approach may be compared to filtering. In some cases filtering a trend component from a series using a filter such as the one by Leser (1961) (often called the Hodrick-Prescott filter), may lead to results similar to ones obtained by modelling the shifting mean using QuickShift. An essential difference between the filtering and our approach is, however, that the latter is completely parametric, and modelling the shifting mean and the dynamics around it can be done simultaneously. Another difference is that, contrary to extrapolating filtered series, forecasting with the SM-AR model is a straightforward exercise. It should be pointed out, however, that the SM-AR model is not a feasible tool for very short-term forecasting because of its lack of adaptability. It is, however, well-suited for medium-term forecasting when extraneous information, for example in the form of a central bank inflation target, is available. This will be discussed in Section 4.1.

2.3 Other approaches to modelling inflation

The SM-AR model is an example of a time-varying parameter model, but there are others. For example, one may assume that parameter variation is stochastic; for various types of the stochastic-parameter model see Teräsvirta, Tjøstheim and Granger (2010, Sections 3.10–11). Recently, Stock and Watson (2007) characterised the US inflation with a model based on decomposing the inflation series into two stochastic unobserved components. With constant parameters, the model is simply an ARIMA(0,1,1) model. Parameter variation is introduced by letting the variances of the two unobserved components be nonconstant over time. They are assumed to follow a stochastic volatility model, that is, their logarithms are generated by a first-order autoregressive process, which in this case is a pure random walk. The first one of the two stochastic unobserved components represents the 'trend' or the gradually shifting component of inflation, whereas the second contains all short-run fluctuations.[4] In the SM-AR model the shift component of inflation is deterministic and there is short-run random variation around it.

Other methods to model time-variation in inflation with the aim of describing medium-term inflation developments and changes in the inflation target, include Kozicki and Tinsley (2005) who estimated a model with a VAR-like structure that allowed for shifts in the inflation target and imperfect policy credibility. Kozicki and Tinsley (2006) provided a measure of the private sector's perception of the inflation target of monetary policy and found that it has shifted considerably over time. However, both papers by Kozicki and Tinsley are not concerned with forecasting, as we are in this paper. An early paper related to ours is Cogley (2002) who proposed a simple exponential smoothing to derive a 'core' (or 'underlying' or 'medium-term') inflation measure. Further references to papers that aim at forecasting inflation using time-varying parameter models are discussed in Section 4.2.2.

3 Modelling gradual shifts in inflation

In this section we show how the SM-AR model can be used to model medium-term developments in headline inflation. Our shifting mean inflation measure can be interpreted as an 'underlying' or 'core' measure of inflation, and we shall show that its response to temporary shocks is limited. The recent period of volatile inflation beginning in mid-2007, mainly due to large changes in energy and food inflation, is a case in point.

3.1 Data

The series representing euro area inflation is the seasonally adjusted monthly Harmonised Index of Consumer Prices (HICP). We also estimate SM-AR models for the monthly CPI inflation for the UK and the US based on monthly year-on-year inflation series. What makes modelling and forecasting inflation of the euro area and the UK particularly interesting is the fact that the European Central Bank (ECB) provides an explicit formulation for its aim of price stability, and the Bank of England is one of the inflation targeting central banks. The time series for the euro area covers the period from

[4]Similar ideas of allowing for a shifting trend inflation process modelled as a driftless random walk without or with stochastic volatility in parameter innovations can be found in Cogley and Sbordone (2008) and Cogley, Primiceri and Sargent (2010).

1981(1) to 2010(6). It consists of annual differences of the monthly, seasonally adjusted and backdated Harmonised Index of Consumer Prices, in which fixed euro conversion rates have been used as weights when backdating. The availability of aggregated backdata for the euro area and the launch of the European Monetary System in 1979 determine the beginning of the series. Both the UK and the US year-on-year inflation series begin 1981(1) and end 2010(6). They comprise annual differences of the monthly Consumer Price Index (CPI). The euro area series is provided by the ECB and the other two by OECD.

It should be noted that in December 2003 the Bank of England changed the series according to which the inflation target is defined. The current target is 2% year-on-year measured by the CPI, which is just the name of the HICP in the UK. As already mentioned, this is the series we shall use here.

3.2 Euro area inflation

The euro area inflation series 1981(1)–2010(6) can be found in Figure 1 (the solid curve). In selecting the number of transitions, the original significance level $\alpha_0 = 0.5$, and the remaining ones equal $\alpha_q = 0.5\alpha_{q-1}$, $q \geq 1$. Assuming $p = 0$ in (1) , QuickShift and parameter estimation yield the following result:

$$\begin{aligned}\widehat{\delta}(t) = \underset{(0.078)}{10.74} &- \underset{(0.17)}{8.90}(1+\exp\{-\underset{(-)}{7.54}(t/T-\underset{(-)}{0.11})\})^{-1} \\ &- \underset{(0.28)}{2.22}(1+\exp\{-\underset{(-)}{17.3}(t/T-\underset{(-)}{0.51})\})^{-1} \\ &- \underset{(0.15)}{1.90}(1+\exp\{-\underset{(-)}{30}(t/T-\underset{(-)}{0.96})\})^{-1} \\ &+ \underset{(0.33)}{1.73}(1+\exp\{-\underset{(-)}{30}(t/T-\underset{(-)}{0.27})\})^{-1} \\ &+ \underset{(0.11)}{0.99}(1+\exp\{-\underset{(-)}{30}(t/T-\underset{(-)}{0.67})\})^{-1}\end{aligned} \tag{4}$$

The standard deviation estimates are heteroskedasticity-autoregression robust ones. Since γ_i and c_i, $i = 1, ..., 5$, are 'estimated' by QuickShift, no standard deviation estimates are attached to their estimates.

The maximum value of γ in the grid equals 30, and this limit is reached three times. The estimated switching mean also appears in Figure 1 (the dashed curve). The transitions in (4) appear in the order they are selected by QuickShift. The first transition describes the prolonged decrease in inflation in the first half of the 1980s and reflects the increased preference for high price stability in all European countries: note the negative estimate $\widehat{\delta}_1 = -8.90$. The second one accounts for another downturn in the mid-1990s, whereas

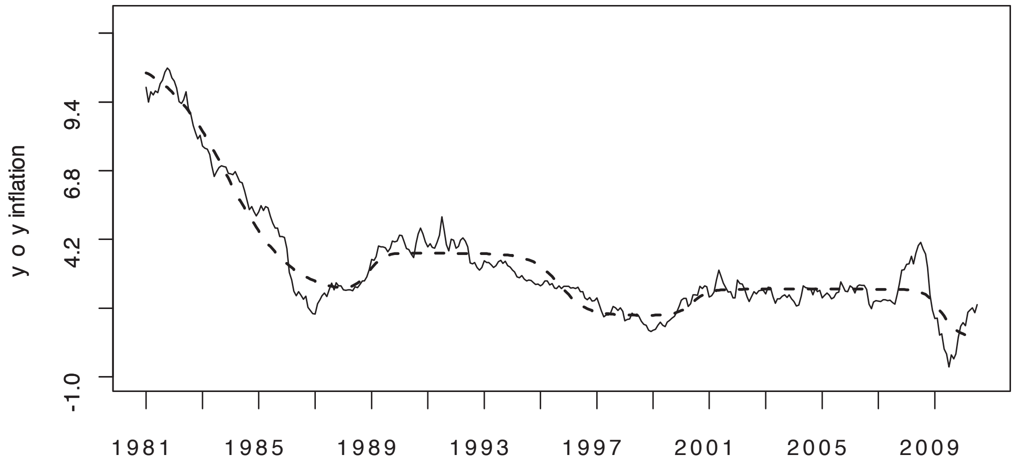

Figure 1: The euro area year-on-year inflation rate 1981(1)–2010(6) (solid curve) and the shifting mean from the SM-AR model (4) (dashed curve)

the fourth one describes the increase at the end of the 1980s ($\widehat{\delta}_4 = +1.73$). The increase following the introduction of the euro is captured by the fifth transition. The very recent increase in inflation does not affect the estimate of $\delta(t)$, but the subsequent steep decrease does. It is characterised by the third transition ($\widehat{\delta}_3 = -1.90$). Here our SM-AR model only indicates a limited response to short-term fluctuations in all items inflation that in this case may be caused by strong movements in energy and food inflation. We therefore interpret the shifting mean as a measure of 'underlying' or 'core' inflation. The final level of the shifting mean equals $\sum_{j=0}^{5} \widehat{\delta}_j = 0.43$.

3.3 UK inflation

The monthly year-on-year UK inflation series from 1981(1) to 2010(6) is graphed in Figure 2 together with the shifting intercept from an estimated SM-AR model. The model has $p = 0$, and the shifting mean has the following form:

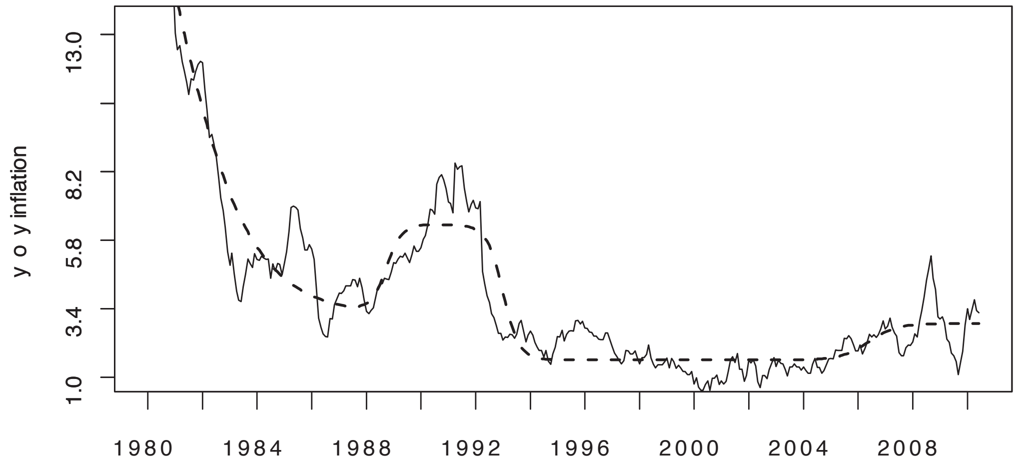

Figure 2: The UK year-on-year inflation rate 1981(1)–2010(6) (solid curve) and the shifting mean from the SM-AR model (4) (dashed curve)

$$
\begin{aligned}
\delta(t) \quad = \quad & \underset{(1.18)}{33.45} - \underset{(1.48)}{30.27}(1+\exp\{-5.72(t/T-0.01)\})^{-1} \\
& -\underset{(0.38)}{4.70}(1+\exp\{-30(t/T-0.43)\})^{-1} \\
& +\underset{(0.53)}{3.13}(1+\exp\{-30(t/T-0.29)\})^{-1} \\
& +\underset{(0.27)}{1.27}(1+\exp\{-30(t/T-0.87)\})^{-1}. \qquad (5)
\end{aligned}
$$

As is seen from (5), four transitions are needed to characterise the shifting mean of the UK series; see also Figure 2. The role of the first one is to describe the decrease in inflation in the 1980s. Note the low estimate of the location parameter: $\widehat{c}_1 = 0.01$ and the high $\widehat{\delta}_0$ and low $\widehat{\delta}_1$. They are due to the fact that less than one half of the logistic function is required to describe the steep early decline in the in-sample shifting mean. The sum $\widehat{\delta}_0 + \widehat{\delta}_1 = 3.18$ is the value of the shifting mean when the first transition is complete, provided that at that time the remaining transition functions still have value zero. The next two rather steep transitions handle the outburst in inflation around 1990-92 and the decline following it, and the last one accounts for the late increase beginning 2005. The final value of the shifting mean equals $\sum_{j=0}^{5} \widehat{\delta}_j = 2.78$, which clearly exceeds the 2% target. It should be noted, however, that similarly to the estimated SM-AR model for the euro area inflation, the shifting mean of the UK inflation process only exhibits a few smooth shifts. It is not affected by transitory movements in headline inflation, so it can be interpreted as a measure of 'underlying' inflation.

3.4 US inflation

The monthly year-on-year US inflation series comprises the period from 1980(1) to 2010(6), and the series is graphed in Figure 3. The series has a structure similar to its European counterparts. The shifting mean of the SM-AR model with $p = 0$ fitted to this series has the following form:

$$
\begin{aligned}
\widehat{\delta}(t) = \underset{(0.15)}{12.48} &- \underset{(0.38)}{9.42}(1 + \exp\{-\underset{(-)}{17.3}(t/T - \underset{(-)}{0.04})\})^{-1} \\
&- \underset{(0.24)}{2.15}(1 + \exp\{-\underset{(-)}{17.3}(t/T - \underset{(-)}{0.41})\})^{-1} \\
&- \underset{(0.37)}{3.01}(1 + \exp\{-\underset{(-)}{30}(t/T - \underset{(-)}{0.95})\})^{-1} \\
&+ \underset{(0.33)}{1.49}(1 + \exp\{-\underset{(-)}{30}(t/T - \underset{(-)}{0.23})\})^{-1} \\
&+ \underset{(0.18)}{1.05}(1 + \exp\{-\underset{(-)}{30}(t/T - \underset{(-)}{0.81})\})^{-1}.
\end{aligned} \tag{6}
$$

Even this model contains five transitions. The first one accounts for the rapid decrease of the inflation rate in the early 1980s, and the Volcker disinflation period. As Figure 3 also shows, the mean is shifting upwards again in the late 1980s before the Gulf War ($\widehat{\delta}_4 = 1.49$). The next downward shift occurs around 1992–1993. After that the mean remains constant until around 2004 when the inflation rate again increases. The last transition around 2009 ($\widehat{\delta}_3 = -3.01$) corresponds to the steep decrease in the inflation rate that year. The final level, $\sum_{j=0}^{5} \widehat{\delta}_j = 0.44$, is almost exactly the same as for the euro area model. Overall, the shifting mean is quite similar to the one estimated for the UK inflation series, except for the latest development. The locations of the first four transitions match each other quite well, but the final downturn does not have a counterpart in the UK model.

4 Forecasting inflation with the SM-AR model using exogenous information

4.1 Penalised likelihood

The SM-AR model may not only be used for describing series that are assumed to be strongly influenced by unobserved or insufficiently observed

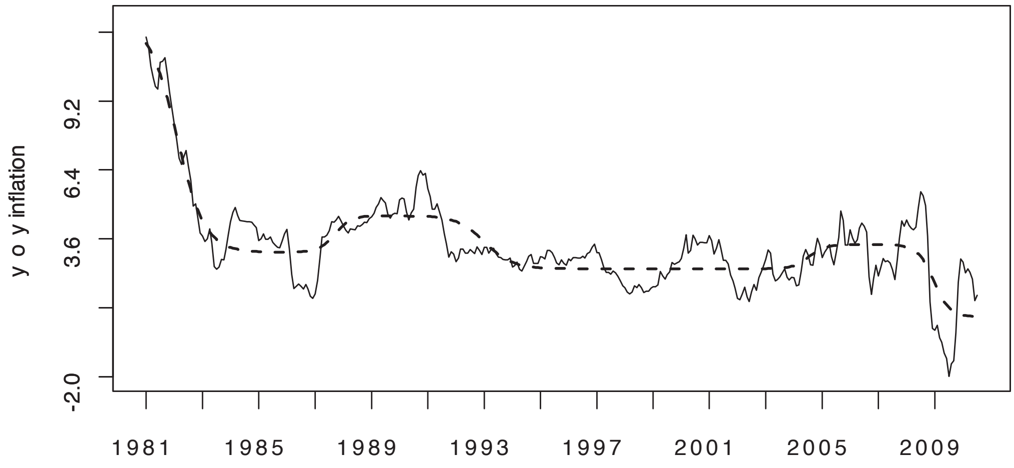

Figure 3: The US year-on-year inflation rate 1981(1)–2010(6) (solid curve) and the shifting mean from the SM-AR model (4) (dashed curve)

events. It may also be used for forecasting. The model offers an excellent possibility of making use of exogenous information in forecasting, such as inflation targets of central banks or inflation expectations of economic agents. Since central banks with an inflation target aim at keeping inflation close to the target value, the target contains information that should be incorporated, if not in short-term, at least in medium-term forecasts. It should be noted, howeer, that it may suffer from the same problem as autoregressive models with a linear trend, namely, that extrapolating the deterministic component may not yield satisfactory short-term forecasts. For very short-term forecasts, more flexible models than the SM-AR model may therefore be preferred; see, for example, Clements and Hendry (1999, Chapter 7) for discussion.

Our idea may be characterised as follows. Assuming T observations, the log-likelihood function of the SM-AR model has the following general form:

$$\ln L_T = \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}; y_t|\mathcal{F}_{t-1}) \tag{7}$$

where $\ell_t(\boldsymbol{\theta}; y_t|\mathcal{F}_{t-1})$, is the log-likelihood for observation t, $\boldsymbol{\theta}$ is the vector of parameters, and $\mathcal{F}_{t-1}$ is the σ-algebra defined by the past information up until $t-1$. Suppose the annual inflation target of the central bank is x and that the observations are year-on-year differences of the logarithmic price level p_t, $y_t = p_t - p_{t-12}$. Assume that one estimates the SM-AR model from data until time T and wants to forecast τ months ahead from T, for example $\tau = 24$ or 36. Ideally, from the point of view of the bank, $y_{T+\tau} = x$. Following the original suggestion of Good and Gaskins (1971), this target may now be incorporated into the forecast by penalising the likelihood. The penalised log-likelihood equals

$$\ln L_T^{\text{pen}} = \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}; y_t|\mathcal{F}_{t-1}) - \lambda\{\delta(T+\tau) - (1 - \sum_{j=1}^{p} \theta_j)x\}^2 \tag{8}$$

where τ is the forecast horizon of interest. The size of the penalty is determined by the nonnegative multiplier λ. When $\lambda \to \infty$, $\delta(T+\tau) \to (1 - \sum_j^p \theta_j)x$, that is, $\mathsf{E}_{T+\tau} y_{T+\tau} \to x$. The smoothly shifting mean, $\delta(t)$, will thus equal the target at time $T+\tau$. More generally, depending on λ, the forecast which is the conditional mean of $y_{T+\tau}$ at time $T+\tau$, lies in a neighbourhood of the target x.

The role of the penalty component is twofold. First, it is useful in preventing the extrapolated conditional mean from settling on values considered unrealistic. Second, as already mentioned, the penalised log-likelihood makes it possible to combine exogenous information about future inflation with what the model suggests. This bears some resemblance to the recent approach by Manganelli (2009). The difference is, however, that in his approach, the exogenous forecast is retained unless there is enough information in the data to abandon it. In our approach, the sample information always modifies the exogenous forecast or information in the form of the target, unless $\lambda \to \infty$ in (8).

It should be noted that if the SM-AR model is used simply for describing the in-sample behaviour of inflation, no penalty on the log-likelihood should be imposed. There is no contradiction, because time series models can be used for both data description and forecasting, and the estimated models for these two purposes need not be identical.

As already mentioned, it is assumed in equation (8) that y_t is directly the year-on-year inflation rate to be forecast, $y_t = p_t - p_{t-12}$. One may, however, model the monthly inflation rate $u_t = p_t - p_{t-1}$ and forecast the year-on-year inflation from the monthly SM-AR model. In this case, $y_t = \sum_{s=0}^{11} u_{t-s}$ and, accordingly, deviations of $\sum_{s=0}^{11} \mathsf{E}_{T+\tau-s} u_{T+\tau-s}$ from x are being penalized. Thus,

$$\ln L_T^{\text{pen}} = \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}; y_t|\mathcal{F}_{t-1}) - \lambda\{\sum_{s=0}^{11} \delta(T+\tau-s) - (1 - \sum_{j=1}^{p} \theta_j)x\}^2.$$

In this paper, however, we only report results obtained using models for the year-on-year inflation series. Since the Federal Reserve does not have an inflation target, we exclude the US inflation from the forecasting exercise. We include the euro area, as the ECB provides an explicit formulation for its aim of price stability, and the UK since the Bank of England is one of the inflation targeting central banks.

It may be argued that the ECB's definition of price stability (the year-on-year inflation 'below but close to 2%') is a target range rather than a point target. The penalised likelihood method still applies, however. In that case x may be taken to represent the mid-point of the range and that the size of the penalty is slightly larger than would be the case if x were a straightforward target. Strictly speaking, this idea is valid only when upward deviations from the range are equally undesirable as downward ones. If this is not the case, one has to construct asymmetric penalty functions. Note that in (8) the loss function of the forecaster is assumed to be quadratic. Other loss functions are possible as well. For example, Boinet and Martin (2005) and Orphanides and Wieland (2000), among others, consider nonlinear loss functions that they argue are applicable to central banks with an inflation target. According to the authors, these functions resemble a target zone function in that they are flat in a neighbourhood of the target. Note, however, that nonlinear loss functions imply numerical estimation algorithms, as the estimation problem no longer has an analytic solution.

It may be mentioned that information about the target could also be used in the analysis by applying Bayesian techniques. One would then have to choose a prior distribution for the target instead of choosing a value for the penalty term λ. Nevertheless, in the case of the SM-AR model the classical framework is well suited for the purpose of incorporating this information in the forecast.

4.2 Modification of penalised likelihood

We are going to make use of the following slight modification of the penalised likelihood:

$$\ln L_T^{\text{pen}} = \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}; y_t|\mathcal{F}_{t-1}) - \lambda \sum_{t=T+1}^{T+\tau} \rho^{T+\tau-t}\{\delta(t) - (1 - \sum_{j=1}^{p} \theta_j)x\}^2 \qquad (9)$$

where $0 < \rho < 1$. The penalty now involves all points of time from $T+1$ to $T+\tau$. The weights are geometrically decaying (other weighting schemes could be possible as well) from $T+\tau$ backwards. The geometrically (into the past) declining weights represent the idea that the forecast inflation path will gradually approach the target. But then, a rapid decay, $\rho = 0.8$, say, would give negligible weights to most observations in the penalty component preceding $T+\tau$, unless λ is very large. Even then, the first values following T would be negligible weights compared to the weight of the observation $T+\tau$. In that case, the results would be similar to the ones obtained by maximising (8).

We may also construct a slightly different modification by defining the standard weighted log-likelihood function as follows:

$$\begin{aligned}\ln L_T^{\text{pen}} &= c - \{(T-p)/2\}\ln\sigma^2 - (1/2\sigma^2)\sum_{t=p+1}^{T}(y_t - \delta(t) - \sum_{j=1}^{p}\theta_j y_{t-j})^2 \\ &\quad -\lambda(1/2\sigma^2)\sum_{t=T+1}^{T+\tau}\rho^{T+\tau-t}(y_t^* - \delta(t) - \sum_{j=1}^{p}\theta_j y_{t-j}^*)^2 \qquad (10)\end{aligned}$$

where

$$y_{T+k}^* = (1-k/\tau)y_T + (k/\tau)(1-k/\tau)x,\ k=1,...,\tau \qquad (11)$$

for $t = T+1,...,T+\tau$. In this case there is a set of artificial observations (11) obtained by linear interpolation between the last observation and the target. The previous log-likelihood (9) does not contain such observations. The forecast of $y_{T+\tau}$ equals

$$\text{est}\mathsf{E}(y_{T+\tau}|x) = (1 - \sum_{j=1}^{p}\widehat{\theta}_j\mathsf{L}^j)^{-1}\widehat{\delta}(T+\tau).$$

4.2.1 Parameter constancy test

When the time series are extended to contain the artificial observations $y_{T+1}^*,...,y_{T+\tau}^*$, the question is how to modify the linearity test. This can be done by using the weighted auxiliary regressions whose weights originate from equation (10). This is equivalent to assuming that there is heteroskedasticity of known form in the errors, and that it is accounted for in the test. The auxiliary regression based on the third-order Taylor expansion has the form; see, for example, Teräsvirta (1998):

$$\tilde{y}_t = \delta_0 + \delta_1\tilde{t}^* + \delta_2\tilde{t}^{*2} + \delta_3\tilde{t}^{*3} + \widetilde{\mathbf{w}}_t'\boldsymbol{\beta} + \varepsilon_t^* \qquad (12)$$

$\tilde{y}_t = y_t$ for $t = 1,...,T$; $\tilde{y}_t = \omega_t y_t^*$ for $t = T+1,\ldots,T+\tau$, where $\omega_t = \sqrt{\lambda\rho^{T+\tau-t}}$ for $t = T+1,\ldots,T+\tau$, and y_t^* is defined as in (11). Furthermore, $\tilde{t}^* = t/(T+\tau)$ for $t = 1,...,t$; $\tilde{t}^* = \omega_t t/(T+\tau)$ for $t = T+1,...,T+t$, and, finally, $\widetilde{\mathbf{w}}_t = (\tilde{y}_{t-1},\ldots,\tilde{y}_{t-p})'$. The QuickShift test sequence is carried out in the same way as in the case where the idea is merely to describe inflation, not to forecast it.

Another possibility is not to rerun the test sequence but rather retain the same number of shifts as is obtained by normal modelling of observations $y_1, ..., y_T$. In forecasting, the parameters of this model would simply be re-estimated by penalised likelihood, and the estimated shifting mean would then be used for forecasting. This short-cut would save computer time, but in our simulations we have respecified the model for each realisation.

4.2.2 Other methods for forecasting inflation

There is a large literature on forecasting inflation using time-varying parameter models. For a general discussion of time-varying parameter models, see Teräsvirta, Tjøstheim and Granger (2010, Sections 3.10–11). In the following, we shall only highlight a few papers containing different types of time-varying parameter models that have been used for forecasting inflation.

Stock and Watson (2007) found that a constant parameter MA(1) for the US inflation based on a rolling estimation window is a good approximation of their unobserved component model with a trend-cycle decomposition described in Section 2.3. They also reported that the rolling MA(1) performed very well in comparison with the other models in terms of forecasting up to four quarters ahead, while the AR model was slightly better for eight-quarters-ahead forecasts. Nevertheless, the rolling MA(1) model was still better than many other models they considered for that horizon. Ang, Bekaert and Wei (2007) compared term structure models, including nonlinear regime-switching specifications, with ARIMA models, Phillips curve type models and survey based forecasts, and found that surveys provided the most accurate forecasts for US inflation, whereas the term structure specifications fared relatively poorly in terms of one-quarter-ahead forecasts for different sample periods until 2002. Koop and Potter (2007) proposed a multiple-regime model in which the duration of a regime is generated from a Poisson process. They found that for US inflation a time-varying parameter model with a change-point provided somewhat better forecasts than their model for one-quarter-ahead forecasts over an evaluation period of two years (2004-2005). None of those papers, except for Stock and Watson (2007) considered 24-months-ahead forecasts, which will be the forecast horizon in our empirical analysis. Furthermore, none of them is designed for incorporating exogenous information into forecasts from the model. This feature separates our model with the others and is a key advantage of our methodology.

4.3 Monte Carlo experiments

4.3.1 The data-generating process

In order to illustrate forecasting with the SM-AR model in the presence of exogenous information, we conduct a small simulation experiment. The data are generated from models with and without autoregressive structure. The DGP has the following form:

$$y_t = \delta_0 + \sum_{i=1}^{3} \delta_i G(\gamma_i, c_i, t/(T+\tau)) + \mathbf{w}_t' \boldsymbol{\phi} + \varepsilon_t \tag{13}$$

$t = 1, ..., T$, where $(\delta_0, \delta_1, \delta_2, \delta_3) = (0.9, 0.2, 0.3, -0.4)$ with $\sum_{i=0}^{3} \delta_i = 1$. This means that the final value of the shifting mean equals unity. The transition functions are logistic functions of time as before:

$$G(\gamma_i, c_i, t/(T+\tau)) = (1 + \exp\{-\gamma(t/T - c_i)\})^{-1}, \quad \gamma_i > 0 \tag{14}$$

with (γ_i, c_i), $i = 1, 2, 3$, given by the pairs $(2, 0.3)$, $(6, 0.5)$ and $(4, 0.9)$. Furthermore, either $\mathbf{w}_t = (y_{t-1}, y_{t-2})'$, and $\boldsymbol{\phi} = (0.5, 0.3)'$ or $\boldsymbol{\phi} = \mathbf{0}$ (no autoregressive structure). In each realization, $T + \tau$ observations are generated, where T is the size of the estimation sample and τ the forecasting horizon. The artificial observations y^*_{T+k}, $k = 1, ..., \tau$, are defined as in (11). Time is rescaled into the zero-one interval such that $T + \tau$ now corresponds to value one. Two sample sizes, $T = 120, 240$, are considered. The target $x = 2, 4$, the forecast horizon $\tau = 36$, and the discount factor $\varrho = 0.9$. The number of replications equals 1000, and six different penalties are applied. The quantity reported for each replication is the point forecast. The model selection by QuickShift is performed for each replication. In these simulations, the initial significance level $\alpha_0 = 0.5$ and $\nu = 0.5$.

The target is higher than the final value of the logistic function. The greater the distance between the two, the higher the probability of obtaining a bimodal density forecast, *ceteris paribus*. A bimodal density results when the information in the time series as conveyed by the model sufficiently strongly deviates from the information provided by the target. Conversely, if the target and the final value of the shifting mean are close to each other, the density is more likely to be unimodal. The shape of the density also depends on the variance of the error distribution and the penalty. If the error variance is small and the penalty high, a conflict and thus a bimodal density is more likely than it is when the opposite is true. When the penalty approaches zero, the density becomes unimodal and symmetric.

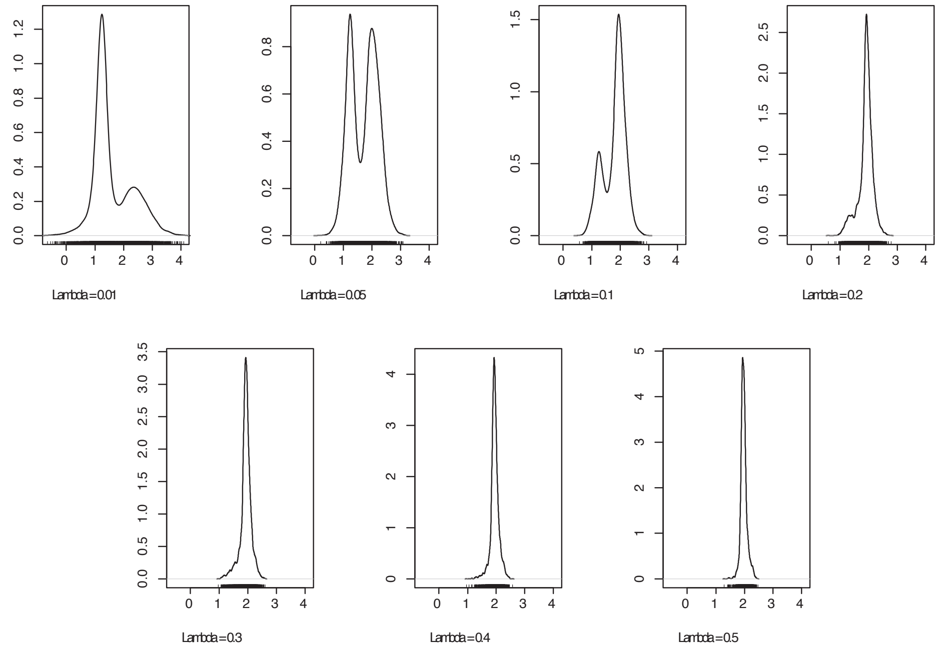

Figure 4: Density forecasts 36 periods ahead with target $x = 2$ and various penalties, $T = 120$. The penalty increases from left to right and from the first row to the second

4.3.2 Results

We shall only report results of the experiment with $p = 0$ and $x = 2$ because they are already sufficiently informative. The results for $T = 120$ are shown in Figure 4 by the estimated density function based on the 1000 point forecasts. As a whole, the results are quite predictable. When the penalty is small as it is in the top-row figures to the left, the density is bimodal but the second peak is relatively small. The mode of the distribution is slightly greater than one, the final value of the shifting mean, and there is a secondary peak somewhat to the right of the target. This is because even a small positive penalty already shifts the whole density to the right. When the penalty is increased, the leftmost peak decreases and eventually disappears as the forecasts on the average approach the target. In general, as already mentioned, the density is bimodal when the target and the shifting mean at the end of the sample are sufficiently different from each other, and the penalty is neither very small nor very large. Finally, when the penalty becomes large, the forecast density first becomes unimodal and then degenerate at the target value x when $\lambda \to \infty$.

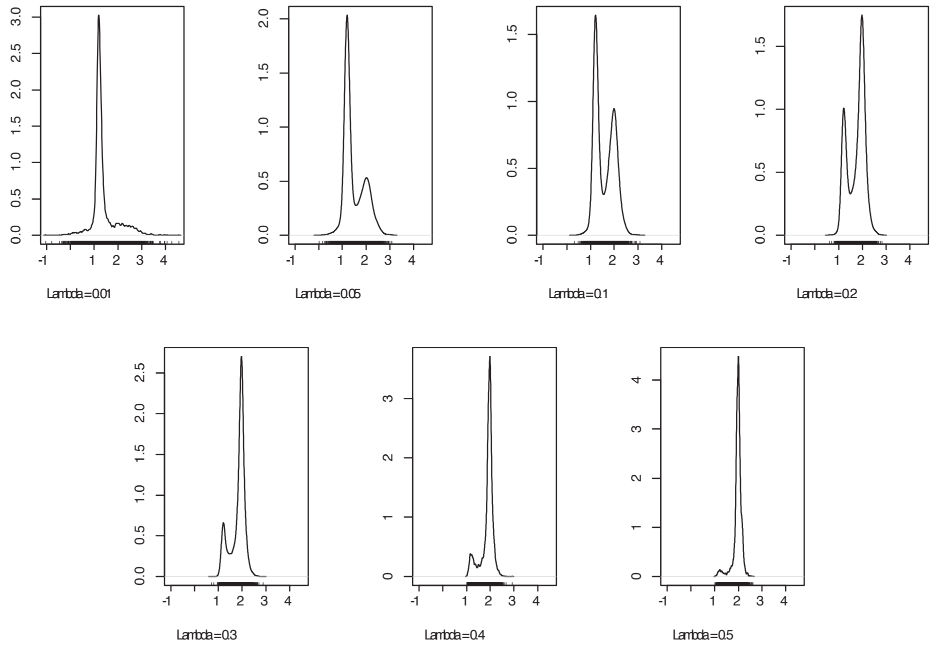

Figure 5: Density forecasts 36 periods ahead with target $x = 2$ and various penalties, $T = 240$. The penalty increases from left to right and from the first row to the second

Figure 5 contains results from the same experiment with $T = 240$. In this case, the sample information weighs more than previously. The density with the smallest penalty is close to unimodal, and the peak in the vicinity of unity disappears later than in the preceding simulation. A heavier penalty is now needed to eliminate it.

5 Application to forecasting inflation

5.1 Constructing point and density forecasts

We apply the SM-AR model and the penalised likelihood approach to forecasting both the euro area and the UK year-on-year inflation 24 months ahead for the period 2003(1)–2010(6). This period is a very relevant one for the euro area, as it begins shortly after the creation of the European Central Bank (ECB) and the introduction of the euro. ECB is aiming to achieve price stability at 'below but close to 2%' in terms of year-on-year inflation. For simplicity, we use a value of 2% for year-on-year inflation for the medium-term horizon when forecasting with the SM-AR model. The inflation target of Bank of England is 2%, and we apply it in our forecasting exercise. The penalty term λ determines the weight of the external information in (8). It reflects the forecaster's subjective assessment of the seriousness of the Bank and chances of success of its policies when it comes to bringing the inflation rate close to the target or holding it there.

In the penalised log-likelihood (8) the penalty is a quadratic function of the deviation from the 2% target. This does not exactly correspond to 'below but close to 2%' of ECB but serves as an approximation. As already mentioned, asymmetric penalty functions would be an alternative but are not considered here. A case could be made for a point target value somewhat below 2%, however, in particular as the penalty function is symmetric around the target.

The last observation of the euro area inflation series is 2010(6), and the forecast horizon equals 24 months. Forecasting starts in the beginning of 2001 and constinues till the end of the series. We report both point and density forecasts. The latter are obtained by a block bootstrap with a varying block size; see Politis and Romano (1994) and Politis and White (2004).

1. Specify and estimate the SM-AR model for the inflation series using 24 artificial observations in addition to the sample information. Obtain the point forecast.

2. Bootstrap the residuals of this model using the block bootstrap and generate a new set of T observations using the estimated model. Add the artificial observations. Repeat Step 1.

3. Repeat Step 2 B times.

4. Obtain the density forecast from the B point forecasts using kernel estimation (Teräsvirta et al. 2010, Section 13.1).

The SM-AR model is first specified and estimated using observations until 2001(1), so the first forecast will be for January 2003, as already indicated. The next observation is then added to the series, and the model is respecified and re-estimated. Respecification comprises selecting the number of transitions by QuickShift. This model is used for forecasting February 2003. New realisations are generated by a block bootstrap that involves respecification and re-estimation of the model for each of the 1000 bootstrap ($B = 1000$) replications. This is the number of replications behind each forecast density.

Variable Specification	Model		RMSFE Euro area	UK
π^{yoy}	SM-AR(λ)	λ		
		1/9	0.893	0.888
		3/7	0.855	0.851
		3/2	0.894	0.889
		9	0.915	0.910
π^{yoy}	AR(p)		0.981	0.797
	MS-AR		1.090	2.913
	2% forecast		0.907	1.747

Table 1: The root mean square error of 24-month forecasts of the Euro area and the UK inflation from the SM-AR model with various penalties, the linear AR(p),p selected by the Rissanen-Schwarz information criterion, and the first-order Markov-switching AR model. Forecasting period: 2003(1)–2010(6).

5.2 Point forecasts

Table 1 contains the root mean square forecast errors (RMSFE) of the point forecasts 24 months ahead from the SM-AR model with four different penalties for both the euro area and the UK inflation forecasts for the period 2003(1)–2010(6). It also contains the RMSFE of the forecasts of three benchmark models, including the linear AR model based on the same transformation as the SM-AR model (i.e., year-on-year inflation, π^{yoy}), the target or 'quantitative aim of price stability' itself (a constant 2% forecast for every period) and a first-order Markov-switching AR (MS-AR) model. The MS-AR model is parameterised as in Tyssedal and Tjøstheim (1988) with the possible extension that the error variance may switch as well. As usual, its parameters are estimated using the EM algorithm of Dempster, Laird and Rubin (1977).

The results in Table 1 show that in forecasting euro area inflation the SM-AR model outperforms the benchmark models considered here. For the euro area inflation forecasts, the size of the penalty does not make a big difference. The RMSFE is larger for the AR than for the SM-AR model regardless of the size of the penalty. The Markov-switching AR (MS-AR) model does not perform well either. An obvious reason for the good performance of our model is that, with the exception of last two years, euro area year-on-year inflation has remained close to 2%. This is also seen from the fact that the RMSFE of the constant 2% forecasts is not much inferior to that of our SM-AR model forecasts.

The situation is different for the UK inflation point forecasts. In terms of the RMSFE, the AR(p) model is the best performer, whereas the forecasts from the MS-AR model are very inaccurate.The UK inflation has been clearly above the target most of the time since the end of 2007. The AR model has been able to forecast these developments rather well, whereas giving weight to the target has not been helpful. This is also obvious from the fact that the pure target forecasts are quite inaccurate. If the years 2008–2010 are excluded from the comparison, the SM-AR model generates the most accurate point forecasts.[5] The size of the penalty does not matter much in the sense that the lightest penalty, $\lambda = 1/99$, already gives too much weight to the target. Nevertheless, investing some trust in the target when forecasting before the year 2006 would have been the right thing to do. A real-time forecaster would probably have adjusted the size of the penalty during the forecasting period according to his or her judgment, but in this experiment it has remained constant throughout the period.

5.3 Forecast densities

5.3.1 Euro area inflation

Forecast densities of the 24-month forecasts for euro area inflation for the period 2003(1)–2010(6) can be found in Figure 6. The figure contains the 50%, 70% and 90% highest density regions (HDR) for $\lambda = 1/99$, $1/9$, $3/7$ and $3/2$. This implies the following relative weights for the penalty: 0.01, 0.1, 0.3, and 0.5. An HDR is the set of intervals within which the density is higher than anywhere else, see Hyndman (1996) or Teräsvirta et al. (2010, Section 15.2). The 90% confidence intervals of the linear AR model are also presented. Since the corresponding forecast densities are symmetric around the mean, they are comparable to the 90% HDR from the SM-AR model.

It is seen from Figure 6 that for the lowest penalty, $\lambda = 1/99$, the forecast density is bimodal until the end of 2006 and then unimodal up to 2010. At the end, this density is bimodal again, with one local model close to 4% and another one in the vicinity of the assumed target value 2%. This reflects the fact that the euro area inflation was high, around 4%, in mid-2008 when the last 24-month forecasts were made. The next panel with $\lambda = 1/9$ shows how the peak around 4% is flattened out and moved close to the 2% mark when the penalty is increased from 1/99 to 1/9. Increasing it further decreases the variance of the distribution even more and concentrates the probability mass around 2%. When inflation is close to the assumed target and the penalty is low ($\lambda \leq 1/9$), the 50% and 70% forecast densities cover the observed values

[5] Results are not reported to save space but are available upon request.

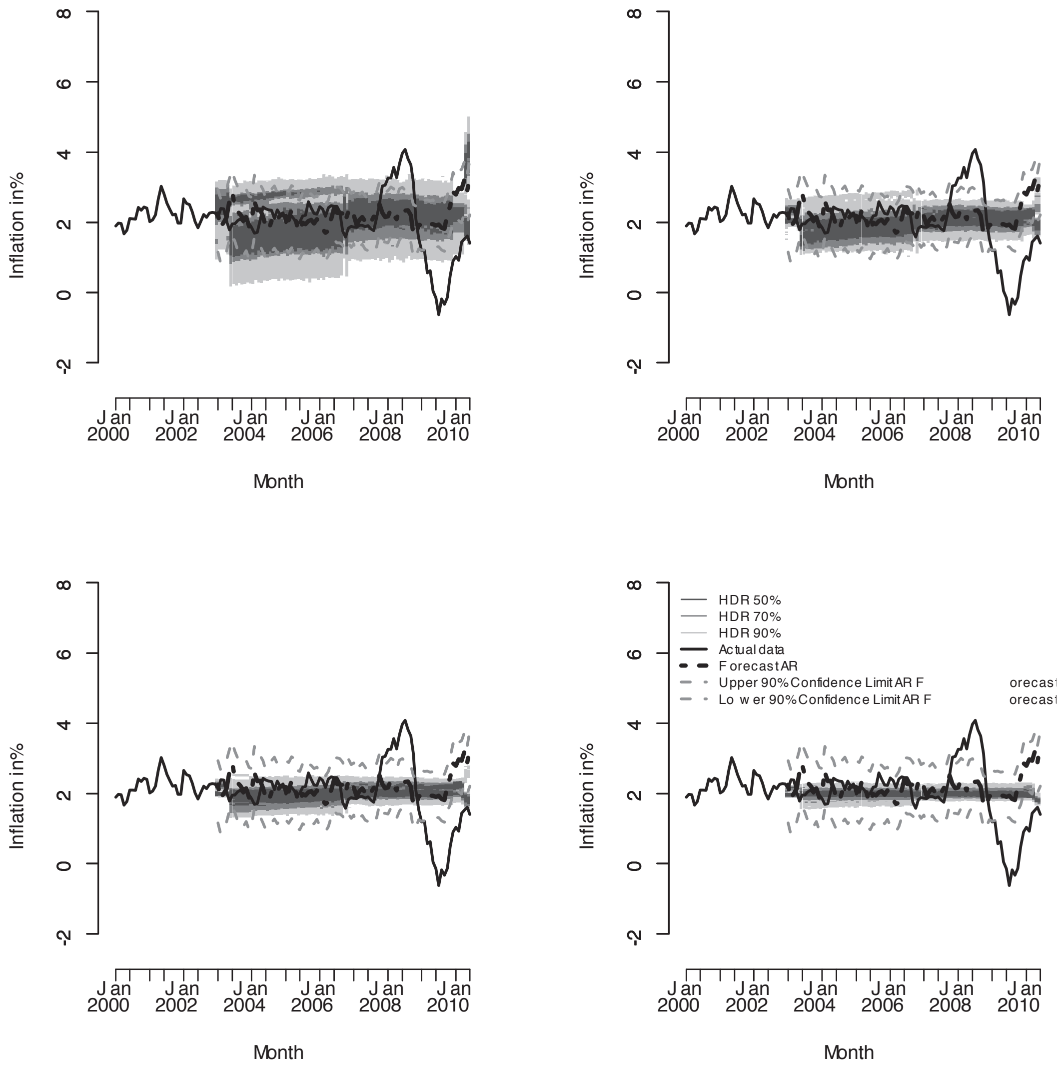

Figure 6: 50%, 70% and 90% highest density regions of 24-month density forecasts of euro area inflation, 2003(1)–2010(6) for $\lambda = 1/100$ (upper left panel), $\lambda = 1/10$ (upper right panel), $\lambda = 3/7$ (lower left panel), and $\lambda = 3/2$ (lower right panel).

quite well until mid-2007, after which the peak values in 2008 and the trough points in 2009 fall outside even the 90% HDRs.

Figure 6 also contains the point forecasts from the linear AR model and their 90% confidence limits. The year-on-year inflation remains inside the band defined by these limits until mid-2007. After this, similarly to the aforementioned HDRs, the band contains neither the high-inflation values until mid-2008 nor the low-inflation ones in 2009. Furthermore, it fails to cover the still low monthly year-on-year values in 2010.

5.3.2 UK inflation

As already mentioned, the Bank of England has an inflation target of 2%. The forecast horizon is again 24 months and the last observation to be forecast is 2010(6). The density forecasts for the year-on-year inflation appear in Figure 7. The four penalties are the same as in the euro area forecasts. The point forecasts from the autoregressive model are upward biased in the beginning of the period due to the fact that before 2005 the unconditional mean of the UK inflation process is higher than the inflation rate. They become more accurate when inflation picks up. The 90% intervals of the linear AR forecasts contain the realised inflation most of the time but fail to do that when inflation is peaking in 2008.

The forecast densities from the SM-AR model for the smallest penalty, $\lambda = 1/99$, are mostly unimodal and widen considerably around the end of 2007. The mode of the density remains below the target. This is due to the fact that when forecasting begins, the estimated conditional mean is clearly below 2%, which can also be seen from Figure 2. It stays there for a long time, because the inflation rate remains very low until early 2005. This explains why the point forecasts from the SM-AR model are less accurate on average than the ones from the linear AR model. Interestingly, when there is enough evidence about the rise in inflation by early 2006, the 24-month density forecasts made for the months of 2008 are skewed with a long upper tail. The 90% HDRs in Figure 7 cover the whole inflation series including the peak in 2008. When $\lambda = 1/9$, this rise in the inflation rate leads to a bimodal or even trimodal forecast densities such that the upper local mode is close to the target, whereas the lower one still reflects the period of low inflation before 2006. In this case, the 90% HDRs no longer cover the 2008 inflation rates. Further increases of the penalty concentrate the density around the 2% target throughout. When $\lambda = 3/7$, weak bi- or multimodality remains, but it almost disappears when $\lambda = 3/2$. In that case, the lower tail of the density is longer than the upper one.

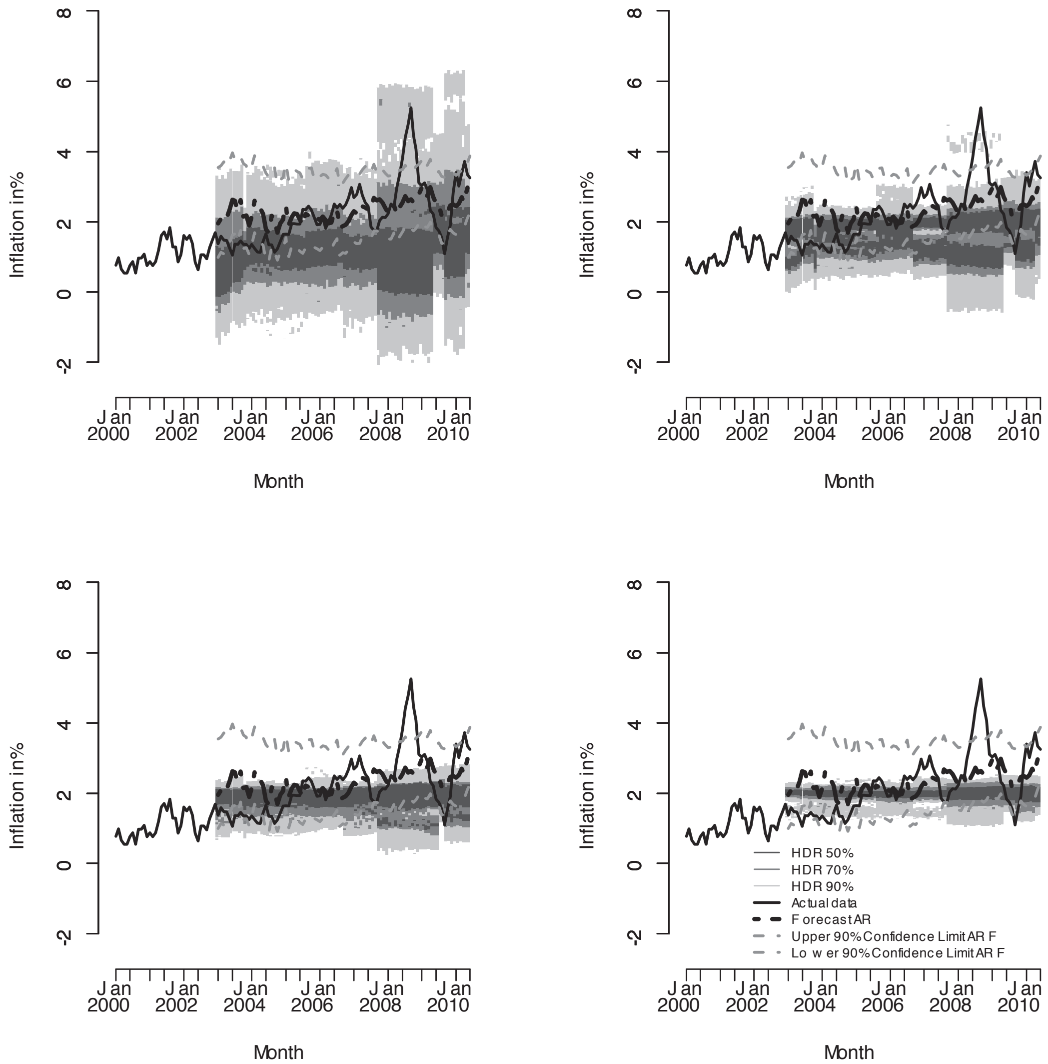

Figure 7: 50%, 70% and 90% highest density regions of 24-month density forecasts of UK inflation, 2003(1)–2010(6) for $\lambda = 1/100$ (upper left panel), $\lambda = 1/10$ (upper right panel), $\lambda = 3/7$ (lower left panel), and $\lambda = 3/2$ (lower right panel).

6 Conclusions

In this work we propose a new method for medium-term forecasting incorporating exogenous information based on a flexible nonstationary autoregressive model, the shifting-mean autoregressive model. In forecasting inflation, a central bank inflation target, if it exists, is a natural example of such exogenous information. Forecasting is carried out using the penalised likelihood in the estimation of the parameters of the model. Another advantage of the shifting-mean autoregressive model worth mentioning is that it is also suitable for describing characteristic features of time series of inflation.

The inflation target is an example of a piece of deterministic *a priori* information. It may also be possible to handle stochastic exogenous information, for example another point forecast. If the uncertainty of this forecast is assumed to be known, that is, if the forecast is a draw from a known probability distribution, this uncertainty can be taken into account when generating density forecasts with the technique described in the paper. That has not, however, been done here.

There is also the possibility of making the model multivariate by including stochastic regressors. They may appear linearly in the usual way or even nonlinearly as arguments of logistic functions. In the latter case they could be included in the pool from which QuickShift selects the appropriate variables for the model. It would also be possible to use other techniques than QuickShift to select the components; see Kock and Teräsvirta (2011) for examples. Such extensions are, however, left for future work.

References

Amemiya, T.: 1985, *Advanced Econometrics*, Blackwell, Oxford.

Ang, A., Bekaert, G. and Wei, M.: 2007, Do macro variables, asset markets, or surveys forecast inflation better?, *Journal of Monetary Economics* **54**, 1163–1212.

Beechey, M. and Österholm, P.: 2010, Forecasting inflation in an inflation-targeting regime: A role for informative steady-state priors, *International Journal of Forecasting* **26**, 248–264.

Benedetti, J. K.: 1977, On the nonparametric estimation of regression functions, *Journal of the Royal Statistical Society, Series B* **39**, 248–253.

Boinet, V. and Martin, C.: 2005, Targets, zones and asymmetries: A flexible nonlinear models of recent UK monetary policy, *Working paper*, Brunel University.

Clark, T. E.: 2001, Comparing measures of core inflation, *Economic Review*, Federal Reserve Bank of Kansas.

Clements, M. P. and Hendry, D. F.: 1999, *Forecasting Non-stationary Economic Time Series*, MIT Press, Cambridge,Massechusetts.

Cogley, T.: 2002, A simple adaptive measure of core inflation, *Journal of Money, Credit and Banking* **34**, 94–113.

Cogley, T., Primiceri, G. E. and Sargent, T. J.: 2010, Inflation gap persistence in the US, *American Economic Review: Macroeconomics* **2**, 43–69.

Cogley, T. and Sbordone, A. M.: 2008, Trend inflation, indexation, and inflation persistence in the New Keynesian Phillips curve, *American Economic Review* **98**, 2101–2126.

Cristadoro, R., Forni, M., Reichlin, L. and Veronese, G.: 2005, A core inflation indicator for the euro area, *Journal of Money, Credit and Banking* **37**, 539–560.

Cybenko, G.: 1989, Approximation by superposition of sigmoidal functions, *Mathematics of Control, Signals, and Systems* **2**, 303–314.

Davidson, J.: 2000, *Econometric Theory*, Blackwell, Oxford.

Dempster, A. P., Laird, N. M. and Rubin, D. B.: 1977, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B,* **39**, 1–38.

Doornik, J. A.: 2008, Encompassing and automated model selection, *Oxford Bulletin of Economics and Statistics* **70**, 915–925.

Doornik, J. A.: 2009, Autometrics, *in* J. L. Castle and N. Shephard (eds), *The Methodology and Practice of Econometrics*, Oxford University Press, Oxford, pp. 88–121.

Funahashi, K.: 1989, On the approximate realization of continuous mappings by neural networks, *Neural Networks* **2**, 183–192.

González, A. and Teräsvirta, T.: 2008, Modelling autoregressive processes with a shifting mean, *Studies in Nonlinear Dynamics and Econometrics* **12**, No. 1, Article 1.

Good, I. J. and Gaskins, R. A.: 1971, Nonparametric roughness penalties for probability densities, *Biometrika* **58**, 255–277.

Hornik, K., Stinchcombe, M. and White, H.: 1989, Multi-layer feedforward networks are universal approximators, *Neural Networks* **2**, 359–366.

Huang, J., Horowitz, J. L. and Ma, S.: 2008, Asymptotic properties of bridge estimators in sparse high-dimensional regression models, *Annals of Statistics* **36**, 587–613.

Hyndman, R. J.: 1996, Computing and graphing highest density regions, *American Statistician* **50**, 120–126.

Kock, A. B. and Teräsvirta, T.: 2011, Nonlinear forecasting of macroeconomic variables using automated model selection techniques, *CREATES Research Paper*, Aarhus University.

Koop, G. and Potter, S. M.: 2007, Estimation and forecasting in models with multiple breaks, *Review of Economic Studies* **74**, 763–789.

Kozicki, S. and Tinsley, P. A.: 2005, Permanent and transitory policy shocks in an empirical macro model with asymmetric information, *Journal of Economic Dynamics and Control* **29**, 1985–2015.

Kozicki, S. and Tinsley, P. A.: 2006, Survey-based estimates of the term structure of expected U.S. inflation, *Working Paper 2006-46*, Bank of Canada.

Lee, T.-H., White, H. and Granger, C. W. J.: 1993, Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests, *Journal of Econometrics* **56**, 269–290.

Lendvai, J.: 2006, Inflation dynamics and regime shifts, *ECB Working Paper 684*, European Central Bank.

Leser, C. E. V.: 1961, A simple method of trend construction, *Journal of the Royal Statistical Society, Series B,* **23**, 91–107.

Lin, C.-F. J. and Teräsvirta, T.: 1994, Testing the constancy of regression parameters against continuous structural change, *Journal of Econometrics* **62**, 211–228.

Manganelli, S.: 2009, Forecasting with judgment, *Journal of Business and Economic Statistics* **27**, 553–563.

Orphanides, A. and Wieland, V.: 2000, Inflation zone targeting, *European Economic Review* **44**, 1351–1387.

Orphanides, A. and Williams, J. C.: 2005, Inflation targeting, *in* B. Bernanke and M. Woodford (eds), *Imperfect Knowledge, Inflation Expectations and Monetary Policy*, University of Chicago Press.

Pesaran, M. H., Pettenuzzo, D. and Timmermann, A.: 2006, Forecasting time series subject to multiple structural breaks, *Review of Economic Studies* **73**, 1057–1084.

Politis, D. N. and Romano, J. P.: 1994, Limit theorems for weakly dependent hilbert space valued random variables with application to the stationary bootstrap, *Statistica Sinica* **4**, 461–476.

Politis, D. N. and White, H.: 2004, Automatic block-length selection for the dependent bootstrap, *Econometric Reviews* **23**, 53–70.

Priestley, M. B. and Chao, M. T.: 1972, Non-parametric function fitting, *Journal of the Royal Statistical Society, Series B* **34**, 385–392.

Schorfheide, F.: 2005, Learning and monetary policy shifts, *Review of Economic Dynamics* **8**, 392–419.

Sims, C. A. and Zha, T.: 2006, Were there regime switches in U.S. monetary policy?, *American Economic Review* **96**, 54–81.

Stock, J. H. and Watson, M. W.: 2007, Why has U.S. inflation become harder to forecast?, *Journal of Money, Credit and Banking* **39**, 3–33.

Teräsvirta, T.: 1998, Modeling economic relationships with smooth transition regressions, *in* A. Ullah and D. E. Giles (eds), *Handbook of Applied Economic Statistics*, Dekker, New York, pp. 507–552.

Teräsvirta, T., Tjøstheim, D. and Granger, C. W. J.: 2010, *Modelling Nonlinear Economic Time Series*, Oxford University Press, Oxford.

Tyssedal, J. S. and Tjøstheim, D.: 1988, An autoregressive model with suddenly changing parameters, *Applied Statistics* **37**, 353–369.

White, H.: 2006, Approximate nonlinear forecasting methods, *in* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Vol. 1, Elsevier, Amsterdam, pp. 459–512.

Appendix

A Selecting sigmoids using QuickShift

We shall give a brief description of QuickShift. QuickNet, its more general version, is used for specifying the number of hidden units in a single hidden-layer feedforward artificial neural network model. The user first fixes the maximum number of 'hidden units', corresponding to transition functions in this work, and selects the units from a large set of predetermined candidate functions. The same is true for QuickShift. The maximum number of transition functions $\bar{q}$ can be set to equal any value such that the model can be estimated, given the sample size. Here, $\bar{q} = 10$. The set of candidate functions is defined by a fixed grid for γ and c. In our applications, the grid will be defined as $\Theta_N = \{(\Gamma_{N_\gamma} \times C_{N_c})\}$ with $\Gamma_{N_\gamma} = \{\gamma_s : \gamma_s = \kappa\gamma_{s-1}, s = 1, \ldots, N_\gamma, \kappa \in (0,1)\}$ and $C_{N_c} = \{c_s : c_s = c_{s-1} + (1/N_c), s = 1, \ldots, N_c\}$. The starting-values are $\gamma_0 = 0.01$ and $c_0 = 0.01$. The final values are $\gamma_N = 30$ and $c_N = 0.99$, and, furthermore $N_c = 100$ and $N_\gamma = 100$. This defines a set of 10000 different transition functions. Since γ is not a scale-free parameter, it is divided by the 'standard deviation' of t/T when constructing the grid. The idea behind all this is to transform the nonlinear model selection and estimation problem into a linear one.

Given $\bar{q}$ and Θ_N, QuickShift consists of the following steps:

1. Estimate model (1) assuming $\delta(t) = \delta_0$, save the residuals $\hat{\varepsilon}_{t,0}$.

2. After selecting $q-1$ transitions, $q>1$, choose the transition function that in absolute terms has the largest correlation with $\hat{\varepsilon}_{t,q-1}$ that is, let

$$(\hat{\gamma},\hat{c})_q = \text{argmax}_{(\gamma_s,c_s)\in\Theta_N}[r(g(\gamma_s,c_s,t/T),\hat{\varepsilon}_{q-1,t})]^2$$

where $r(g(\gamma_s,c_s,t/T),\hat{\varepsilon}_{q-1,t})$ is the sample correlation between $g(\gamma_s,c_s,$ $t/T)$ and

$$\hat{\varepsilon}_{q-1,t} = y_t - \hat{\delta}_0 - \sum_{i=1}^{q-1}\hat{\delta}_i g(\hat{\gamma}_i,\hat{c}_i,t/T) - \sum_{j=1}^{p}\widehat{\phi}_j y_{t-j}.$$

Test the model with $q-1$ transitions against its counterpart with q transitions; for details see González and Teräsvirta (2008). If the null hypothesis is rejected, proceed to Step 3. In order to have the overall significance level of the sequence under control as well as to favour parsimony, the significance level α_q of an individual test is gradually decreased such that $\alpha_q = \nu\alpha_{q-1}$, $q=1,2,...,$ where $0<\nu<1$. The user determines α_0 and ν.

3. Given $(\hat{\gamma},\hat{c})_q$, obtain the estimates $(\hat{\delta}_0,\ldots,\hat{\delta}_q,\widehat{\phi}_1,\ldots,\widehat{\phi}_q)'$ by ordinary least squares. Go back to Step 2.

4. If every null hypothesis is rejected, stop at $q=\bar{q}$. The choice of $\bar{q}$, the maximum number of transitions, is controlled by the user and depends on the modelling problem at hand.

The test used for selecting q is the Taylor expansion based test by Lin and Teräsvirta (1994). Other choices, such as the Neural Network test by Lee, White and Granger (1993), are possible, and one can also apply model selection criteria to this selection problem. In the simulations reported in González and Teräsvirta (2008), the model selection criteria they investigated performed less well than the sequential tests and will not be used here.

B Asymptotic normality of the maximum likelihood estimators of the SM-AR model

Let $\boldsymbol{\theta} = (\boldsymbol{\phi}',\boldsymbol{\delta}',\boldsymbol{\gamma}',\mathbf{c}')' \in \Theta \subset \mathbb{R}^{3q+1}$ where $\boldsymbol{\phi} = (\phi_1,...,\phi_p)'$ is a $p\times 1$ vector, $\boldsymbol{\delta} = (\delta_0,\delta_1,...,\delta_q)'$ is a $(q+1)\times 1$ vector, and $\boldsymbol{\gamma} = (\gamma_1,...,\gamma_q)'$ and $\mathbf{c} = (c_1,...,c_q)'$ are $q\times 1$ vectors. Let $\boldsymbol{\theta}_0 = (\boldsymbol{\delta}_0',\boldsymbol{\gamma}_0',\mathbf{c}_0')'$ be the corresponding true parameter vector. The model for y_t is

$$y_t = \delta(t) + \sum_{j=1}^{p}\phi_j y_{t-j} + \varepsilon_t \tag{15}$$

where $\varepsilon_t \sim \text{iid}(0, \sigma^2)$. Furthermore, $\delta(t)$ is a bounded positive-valued function of time:

$$\delta(t) = \delta_0 + \sum_{i=1}^{q} \delta_i g(\gamma_i, c_i, t/T) \tag{16}$$

such that δ_i, $i = 1, \ldots, q$, are parameters and $g(\gamma_i, c_i, t/T)$, $i = 1, \ldots, q$, are logistic transition functions:

$$g(\gamma_j, c_j, t/T) = g_{jt} = \left(1 + \exp\{-\gamma_j(t/T - c_i)\}\right)^{-1}, \ \gamma_j > 0 \tag{17}$$

for $j = 1, ..., q$, where T is the number of observations. Furthermore, assume $c_1 < c_2 < ... < c_q$. The value of $\delta(t)$ thus changes (possibly nonmonotonically) from δ_0 to $\delta_0 + \sum_{i=1}^{q} \delta_i$ as a function of t. The definition (17) implies that $g(\gamma_j, c_j, t/T)$ and thus $\delta(t)$ is continuous and infinitely many times differentiable in $\boldsymbol{\theta}$. Rescaled time in the argument of $g(\gamma_j, c_j, t/T)$ leaves the relative locations of transitions intact as $T \to \infty$. We make the following assumptions:

Assumption A1. The parameter space Θ is an open subset of $\mathbb{R}^{3q+p+1}$ and $\boldsymbol{\theta}_0$ is an interior point of Θ.

Assumption A2. The roots of the lag polynomial $1 - \sum_{j=1}^{p} \phi_j z^j$ lie outside the unit circle, and $\sum_{j=0}^{\infty} |\theta_j| < \infty$ in

$$(1 - \sum_{j=1}^{p} \phi_j z^j)^{-1} = \sum_{j=0}^{\infty} \theta_j z^j.$$

The quasi log-likelihood function (T observations) of the model is defined as follows:

$$L_T(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \sum_{t=1}^{T} \ell(\boldsymbol{\theta}, \varepsilon_t) \tag{18}$$

where

$$\ell(\boldsymbol{\theta}, \varepsilon_t) = k - (1/2) \ln \sigma^2 + \frac{\varepsilon_t^2}{2\sigma^2}. \tag{19}$$

Lemma A.1. The $(3q + p + 1) \times 1$ score function $\partial \ell(\boldsymbol{\theta}, \varepsilon_t)/\partial \boldsymbol{\theta}$ for observation t has the form

$$\partial \ell(\boldsymbol{\theta}, \varepsilon_t)/\partial \boldsymbol{\theta} = -\frac{\varepsilon_t}{\sigma^2} \frac{\partial \varepsilon_t}{\partial \boldsymbol{\theta}} = \frac{\varepsilon_t}{\sigma^2} \mathbf{g}_t(\boldsymbol{\theta})$$

where $\mathbf{g}_t(\boldsymbol{\theta})$=$(\mathbf{y}'_{t-1}, \mathbf{g}'_{\delta t}, \mathbf{g}'_{\gamma t}, \mathbf{g}'_{ct})'$. The blocks of $\mathbf{g}_t(\boldsymbol{\theta})$ are

$$\begin{aligned}
\mathbf{y}_{t-1} &= \partial g_t/\partial\boldsymbol{\phi} = (y_{t-1}, ..., y_{t-p})' \\
\mathbf{g}_{\delta t} &= \partial g_t/\partial\boldsymbol{\delta} =(1, g_{1t}, ..., g_{qt})' \\
\mathbf{g}_{\gamma t} &= \partial g_t/\partial\boldsymbol{\gamma} =(g_{\gamma 1t}, ..., g_{\gamma qt})' \\
\mathbf{g}_{ct} &= \partial g_t/\partial\mathbf{c} =(g_{c1t}, ..., g_{cqt})'
\end{aligned}$$

where $g_{\gamma jt} = \delta_j g_{jt}(1-g_{jt})(t/T-c_j)$ and $g_{cjt} = -\gamma_j\delta_j g_{jt}(1-g_{jt})$ for $j = 1, ..., q$.

Lemma A.2. The Hessian $\partial^2\ell(\boldsymbol{\theta},\varepsilon_t)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ for observation t equals

$$\partial^2\ell(\boldsymbol{\theta},\varepsilon_t)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = -\frac{1}{\sigma^2}\{\mathbf{g}_t(\boldsymbol{\theta})\mathbf{g}'_t(\boldsymbol{\theta}) + \varepsilon_t(\boldsymbol{\theta})\frac{\partial^2\mathbf{g}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\}$$

where

$$\mathbf{g}_t(\boldsymbol{\theta})\mathbf{g}_t(\boldsymbol{\theta})' = \begin{bmatrix} \mathbf{M}_{\phi\phi t} & \mathbf{M}_{\phi\delta t} & \mathbf{M}_{\phi\gamma t} & \mathbf{M}_{\phi ct} \\ & \mathbf{M}_{\delta\delta t} & \mathbf{M}_{\delta\gamma t} & \mathbf{M}_{\delta ct} \\ & & \mathbf{M}_{\gamma\gamma t} & \mathbf{M}_{\gamma ct} \\ & & & \mathbf{M}_{cct} \end{bmatrix}$$

with

$$\begin{aligned}
\mathbf{M}_{\phi\phi t} &= \mathbf{y}_{t-1}\mathbf{y}'_{t-1},\ \mathbf{M}_{\phi\alpha t} = \mathbf{y}_{t-1}\mathbf{g}'_{\phi\alpha t},\ \alpha = \delta, \gamma, c \\
\mathbf{M}_{\delta\delta t} &= \mathbf{g}_{\delta t}\mathbf{g}'_{\delta t},\ \mathbf{M}_{\delta\alpha t} = \mathbf{g}_{\delta t}\mathbf{g}'_{\alpha t} + \varepsilon_t \mathrm{diag}(g_{\delta\alpha 1t}, ..., g_{\delta\alpha qt}),\ \alpha = \gamma, c \\
\mathbf{M}_{\gamma\gamma t} &= \mathbf{g}_{\gamma t}\mathbf{g}'_{\alpha t} + \varepsilon_t \mathrm{diag}(g_{\gamma\alpha 1t}, ..., g_{\gamma\alpha qt}),\ \alpha = \gamma, c \\
\mathbf{M}_{cct} &= \mathbf{g}_{ct}\mathbf{g}'_{ct} + \varepsilon_t \mathrm{diag}(g_{cc1t}, ..., g_{ccqt})
\end{aligned}$$

and

$$\begin{aligned}
g_{\gamma\gamma jt} &= \delta_j g_{jt}(1 - g_{jt})(1 - 2g_{jt})(t^* - c_j)^2 \\
g_{ccjt} &= \delta_j \gamma_j^2 g_{jt}(1 - g_{jt})(1 - 2g_{jt}) \\
g_{\gamma\delta jt} &= g_{jt}(1 - g_{jt})(t^* - c_j) \\
g_{c\delta jt} &= -\gamma_j g_{jt}(1 - g_{jt}) \\
g_{\gamma cjt} &= -\delta_j\gamma_j g_{jt}(1 - g_{jt})(1 - 2g_{jt})(t^* - c_j)
\end{aligned}$$

for $j = 1, ..., q$.

Lemma A.3. The probability limit

$$\mathbf{M}_{\phi\phi} = \mathrm{plim}_{T\to\infty}(1/T)\sum_{t=1}^{T}\mathbf{M}_{\phi\phi t} = \lim_{T\to\infty}(1/T)\sum_{t=1}^{T}\boldsymbol{\delta}^*(t-1)\boldsymbol{\delta}^*(t-1)' + \mathrm{cov}(\mathbf{y}^*_t)$$

where $\mathbf{y}_t^* = (y_t^*, y_{t-1}^*, ..., y_{t-p+1}^*)'$ with $y_t^* = y_t - \delta(t)$, and

$$\boldsymbol{\delta}^*(t) = (\delta(t), ..., \delta(t-p+1))'.$$

Proof. Consider

$$\begin{aligned} y_t &= (1 - \sum_{j=1}^{p} \phi_j \mathsf{L}^j)^{-1}\delta(t) + (1 - \sum_{j=1}^{p} \phi_j \mathsf{L}^j)^{-1}\varepsilon_t \\ &= \delta^*(t) + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \end{aligned} \tag{20}$$

where $\delta^*(t) = \sum_{j=0}^{\infty} \theta_j \delta(t-j)$ and $\theta_0 = 1$. Then

$$|\delta^*(t)| \leq \sum_{j=0}^{\infty} |\theta_j||\delta(t-j)| \leq \delta_{\max} \sum_{j=0}^{\infty} |\theta_j|.$$

This implies that $\delta^*(t)$ and, consequently, $\delta^{*2}(t)$, are finite.

Next, consider

$$\begin{aligned} y_t^2 &= (\delta^*(t) + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j})^2 \\ &= \delta^{*2}(t) + 2\delta^*(t) \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} + (\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j})^2 \end{aligned}$$

where $\mathsf{E}(\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j})^2 = \sigma^2 \sum_{j=0}^{\infty} \theta_j^2$. Then

$$\begin{aligned} \text{plim}_{T\to\infty}(1/T)(\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j})^2 &= \sigma^2 \sum_{j=0}^{\infty} \theta_j^2 \\ \text{plim}_{T\to\infty}(1/T)\{\sum_{t=0}^{\infty} \delta^*(t)(\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j})\} &= 0 \end{aligned}$$

and

$$\text{plim}_{T\to\infty}(1/T) \sum_{t=0}^{\infty} y_t^2 = \lim_{T\to\infty}(1/T) \sum_{t=0}^{\infty} \delta^{*2}(t) + \sigma^2 \sum_{j=0}^{\infty} \theta_j^2 \tag{21}$$

by Kolmogorov's LLN 1. The limit in (21) is finite since all elements in the first sum are $O(1)$ and by Assumption A2, $\sum_{j=0}^{\infty} \theta_j^2 < \infty$.

Correspondingly,

$$\text{plim}_{T\to\infty}(1/T)\sum_{t=0}^{\infty} y_t y_{t-j} = \lim_{T\to\infty}(1/T)\sum_{t=0}^{\infty}\delta^*(t)\delta^*(t-j) + \sigma^2\sum_{k=1}^{\infty}\theta_k\theta_{k+j}. \tag{22}$$

where the limit is again finite. Expressions (21) and (22) complete the proof. ■

Lemma A.4. The probability limit

$$\mathbf{M}_{\phi\alpha} = \text{plim}_{T\to\infty}(1/T)\sum_{t=1}^{T}\mathbf{M}_{\phi\alpha t} = \lim_{T\to\infty}(1/T)\sum_{t=1}^{T}\boldsymbol{\delta}^*(t-1)\mathbf{g}'_{\alpha t},\ \ \alpha = \delta, \gamma, c. \tag{23}$$

The result follows from (20) and applying the Kolmogorov LLN 1 to

$$\{(1/T)\sum_{t=1}^{T} g_{\alpha kt}(\sum_{m=0}^{\infty}\theta_m\varepsilon_{t-m-j})\}$$

$k = 1, ..., q$; $j = 1, ..., p$, where $\mathsf{E}g_{\alpha kt}(\sum_{m=0}^{\infty}\theta_m\varepsilon_{t-m-j}) = 0$ for all t. All the elements in (23) are finite. ■

Theorem: Let $\widehat{\boldsymbol{\theta}}_T$ be the maximum likelihood estimator of $\boldsymbol{\theta}_0$,

$$\widehat{\boldsymbol{\theta}}_T = \arg\max L_T(\boldsymbol{\theta}, \boldsymbol{\varepsilon})$$

where $L_T(\boldsymbol{\theta}, \boldsymbol{\varepsilon})$ is defined in (18) and (19). Then,

$$T^{1/2}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1}(\boldsymbol{\theta}_0)\mathbf{B}(\boldsymbol{\theta}_0)\mathbf{A}^{-1}(\boldsymbol{\theta}_0))$$

where

$$\mathbf{A}(\boldsymbol{\theta}_0) = -\text{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{T}\frac{\partial^2\ell(\boldsymbol{\theta}, \varepsilon_t)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

and

$$\mathbf{B}(\boldsymbol{\theta}_0) = \text{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{T}\frac{\partial\ell(\boldsymbol{\theta}, \varepsilon_t)}{\partial\boldsymbol{\theta}}\frac{\partial\ell(\boldsymbol{\theta}, \varepsilon_t)}{\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

We prove the results by verifying the assumptions of Theorem 4.1.6 in Amemiya (1985).

Lemma A.5 [Thm 4.1.3, Assumption (A)]. $L_T(\boldsymbol{\theta}, \varepsilon_t)$ continuous in Θ for each $\boldsymbol{\varepsilon}$.

Proof. From Lemma A.1 and the fact that $\delta(t)$ is continuous in $\boldsymbol{\theta}$, it follows that $\ell(\boldsymbol{\theta}, \varepsilon_t)$ is continuous in Θ for each ε_t and thus the same is true for $L_T(\boldsymbol{\theta}, \varepsilon_t)$. ■

Lemma A.6 [Thm 4.1.3, Assumption (B)]. The average Hessian

$$T^{-1}\mathbf{H}_T(\boldsymbol{\theta},\boldsymbol{\varepsilon}) = T^{-1}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}L_T(\boldsymbol{\theta},\boldsymbol{\varepsilon}) = T^{-1}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\sum_{t=1}^{T}\ell(\boldsymbol{\theta},\varepsilon_t)$$

converges to a finite nonsingular matrix $\mathbf{A}(\boldsymbol{\theta}_0)$ for any sequence $\boldsymbol{\theta}_T^*$ such that $\text{plim}_{T\to\infty}\boldsymbol{\theta}_T^* = \boldsymbol{\theta}_0$.

Proof. A straightforward calculation yields

$$T^{-1}\mathbf{H}_T(\boldsymbol{\theta},\boldsymbol{\varepsilon}) = -(\sigma^2 T)^{-1}\sum_{t=1}^{T}\mathbf{g}_t(\boldsymbol{\theta})\mathbf{g}_t'(\boldsymbol{\theta}) + (\sigma^2 T)^{-1}\sum_{t=1}^{T}\varepsilon_t\frac{\partial^2\mathbf{g}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \quad (24)$$

Now, each $\mathbf{M}_{\alpha\beta t} = [g_{\alpha it}g_{\beta jt}]$, $\alpha,\beta = \delta,\gamma,c$, is a matrix of finite elements and $\{T^{-1}\sum_{t=1}^{T}g_{\alpha it}g_{\beta jt}\}$ is a Cesàro summable sequence. It follows that

$$\mathbf{M}_{\alpha\beta} = \lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\mathbf{M}_{\alpha\beta t} \quad (25)$$

is finite. From Lemma A3 it follows that $\mathbf{M}_{\phi\phi} = \text{plim}_{T\to\infty}(1/T)\sum_{t=1}^{T}\mathbf{M}_{\phi\phi t}$ is finite, and Lemma A.4 contains a similar result for $\{(1/T)\sum_{t=1}^{T}\mathbf{M}_{\phi\alpha t}\}$, $\alpha = \delta,\gamma,c$. By Kolmogorov's LLN 1 and uniform convergence of (24),

$$\text{plim}_{T\to\infty}T^{-1}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}L_T(\boldsymbol{\theta},\boldsymbol{\varepsilon})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -\lim_{T\to\infty}(\sigma^2 T)^{-1}\sum_{t=1}^{T}\mathbf{g}_t(\boldsymbol{\theta}_0)\mathbf{g}_t'(\boldsymbol{\theta}_0) = \mathbf{A}(\boldsymbol{\theta}_0) \quad (26)$$

because $\ell(\boldsymbol{\theta},\varepsilon_s)$ and $\ell(\boldsymbol{\theta},\varepsilon_t)$ are independent for $s \neq t$ and because $\ell(\boldsymbol{\theta},\varepsilon_t)$ is twice continuously differentiable for all $\boldsymbol{\theta}$ and every ε_t. $\mathbf{A}(\boldsymbol{\theta}_0)$ is finite because every element of $\mathbf{M}_{\alpha\beta}$, $\alpha,\beta = \delta,\gamma,c$, is finite, and positive definite because $\mathbf{g}_t(\boldsymbol{\theta}_0)\mathbf{g}_t'(\boldsymbol{\theta}_0)$ is positive semidefinite.

Since the convergence of (24) is uniform and continuous, applying Theorem 4.1.5 in Amemiya (1985, p. 113) yields

$$\text{plim}_{T\to\infty}T^{-1}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}L_T(\boldsymbol{\theta},\boldsymbol{\varepsilon})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_T^*} = \text{plim}_{T\to\infty}(\sigma^2 T)^{-1}\sum_{t=1}^{T}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ell_T(\boldsymbol{\theta},\varepsilon_t)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

for any sequence $\{\boldsymbol{\theta}_T^*\}$ such that $\text{plim}_{T\to\infty}\boldsymbol{\theta}_T^* = \boldsymbol{\theta}_0$. From (26) it follows that

$$\text{plim}_{T\to\infty}T^{-1}\mathbf{H}_T(\boldsymbol{\theta}_T^*,\boldsymbol{\varepsilon}) = -\text{lim}_{T\to\infty}(\sigma^2 T)^{-1}\sum_{t=1}^{T}\mathbf{g}_t(\boldsymbol{\theta}_0)\mathbf{g}_t'(\boldsymbol{\theta}_0) = \mathbf{A}(\boldsymbol{\theta}_0)$$

for $\boldsymbol{\theta}_T^* \to \boldsymbol{\theta}_0$. $\mathbf{A}(\boldsymbol{\theta}_0)$ is a negative definite matrix [Thm 4.1.6, Assumption (B)].■

Let $\mathbf{s}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \partial L_T(\boldsymbol{\theta}, \boldsymbol{\varepsilon})/\partial \boldsymbol{\theta}$ be the score of (18). We have

Lemma A.7 [Thm 4.1.3, Assumption C].

$$\begin{aligned} T^{-1/2}\mathbf{s}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) &= T^{-1/2}\sum_{t=1}^{T}\{\partial \ell(\boldsymbol{\theta},\varepsilon_t)/\partial \boldsymbol{\theta}\}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &\xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{B}(\boldsymbol{\theta}_0)). \end{aligned}$$

Proof. From Lemma A.1 one obtains

$$T^{-1/2}\mathbf{s}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = T^{-1/2}\sum_{t=1}^{T}\frac{\varepsilon_t}{\sigma^2}\mathbf{g}_t(\boldsymbol{\theta}).$$

We have

$$\mathsf{E}\{\partial \ell(\boldsymbol{\theta},\varepsilon_t)/\partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\} = \mathbf{0}$$

and

$$\begin{aligned} \mathrm{cov}\{\partial \ell(\boldsymbol{\theta}_0,\varepsilon_t)/\partial \boldsymbol{\theta}\} &= \mathsf{E}\frac{\varepsilon_t^2}{\sigma^4}\mathsf{E}\mathbf{g}_t(\boldsymbol{\theta})\mathbf{g}_t'(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= (\sigma^2 T)^{-1}\mathsf{E}\mathbf{g}_t(\boldsymbol{\theta}_0)\mathbf{g}_t'(\boldsymbol{\theta}_0). \end{aligned}$$

Then

$$\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathrm{cov}\{\partial \ell(\boldsymbol{\theta},\varepsilon_t)/\partial \boldsymbol{\theta}\}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \lim_{T\to\infty}(\sigma^2 T)^{-1}\sum_{t=1}^{T}\mathsf{E}\mathbf{g}_t(\boldsymbol{\theta}_0)\mathbf{g}_t'(\boldsymbol{\theta}_0).$$

Let

$$x_{jt} = \partial \ell(\boldsymbol{\theta},\varepsilon_t)/\partial \theta_j|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

so $\mathsf{E}x_{jt} = 0$ and $\mathrm{var}(x_{jt}) = \mathsf{E}x_{jt}^2 = \mathsf{E}\{\partial \ell(\boldsymbol{\theta},\varepsilon_t)/\partial \theta_j\}^2|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$. Next assume that

$$\max_{t=1,\ldots,T}|x_{jt}|^{2+\phi} = O_p(1). \tag{27}$$

It follows that the p-norm

$$||x_{jt}||_{2+\phi} = (\mathsf{E}|x_{jt}|^{2+\phi})^{1/(2+\phi)} = O(1).$$

From (27) we have

$$x_{jt}^2 = \{\partial\ell(\boldsymbol{\theta},\varepsilon_t)/\partial\theta_j|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\}^2 < \infty$$

for all t and nearly all of these terms are positive as $T \to \infty$. It then follows that

$$\overline{\sigma}_{jT} = \{T^{-1}\sum_{t=1}^{T}\text{var}(x_{jt})\}^{1/2} = O(1)$$

and, consequently, for some $\phi > 0$,

$$\frac{\max_{j=1,...,T}||x_{jt}||_{2+\phi}}{\overline{\sigma}_{jT}} \leq M < \infty,\ T \geq 1.$$

Thus x_{jt}, which is martingale difference sequence with respect to the conditioning information $\mathcal{F}_{t-1}$ defined by the structure of the likelihood, satisfies the assumptions of Theorems 6.2.2 and 6.2.3 in Davidson (2000, pp. 123-124), which proves that $(T^{-1/2}\sum_{t=1}^{T}x_{jt})/\overline{\sigma}_{jT} \xrightarrow{D} x_j \sim \mathcal{N}(0,1)$, $j = 1,...,3q+1$. It follows that for all linear combinations $\boldsymbol{\lambda}'\mathbf{x}_t$ with $\boldsymbol{\lambda} \neq \mathbf{0}$ one obtains $\boldsymbol{\lambda}'\mathbf{x}_t \xrightarrow{D} \boldsymbol{\lambda}'\mathbf{x}$, where $\mathbf{x} = (x_1,...,x_{3q+1})'$. Theorems 3.3.3 and 3.3.4 in Davidson (2000, p. 46) then yield $\mathbf{x} \sim \mathcal{N}(\mathbf{0},\mathbf{B}(\boldsymbol{\theta}_0))$. ■

Lemma A.8 [Thm 4.1.6, Assumption (A)]. Function $T^{-1}L_T(\boldsymbol{\theta},\boldsymbol{\varepsilon})$ converges to a nonstochastic function $L(\boldsymbol{\theta})$ in probability uniformly in $\boldsymbol{\theta}$ (in a neighbourhood of $\boldsymbol{\theta}_0$).

Proof. We have

$$T^{-1}L_T(\boldsymbol{\theta}_0,\boldsymbol{\varepsilon}) = T^{-1}L_T(\boldsymbol{\theta}_0,\boldsymbol{\varepsilon}) + T^{-1}\frac{\partial}{\partial\boldsymbol{\theta}}L_T(\boldsymbol{\theta},\boldsymbol{\varepsilon})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + (2T)^{-1}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}L_T(\boldsymbol{\theta},\boldsymbol{\varepsilon})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$$

where each element of $\boldsymbol{\theta}^*$ lies in the interval joining the corresponding elements of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. Then

$$\begin{aligned} T^{-1}L_T(\boldsymbol{\theta}_0,\boldsymbol{\varepsilon}) &\to k - (1/2)\ln\sigma^2 - \text{plim}_{T\to\infty}(1/2T)\sum_{t=1}^{T}\frac{\varepsilon_t^2}{\sigma^2} \\ &\quad +\text{plim}_{T\to\infty}(2T)^{-1}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}L_T(\boldsymbol{\theta},\boldsymbol{\varepsilon})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \\ &= k - (1/2)\ln\sigma^2 - (\sigma_0^2/2\sigma^2) + (1/2)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\mathbf{H}(\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) \end{aligned}$$

where

$$\mathbf{H}(\boldsymbol{\theta}^*) = \text{plim}_{T\to\infty}(1/T)\sum_{t=1}^{T}\frac{\partial^2\varepsilon_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}.$$

This follows from the Kolmogorov LLN 2 applied to the sequence of independent, identically distributed variables ε_t^2, with $\mathsf{E}\varepsilon_t^2 = \sigma_0^2$. The convergence is uniform because ε_t^2 is continuous for all $\boldsymbol{\theta} \in \Theta$. ■

Lemma A.9 [Thm 4.1.6, Assumption (C)]. The probability limit

$$\text{plim}_{T\to\infty} T^{-1}\mathbf{H}_T(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \ell_T(\boldsymbol{\theta}, \boldsymbol{\varepsilon})$$

exists and is continuous in a neighbourhood of $\boldsymbol{\theta}_0$.

Proof. The probability limit of the average Hessian is given in (26). It exists and is continuous for all $\boldsymbol{\theta}$. The continuity is a consequence of the fact that $g_t(\boldsymbol{\theta})$ is bounded and infinitely many times differentiable in Θ. ■

When $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, (26) becomes

$$\text{plim}_{T\to\infty} T^{-1}\mathbf{H}_T(\boldsymbol{\theta}_0, \boldsymbol{\varepsilon}) = \mathbf{A}(\boldsymbol{\theta}_0).$$

which is a negative definite matrix.

Proof of Theorem. The result follows from the fact that $\mathbf{A}(\boldsymbol{\theta}_0)$ is negative definite and from Lemmata A.5– A.9. ■