Second Order Time Dependent Inflation Persistence in the United States: a GARCH-in-Mean Model with Time Varying Coefficients

A. Canepa
University of Turin
and Brunel University London

M. G. Karanasos*
Brunel University London*

A. G. Paraskevopoulos
University of Piraeus

November 27, 2019

Abstract

In this paper we investigate the behavior of inflation persistence in the United States. To model inflation we estimate an autoregressive GARCH-in-mean model with variable coefficients and we propose a new measure of second-order time varying persistence, which not only distinguishes between changes in the dynamics of inflation and its volatility, but it also allows for feedback from nominal uncertainty to inflation. Our empirical results suggest that inflation persistence in the United States is best described as unchanged. Another important result relates to the Monte Carlo experiment evidence, which reveals that if the model is misspecified, then commonly used unit root tests will misclassify inflation as being a nonstationary, rather than a stationary process.

Keywords: GARCH-in Mean, Inflation persistence, Monte Carlo simulations, Optimal forecasts, Structural breaks.

JEL Classification: C13, C22, C32, E17, E31, E5

*Address for correspondence: Menelaos Karanasos, Economics and Finance, Brunel University London, UB8 3PH, UK; email: menelaos.karanasos@brunel.ac.uk, tel: +44(0)1895265284, fax: +44 (0)1895269770.

*†Address for correspondence: Menelaos Karanasos, Economics and Finance, Brunel University London, UB3 3PH, UK; email: menelaos.karanasos@brunel.ac.uk, tel: +44(0)1895265284, fax: +44 (0)1895269770.

1 Introduction

The behavior of inflation has long been an object of interest to economists, but especially to central banks, which are bounded by statutory mandate to maintain price stability, thus promoting sustainable growth. A critical aspect of inflation is persistence, this is because the degree of inflation persistence determines to what extent monetary policy authorities can control inflation. Broadly speaking, inflation persistence measures the speed at which the inflation rate returns to its equilibrium level after an inflationary shock: the faster inflation returns to its equilibrium level after a macroeconomic shock, the more effective monetary policy action can be, all else equal. As a result, optimal monetary policy crucially depends on the knowledge of inflation dynamics. For example, high inflation persistence may require a bolder monetary policy action to bring inflation under control. On the other side, a low level of inflation persistence may require a weaker or no action by monetary authorities in response to an exogenous shock.

In the light of their mandate Central Banks are also interested in addressing the issue of whether inflation persistence varies over time. Changes in the structural characteristics of the economy or monetary policy actions can affect the characteristic features of the stochastic process, generating time-variation in inflation persistence. There is now growing consensus in the literature that substantial changes over time in monetary regimes may leave econometric models exposed to the Lucas critique (see for example Taylor, 2000 and the references therein).

The literature on inflation persistence in the United States is quite extensive, but there is considerable disagreement regarding the empirical findings. From the methodological point of view two strands of literature are clearly defined, depending on the measure of persistence. Empirical work in the first strand relies on the analysis of the order of integration of the process as the measure of inflation persistence, using unit root tests to classify inflation as either an I(0) or an I(1) process. For example Chandler and Polonik (2006), Beran (2009), and Palma and Olea (2010) find strong evidence for nonstationarity in the U.S. inflation. On the other hand, Rose (1988) indicated that monthly U.S. inflation was an I(0) process from 1947 to 1986. Mixed evidence was provided by Brunner and Hess (1993). They concluded that the inflation rate was I(0) before the 1960's but that it is characterized as I(1) since that time. Other studies include Barsky (1987), Ball and Cecchetti (1990) and Kim *et al.*(2004) among others. A second strand of literature uses autoregressive (AR) model-based measures such as the largest autoregressive root (LAR) and the sum of the autoregressive coefficients (SAR) to measure persistence. For instance, in his seminal paper Taylor (2000) concluded that U.S. inflation persistence during the Volcker-Greenspan era was substantially lower than during the previous two decades. Similarly, Levin and Piger (2003) showed that high inflation persistence is not an inherent characteristic of industrial economies over the period 1984-2002. On the other hand, the work based on the SAR approach by Batini (2006) suggests relatively little evidence of shifts in inflation persistence for the Euro area.

With this background in mind, in this paper we try and reconcile disagreements in the conclusions of empirical studies by considering a model that allows for time varying parameters for both the intrinsic and uncertainty persistence. We also argue that, perhaps, one possible explanation of divergent results of so many empirical works is the presence of structural breaks in the inflation process that would bias the findings for most of the commonly used unit root tests. From the theoretical point of view whether inflation follows a stationary or nonstationary process has important theoretical implications. In the literature textbook treatments of inflation, such as Blanchard (2000) for example, assume that inflation is stationary. Also, in their seminal paper Blanchard and Gali (2007) suggest that inflation persistence captures structural characteristics of the economy that do not likely respond to policy actions, which implies that a policy of inflation targeting should exert no effect on inflation persistence. On the other side, the works by Cogley and Sargent (2005), Beechey and Österholm (2009), and Cogley and Sbordone (2008) support the view that inflation persistence varies across monetary regimes, therefore supporting the Lucas critique.

The contribution of this paper is threefold. First, we extend the literature on inflation persistence along the second strand of the literature by proposing a new measure of inflation persistence. Unlike most related studies our measure of persistence is grounded on economic, rather than statistical theory. In particular, we estimate an autoregressive (AR) asymmetric power (AP) GARCH in-mean (M) model with variable coefficients and we compute a measure of second-order time varying persistence, which not only distinguishes between changes in the dynamics of inflation and its volatility, but it also allows for feedback from nominal uncertainty to inflation.

In the related literature empirical works that document changes over time of inflation persistence in the United States are the works by Cogley and Sargent (2001, 2005), where a Bayesian state-space VAR model is used to model inflation dynamics. The authors conclude that there has been a change in the underlying characteristics of inflation reflecting a change in the structural characteristics of the economy and, possibly, a more active inflation targeting policy. In sharp contrast, Stock (2001) estimates the LAR using a rolling window estimation method and concludes that there is no indication of a marked decline in the persistence. A similar result is found in Pivetta and Reis (2007), where the LAR and the SAR are estimated using both Bayesian and rolling window estimation methods. The authors conclude that inflation persistence has been high in the United States and approximately unchanged over the entire post-war period.

Modeling inflation crucially relies on the properties of the inflation process. Accordingly, the second task of this paper is to fill a gap in the first strand of the literature by investigating to what extent commonly used unit root tests are robust to structural breaks in the times series process. It is well known that the performance of such tests depends on a number of factors that are not easily observed by applied economists trying to discriminate between stationarity and non stationarity. In addition, empirical research has often found evidence of GARCH effects with highly persistent volatility in situations where the conditional second moment affects the level of the series. However, the performance of the unit root tests for these types of stochastic processes has not been widely investigated. Thus, we consider a 'time varying' AR-APGARCH-M specification and we carry out an extensive Monte Carlo experiment in order to examine the size and power of these tests in the presence of abrupt breaks in the in-mean parameter. The results indicate that the performance of the test statistics under consideration is severely affected by these breaks. The above considerations reinforce the argument (and extend it to a dynamic environment) made by Conrad and Karanasos (2015a) that conventional time invariant measures of persistence, such as unit roots, might result in misleading conclusions regarding the persistence in the level. Similarly, it is well known that unexpected shifts in a time series can lead to huge forecasting errors and unreliability of the model in general. Therefore, in a companion exercise we use simulated data to evaluate the reliability of the out-of-sample forecasts in the context of the AR-APGARCH-M specification with abrupt breaks and find that the location and the magnitude of the breaks severely affects the forecasting performance of the models.

Research over the past decade has documented considerable instability in inflation forecasting models, see for example Stock and Watson (2007) or Stock and Watson (2009) for an excellent survey on the related literature. This instability has created major headaches for inflation forecasters. Accordingly, the third contribution of this work relates to the forecasting properties of the proposed model.

The outline of the paper is as follows. Section 2 introduces the model and some related literature. Section 3 considers the performance of commonly used unit root tests when the data generating process is an AR-APGARCH-M process with deterministic abrupt breaks. Section 4 derives the optimal forecasts and the second moments of this construction which we utilize in order to obtain a new time varying measure of second-order persistence. Section 5 presents an empirical study on inflation persistence in the United States. Finally, Section 6 presents some concluding remarks.

2 Theory and Model

2.1 Theoretical Background

In the literature economists have placed considerable emphasis on the impact of inflation uncertainty on both inflation and output growth. Friedman (1977) states that nominal uncertainty causes an adverse output effect. This argument is based on the viewpoint that uncertainty about future inflation distorts the allocative efficiency aspect of the price mechanism (for details, see for example, Fountas *et al.*, 2006; or Fountas and Karanasos, 2007). Following the influential work of Friedman a rich literature highlights the importance of nominal uncertainty for macroeconomic modeling and policy making. In particular, according to Cukierman and Meltzer (1986) in the presence of uncertainty about the rate of monetary growth and, therefore, inflation, the policymaker applies an expansionary monetary policy in order to surprise the agents and enjoy output gains. The argument that Central Banks tend to create inflation surprises in the presence of more inflation uncertainty (hereafter, termed the Cukierman and Meltzer hypothesis) implies a positive causal effect from inflation uncertainty to inflation (see for example, Conrad and Karanasos, 2015b).

One of the first papers to test for the Cukierman and Meltzer hypothesis in a context of a GARCH-M model was Baillie *et al.* (1996); see also Brunner and Hess (1993). However, the econometric specifications which are employed in these studies do not take into consideration the time dependent characteristics of inflation. Time variation may explain why inflation in the United States has become harder to be modeled and forecasted in recent years (see, for example, Stock and Watson, 2007). In this respect, our work is more closely related to the econometric framework in Evans (1991) and recently used by Berument *et al.* (2005), Caporale and Kontonikas (2009) and Caporale *et al.* (2010) where time varying AR–APGARCH-M models are estimated to accommodate for structural changes in the economy and the resulting shifts in the private sector behavior.

2.2 The Model

In this section, we consider an AR(1) process with GARCH$(1,1)$-in mean effects, that is, a model in which the conditional variance affects the conditional mean, and two deterministic abrupt breaks (hereafter, DAB-AR$(1;2)$-M model). In particular, we will examine the case of two breaks ($N=2$) which occur at times $t-k_1$ and $t-k_2$ (with $k_2>k_1$, $k_2\in\mathbb{Z}_{>0}$ (the set of positive integers)); of course when $k_2=k_1$ we have the case of one break), where the switch from one set of parameters to another is abrupt. The time invariant version of the model was introduced by Engle *et al.* (1987) and applied in Glosten *et al.* (1993), Christensen and Nielsen (2007) and Conrad and Karanasos (2015a), among others.

The DAB-AR$(1;2)$-M model is given by

$$y_t=\varphi(t)+\phi(t)y_{t-1}+\varsigma(t)\sigma_t^{\delta}+\varepsilon_t, \tag{1}$$

where $\varepsilon_t=e_t\sigma_t$, and the vector of the three deterministically varying coefficients, $\mathbf{m}(\tau)'=(\varphi(\tau),\ \phi(\tau),\ \varsigma(\tau))$ is given by

$$\mathbf{m}(\tau)'=\begin{cases}(\varphi_1,\phi_1,\varsigma_1) & \text{if} \quad \tau>t-k_1,\\ (\varphi_2,\phi_2,\varsigma_2) & \text{if} \quad t-k_2<\tau\leq t-k_1,\\ (\varphi_3,\phi_3,\varsigma_3) & \text{if} \quad \tau\leq t-k_2.\end{cases}$$

with φ_n, ϕ_n, $\varsigma_n\in\mathbb{R}$ (the set of real numbers), $n=1,2,3$, $\delta\in\mathbb{R}_{>0}$ (the set of positive real numbers), $\{e_t\}$ is a sequence of independent and identically distributed (*i.i.d*) random variables with zero mean and variance, $\mathbb{E}(e_t^2)$, and σ_t^2 is the conditional variance of y_t. [1] The time dependent autoregressive coefficient $\phi(t)$ naturally measures the intrinsic persistence in the level of y_t. By including σ_t^{δ} in the conditional mean we allow for feedback from the power transformed conditional variance of y_t to its level, captured by the deterministically varying in-mean coefficient $\varsigma(t)$. We denote the size of the breaks by $\Delta\phi_n=\phi_n-\phi_{n-1}$ and $\Delta\varsigma_n=\varsigma_n-\varsigma_{n-1}$, for $n=2,3$. For example, $\phi_2=\phi_3-\Delta\phi_3$ and $\phi_1=\phi_3-\Delta\phi_3-\Delta\phi_2$.

[1] Within the class of ARMA processes this specification is quite general and allows for intercept and slope shifts (see also Pesaran and Timmermann, 2005, Pesaran et al., 2006, and Koop and Potter, 2007).

The power transformed conditional variance, σ_t^δ, is positive with probability one and is a measurable function of $\mathcal{F}_{t-1}$, which in turn is the sigma-algebra generated by $\{y_{t-1}, y_{t-2}, \ldots\}$. We assume that σ_t^δ is specified as a time invariant APGARCH$(1,1)$ process:

$$(1-\beta B)\sigma_t^\delta = \omega + \alpha f(\varepsilon_{t-1}), \tag{2}$$

with

$$f(\varepsilon_{t-1}) = (|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1})^\delta,$$

where $|\gamma| < 1$ (for the APGARCH model with time invariant parameters see, for example, Ding *et al.*, 1993, and Karanasos and Kim, 2006). The following conditions are necessary and sufficient for $\sigma_t^\delta > 0$, for all t: $\omega > 0$, $\alpha, \beta \geq 0$.

Next we will introduce some important notation.

Notation 1 *i) We denote the time invariant r-th moment ($r \in \mathbb{Z}_{>0}$) of the power transformed variance by $\mu_r = \mathbb{E}(\sigma_t^{\delta r})$.*
ii) Similarly, κ_r denotes the r-th moment of $f(e_t)$: $\kappa_r = \mathbb{E}[[f(e_t)]^r]$.

Clearly for $\delta \geq 1$, $\mu_{2/\delta} = \mathbb{E}(\sigma_t^2)$ is not a fractional moment only if δ is equal to 1 or 2. In all other cases $\mu_{2/\delta}$ has to be calculated numerically. However, if $\delta > 2$, the existence of the first moment, μ_1 guarantees that of $\mu_{2/\delta}$. Similarly, $\mu_{1+1/\delta} = \mathbb{E}(\sigma_t^{\delta+1})$ is not a fractional moment only if $\delta = 1/\lambda$ where $\lambda \in \mathbb{Z}_{>0}$. In all other cases $\mu_{1+1/\delta}$ has to be calculated numerically.

The APGARCH$(1,1)$ formulation in eq. (2) can readily be interpreted as having an ARMA$(1,1)$ representation for the conditional variance:

$$(1-cB)\sigma_t^\delta = \omega + \alpha v_{t-1}, \tag{3}$$

where

$$c = \alpha\kappa_1 + \beta, \quad \text{and} \quad v_t = f(\varepsilon_t) - \mathbb{E}[f(\varepsilon_t)\,|\mathcal{F}_{t-1}] = f(\varepsilon_t) - \kappa_1\sigma_t^\delta,$$

and v_t is, by construction, an uncorrelated term with expected value 0. While the ε_t are the innovations to the level of y_t, the v_t can be considered the 'innovations' to the power transformed conditional variance of y_t. Note that the parameter c measures the *intrinsic* memory or persistence in the conditional variance.

Next we will define the covariance matrix of the two 'shocks' ε_t and v_t, $\mathbf{\Sigma} = \mathbf{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$, where $\mathbf{E}(\cdot)$ denotes the elementwise expectation operator. First, we will denote the variances of the two 'shocks' and their covariance by

$$\sigma_\varepsilon = \mathbb{E}(\varepsilon_t^2), \sigma_v = \mathbb{E}(v_t^2), \sigma_{\varepsilon v} = \mathbb{E}(\varepsilon_t v_t).$$

The covariance matrix $\mathbf{\Sigma}$ is given by

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_\varepsilon & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v \end{bmatrix} = \begin{bmatrix} \mu_{2/\delta}\mathbb{E}(e_t^2) & \mu_{1+1/\delta}\widetilde{\kappa} \\ \mu_{1+1/\delta}\widetilde{\kappa} & \mu_2\kappa \end{bmatrix}, \tag{4}$$

where

$$\kappa = (\kappa_2 - \kappa_1^2), \ \widetilde{\kappa} = \mathbb{E}[e_t f(e_t)].$$

In the following corollary we present expressions for κ_r and $\widetilde{\kappa}$ under the assumption of Normality (see also Karanasos and Kim, 2006).

Corollary 1 *Consider the case where the term e_t is standard normal. Then $\mathbb{E}(e_t^2) = 1$, and κ_r, $\widetilde{\kappa}$ are given by*

$$\kappa_r = \frac{1}{\sqrt{\pi}}\left[(1-\gamma)^{r\delta} + (1+\gamma)^{r\delta}\right]2^{(\frac{r\delta}{2}-1)}\Gamma\left(\frac{r\delta+1}{2}\right),$$
$$\widetilde{\kappa} = \frac{1}{\sqrt{2\pi}}\left[[1-\gamma]^{\delta} - [1+\gamma]^{\delta}\right]2^{(\delta/2)}\Gamma\left(\frac{\delta}{2}+1\right),$$

where $\Gamma(\cdot)$ is the Gamma function.
When $\delta = 1$ the above expressions reduce to $\widetilde{\kappa} = -\gamma$, $\kappa_1 = \sqrt{\frac{2}{\pi}}$, $\kappa_2 = 1+\gamma^2$ and therefore $\kappa = (\kappa_2 - \kappa_1^2) = 1+\gamma^2 - \frac{2}{\pi}$, which implies that $\mathbf{\Sigma}$ becomes

$$\mathbf{\Sigma} = \mu_2 \begin{bmatrix} 1 & -\gamma \\ -\gamma & 1+\gamma^2-\frac{2}{\pi} \end{bmatrix}. \tag{5}$$

Having defined the deterministically varying extension of the AR-APGARCH-M model, in the next section we will present its bivariate vector autoregressive moving average (BVARMA) formulation.

2.3 VAR Formulation

To obtain the optimal predictors and the variance of y_t for the DAB-AR-M model in eqs. (1) and (2) in the next lemma we will express eqs. (1) and (3) in a matrix form.

Lemma 1 *Eqs. (1) and (3) can be expressed in a matrix form as*

$$\mathbf{y}_\tau = \boldsymbol{\varphi}(\tau) + \mathbf{\Phi}(\tau)\mathbf{y}_{\tau-1} + \mathbf{J}\boldsymbol{\varepsilon}_\tau + \mathbf{Z}(\tau)\boldsymbol{\varepsilon}_{\tau-1}, \tag{6}$$

with $\mathbf{y}_\tau = (y_\tau \ \sigma_\tau^\delta)'$, $\boldsymbol{\varepsilon}_\tau = (\varepsilon_\tau \ v_\tau)'$, $\mathbf{J} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, where the three time varying coefficient matrices, $\boldsymbol{\varphi}(\tau)$, $\mathbf{\Phi}(\tau)$, and $\mathbf{Z}(\tau)$ are time invariant in each of the three segments:

$$\boldsymbol{\varphi}_n = \begin{bmatrix} \varphi_n + \varsigma_n\omega \\ \omega \end{bmatrix}, \mathbf{\Phi}_n = \begin{bmatrix} \phi_n & \varsigma_n c \\ 0 & c \end{bmatrix}, \mathbf{Z}_n = \begin{bmatrix} 0 & \varsigma_n\alpha \\ 0 & \alpha \end{bmatrix}, \begin{cases} n=1 & \text{if} \quad \tau > t-k_1, \\ n=2 & \text{if} \quad t-k_2 < \tau \le t-k_1, \\ n=3 & \text{if} \quad \tau \le t-k_2. \end{cases}$$

For notational convenience we will interchangeably use $\mathbf{\Phi}_3$ or $\mathbf{\Phi}$ and $\mathbf{Z}_3$ or $\mathbf{Z}$. We will term the deterministically varying bivariate expression in eq. (6) the DAB-BVARMA$(1,1;2)$ representation.[2]

In what follows we will employ the above representation to derive explicit formulas for the optimal predictors and the variance of y_t and σ_t^δ in eqs. (1) and (2), respectively.[3] These are needed in order to obtain time varying first and second-order measures of persistence. But first, since the LAR has been commonly used as a measure of persistence in the context of testing for the presence of unit roots, we will use Monte Carlo simulations to examine the performance of unit root tests when the data are generated from an AR-(APGARCH) M process with unknown structural breaks in the in-mean coefficient.

3 Monte Carlo Experiment

A decision whether a series is treated as integrated of order zero, $I(0)$, or $I(1)$ has important implications for the subsequent modeling, hypothesis testing and forecasting. A frequent criticism of unit root tests concerns the poor power and size properties that many such tests exhibit. Since standard unit root tests are based on the assumption that some type of heteroscedasticity is present but ignore the possibility that the volatility has a direct impact on the level, we investigate the size and power properties of common unit root tests in the presence of GARCH-M effects and unknown structural breaks in the in-mean parameter.

The two unit root tests considered are the Dickey-Fuller test (DF) proposed by Dickey and Fuller (1981) and the M test proposed by Sims *et al.* (1990) and Perron and Ng (1996). As far as the estimation of the autoregressive parameter ϕ is concerned both the ordinary least squared method (OLS) and the generalized least squared method (GLS) suggested by Elliott *et al.* (1996) are considered. This gives us two DF statistics, which we define as DF_{OLS} or DF_{GLS} depending on the estimation method used for ϕ. Likewise, the M tests are defined as M_{OLS} and M_{GLS} respectively.

To examine the properties of these tests we consider the DAB-AR$(1;2)$-M model (data generating process, DGP) in eqs. (1) and (2) for the Monte Carlo simulation experiment where

$$\varphi(t) = \phi(t) = 1 \text{ for all } t,\ \delta = 1,\ \omega = 1 - \alpha - \beta,\ \alpha = 0.1,\ \beta = 0.70,\ \gamma = 0, \tag{7}$$

and there are two abrupt breaks in the time varying in-mean coefficient, $\varsigma(t)$, at times $t-k_1$ and $t-k_2$. In particular, $\varsigma(\tau) = \varsigma_1$ for $\tau < t - k_2$ and $\tau > t - k_1$, whereas $\varsigma(\tau) = \varsigma_2 = \varsigma_1 + \Delta_\varsigma$ for $t - k_2 \leq \tau \leq t - k_1$. The magnitude of the break is denoted by Δ_ς and the length of the break by $\Delta_k = k_2 - k_1$. Therefore, time variation is caused only by the in-mean coefficient. We also set the sample size k equal to $1,000$. Finally, $\{e_t\}$ are $i.i.d \sim N(0,1)$ random variables.

[2] As pointed out by Conrad and Karanasos (2015a) the AR(1)-[APGARCH$(1,1)$]-M model is observationally equivalent to an ARMA$(2,1)$ process, with the largest autoregressive root (LAR) being close to one. Clearly, if $\phi = 0$, $c = 1$ and there are no breaks the reduced form representation of the AR-M specification coincides with the IMA$(1,1)$ model proposed by Stock and Watson (2007).

[3] Notice that, as pointed ouy by Pivetta and Reis (2007), including other variables would lead to an assessment of predictability. Since here we focus on persistence, not predictability, we work with a univariate GARCH-M model.

3.1 Empirical Sizes

The Monte Carlo simulation experiment design is targeted at investigating the effect of the in-mean breaks on the empirical sizes of the test statistics under consideration. However, as the magnitude of the in-mean parameter itself is likely to affect the performance of the test statistics we investigate this latter issue before considering the former. Accordingly, the Monte Carlo experiment is aimed at investigating the effects on the empirical sizes of *i*) the magnitude of the in-mean parameter, *ii*) the magnitude of the break, Δ_ς, and *iii*) the timing (k_1, k_2) and the length or duration (Δ_k) of the breaks as a fraction of the sample size, k. To address point *i)* a set of simulation experiments was undertaken with the *DGP* in eqs. (1), (2) and (7) with increasing magnitude of the in-mean parameter, namely $\varsigma_1 \in \{0.1, 0.3, 0.9\}$. Similarly, to investigate point *ii*) simulation experiments were undertaken with $\Delta_\varsigma \in \{0.07, 0.25, 0.50\}$ with the case of $\Delta_\varsigma = 0.00$ set as benchmark. Finally, to tackle point *iii*) in the experiment design we considered the above *DGP* with $k_1/k = (k - k_2)/k \in \{0.100, 0.333, 0.450\}$, that is $k_1 = (k - k_2) \in \{100, 333, 450\}$. In other words, we consider three values for the length of the in-mean break: $\Delta_k/k = (k_2 - k_1)/k \in \{0.80, 0.333, 0.10\}$ or $\Delta_k \in \{800, 333, 100\}$.

Note that all experiments were performed over $10,000$ Monte Carlo replications using, as noted earlier, a sample size $k = 1,000$, with a further 50 observations created and discarded in order to avoid the influence of the initial values. The sequence $\{e_t\}$ was generated using pseudo *i.i.d*$\sim N(0,1)$ random numbers from the RNDNS procedure in GAUSS with the value of y_0 set as a $N(0,1)$ random number.

Table 1 reports the results for the empirical sizes of the inference procedures under consideration for the 5% nominal significance level. The top panel reports the empirical sizes resulting from the simulation experiment with the aforementioned *DPG* with $\Delta_k = 800$, whereas the results for $\Delta_k = 333$ and $\Delta_k = 100$ are given in the middle and bottom panel, respectively.

Table 1. Empirical sizes of unit root tests: the case of two unknown breaks in the in-mean parameter.

		DF_{OLS}	DF_{GLS}	M_{OLS}	M_{GLS}
		$\Delta_k = 800$ or $\Delta_k/k = 0.80$			
$\varsigma_1 = 0.1$	$\Delta_\varsigma = 0.00$	0.049	0.054	0.054	0.054
	$\Delta_\varsigma = 0.07$ (ς_2=0.17)	0.048	0.042	0.046	0.040
	$\Delta_\varsigma = 0.25$	0.048	0.037	0.037	0.037
	$\Delta_\varsigma = 0.50$	0.017	0.011	0.012	0.011
$\varsigma_1 = 0.3$	$\Delta_\varsigma = 0.00$	0.049	0.040	0.042	0.040
	$\Delta_\varsigma = 0.07$	0.045	0.028	0.030	0.028
	$\Delta_\varsigma = 0.25$	0.029	0.014	0.014	0.013
	$\Delta_\varsigma = 0.50$	0.012	0.007	0.005	0.006
$\varsigma_1 = 0.9$	$\Delta_\varsigma = 0.00$	0.015	0.001	0.005	0.001
	$\Delta_\varsigma = 0.07$	0.013	0.001	0.001	0.001
	$\Delta_\varsigma = 0.25$	0.016	0.000	0.000	0.000
	$\Delta_\varsigma = 0.50$	0.008	0.000	0.000	0.000
		$\Delta_k = 333$ or $\Delta_k/k = 0.333$			
$\varsigma_1 = 0.1$	$\Delta_\varsigma = 0.07$	0.047	0.045	0.046	0.045
	$\Delta_\varsigma = 0.25$	0.049	0.042	0.043	0.041
	$\Delta_\varsigma = 0.50$	0.020	0.027	0.022	0.025
$\varsigma_1 = 0.3$	$\Delta_\varsigma = 0.07$	0.042	0.030	0.032	0.030
	$\Delta_\varsigma = 0.25$	0.037	0.025	0.024	0.025
	$\Delta_\varsigma = 0.50$	0.013	0.010	0.009	0.010
$\varsigma_1 = 0.9$	$\Delta_\varsigma = 0.07$	0.013	0.001	0.002	0.001
	$\Delta_\varsigma = 0.25$	0.009	0.000	0.000	0.000
	$\Delta_\varsigma = 0.50$	0.003	0.000	0.000	0.000
		$\Delta_k = 100$ or $\Delta_k/k = 0.10$			
$\varsigma_1 = 0.1$	$\Delta_\varsigma = 0.07$	0.049	0.053	0.043	0.053
	$\Delta_\varsigma = 0.25$	0.052	0.038	0.044	0.037
	$\Delta_\varsigma = 0.50$	0.037	0.045	0.046	0.045
$\varsigma_1 = 0.3$	$\Delta_\varsigma = 0.07$	0.048	0.038	0.031	0.038
	$\Delta_\varsigma = 0.25$	0.050	0.036	0.031	0.035
	$\Delta_\varsigma = 0.50$	0.022	0.023	0.019	0.023
$\varsigma_1 = 0.9$	$\Delta_\varsigma = 0.07$	0.013	0.001	0.002	0.001
	$\Delta_\varsigma = 0.25$	0.013	0.000	0.000	0.000
	$\Delta_\varsigma = 0.50$	0.008	0.001	0.001	0.001

Note: The DGP is $y_t = 1 + y_{t-1} + \varsigma(t)\sigma_t + e_t\sigma_t$ and $\sigma_t = 0.2 + 0.1\,|e_{t-1}\sigma_{t-1}| + 0.7\sigma_{t-1}$, where $\varsigma(\tau) = \varsigma_1$ if $\tau > t - k_1$ or $\tau < t - k_2$, and $\varsigma(\tau) = \varsigma_2 = \varsigma_1 + \Delta_\varsigma$ otherwise with $\varsigma_1 \in \{0.1, 0.3, 0.9\}$, $\Delta_\varsigma \in \{0.07, 0.25, 0.50\}$, $k = 1,000$, $k_1 = (k - k_2) \in \{100, 333, 450\}$ or $\Delta_k \in \{800, 333, 100\}$.

Looking at the results in Table 1 we first notice that all inference procedures appear to be robust to small values of the in-mean parameter ($\varsigma_1 = 0.1$) and of the breaks ($\Delta_\varsigma = 0.07$). However, the magnitude of the in-mean parameter appears to have a significant effect on the size distortion of all test statistics as, even when $\Delta_\varsigma = 0.00$, for $\varsigma_1 = 0.9$ all the test statistics are severely undersized. Similarly, both the magnitude and the location of the breaks affect the size properties of the inference procedures under consideration as from the top panel of Table 1 it is clear that the worst case scenario appears to be when $\Delta_k = 800$ and $\Delta_\varsigma \geq 0.25$. In this case the break occurs very early and the stochastic process stays in the second regime for 80% of the time period, only to go back to the first regime for the last 100 observations.[4]

Looking now at the performance of the individual tests, it appears that the *OLS* based test are more robust to regime shifts in the in-mean parameter than the *GLS* based tests, as both DF_{OLS} and M_{OLS} enjoy smaller size distortion than their *GLS* based counterparts.

3.2 Empirical Power

The empirical sizes of the unit root tests presented in Table 1 are constructed to generate a test with asymptotic size of 5% under the null hypothesis of a unit root. We now focus on examining the power of the inference procedures to reject the null hypothesis of $\phi(t) = 1$ for all t when in fact the process is second-order, that is $\phi(t) = \phi$ with $|\phi| < 1$ for all t.

As for the size, the Monte Carlo experiment design is meant to investigate the effects for points *i*) - *iii*) above. With this target in mind, the asymptotic local power functions for the 5% nominal level test have been calculated. To model the sequence of stationary alternatives near the null hypothesis of unit root, we consider the aforementioned *DGP* but now with $\phi(t) = 1 - \frac{l}{k}$ for all t (instead of $\phi(t) = 1$) in eq. (1) where $l = 30, 29, ..., 1, 0$ controling the size of the departure from a unit root.

To investigate the issue in point *i*) simulation experiments were undertaken setting different values of the in-mean parameter under the alternative hypothesis. The simulation results are summarized in Figure 1, where the asymptotic local power curves are plotted for the *DGP* when the magnitude of the parameter is increased from the modest value of $\varsigma_1 = 0.1$ to a relatively large value $\varsigma_1 = 0.9$, with the break parameter fixed at $\Delta_\varsigma = 0$. In the x-axis the value taken by l is reported, whereas in the y-axis the empirical rejection frequencies are reported. Looking at the plot of the asymptotic power curves for the tests under consideration from Figure 1 it appears that all test statistics are sensitive to the magnitude of the in-mean parameter. However, it is clear that DF_{OLS} and M_{OLS} are less sensitive to the magnitude of ς than the *GLS* based counterparts.

Coming to target point *ii*), in Figure 2 we report the results of simulation experiments obtained by fixing the in-mean parameter at 0.9 and $\Delta_k = 800$, then comparing the resulting power curves of the test statistics when $\Delta_\varsigma = 0$ and $\Delta_\varsigma = 0.5$. Interestingly enough, the DF_{OLS} procedure appears to be the most robust to the regime shift of the in-mean parameter. By contrast both *GLS* based statistics are severely affected by the magnitude of the break.

[4] We also find (results not reported) that in the presence of asymmetries the size distortion of the unit root tests is stronger.

Finally, we consider the issue of the timing and duration of the in-mean regime shift as stated in target point *iii*). In this case the simulation experiment was undertaken with $\Delta_k \in \{800, 100\}$ and Δ_ς fixed at the smallest value 0.07. Figure 3 plots the asymptotic local power function for DF_{OLS}, DF_{GLS}, M_{OLS} and M_{GLS} respectively. From the results in Figure 3 it appears that the empirical power of all inference procedures is less affected by the timing and the duration of the regime shift than the size reported in Table 1. Note that in the interest of brevity not all the values of the parameter space considered in Table 1 have been reported, but results are available upon request.

In the next section we use Monte Carlo simulations to examine the out-of-sample performance of the model under consideration.

3.3 Forecasting

In this section we investigate the out-of-sample forecasting performance of the model in eqs. (1)-(2). The DGP was generated by Monte Carlo simulation as explained in Section 3 with $\phi(t) = \phi = 0.8$, $\varsigma_1 = 0.3$ and the other parameters as specified in eq. (7). [5]

In order to investigate the effects of the time varying in-mean parameter the model with $\Delta_\varsigma = 0.00$ was considered as a benchmark and then the magnitude of the break increased as in Table 1. Similarly, the duration of the regime shift was decreased from $\Delta_k = 800$ to $\Delta_k = 100$.

The evaluation of the out-of-sample forecast exercise does not rely on a single criterion; for robustness we compare the results of three different forecasting measures, namely, the mean square error (MSE), the mean absolute error (MAE) and the root mean square forecast error (RMSE). Table 2 reports the results of the forecasting exercise. In columns 1 and 2 the forecasting horizon and the break magnitude under consideration are reported, respectively, whereas in columns 3-8 the forecasting results for the conditional mean and the conditional variance are reported.

[5] See Elliott and Timmermann (2008) for an excellent review on economic forecasting.

Table 2. Forecasting with a DAB-AR(1; 2)-M model. Point predictive performances.

Forecast Horizon	Break size	Conditional Mean			Conditional Variance		
		MSE	MAE	$RMSE$	MSE	MAE	$RMSE$
		$\Delta_k = 800$					
1	$\Delta_\varsigma = 0.00$	0.002	0.040	0.040	0.016	0.127	0.126
	$\Delta_\varsigma = 0.07$	0.002	0.045	0.044	0.018	0.134	0.139
	$\Delta_\varsigma = 0.25$	0.002	0.048	0.048	0.018	0.137	0.137
	$\Delta_\varsigma = 0.50$	0.002	0.050	0.050	0.019	0.137	0.137
5	$\Delta_\varsigma = 0.00$	0.009	0.082	0.098	0.022	0.142	0.148
	$\Delta_\varsigma = 0.07$	0.011	0.084	0.101	0.023	0.145	0.150
	$\Delta_\varsigma = 0.25$	0.012	0.103	0.107	0.024	0.150	0.153
	$\Delta_\varsigma = 0.50$	0.013	0.104	0.118	0.026	0.153	0.155
10	$\Delta_\varsigma = 0.00$	0.117	0.300	0.343	0.026	0.158	0.164
	$\Delta_\varsigma = 0.07$	0.131	0.319	0.361	0.028	0.162	0.162
	$\Delta_\varsigma = 0.25$	0.141	0.333	0.376	0.029	0.169	0.172
	$\Delta_\varsigma = 0.50$	0.174	0.381	0.417	0.032	0.175	0.177
		$\Delta_k = 333$					
1	$\Delta_\varsigma = 0.07$	0.002	0.046	0.046	0.018	0.135	0.135
	$\Delta_\varsigma = 0.25$	0.002	0.048	0.048	0.018	0.136	0.138
	$\Delta_\varsigma = 0.50$	0.002	0.049	0.049	0.019	0.137	0.137
5	$\Delta_\varsigma = 0.07$	0.009	0.083	0.100	0.022	0.144	0.149
	$\Delta_\varsigma = 0.25$	0.010	0.084	0.103	0.023	0.147	0.151
	$\Delta_\varsigma = 0.50$	0.011	0.087	0.107	0.025	0.150	0.153
10	$\Delta_\varsigma = 0.07$	0.135	0.325	0.368	0.027	0.160	0.165
	$\Delta_\varsigma = 0.25$	0.140	0.333	0.375	0.030	0.164	0.168
	$\Delta_\varsigma = 0.50$	0.144	0.336	0.379	0.029	0.168	0.171
		$\Delta_k = 100$					
1	$\Delta_\varsigma = 0.07$	0.002	0.048	0.049	0.018	0.137	0.137
	$\Delta_\varsigma = 0.25$	0.002	0.049	0.049	0.019	0.137	0.137
	$\Delta_\varsigma = 0.50$	0.002	0.050	0.050	0.019	0.137	0.137
5	$\Delta_\varsigma = 0.07$	0.009	0.080	0.092	0.019	0.140	0.143
	$\Delta_\varsigma = 0.25$	0.010	0.082	0.099	0.022	0.144	0.149
	$\Delta_\varsigma = 0.50$	0.011	0.083	0.100	0.027	0.145	0.150
10	$\Delta_\varsigma = 0.07$	0.140	0.332	0.374	0.027	0.159	0.164
	$\Delta_\varsigma = 0.25$	0.143	0.335	0.378	0.027	0.160	0.165
	$\Delta_\varsigma = 0.50$	0.145	0.337	0.380	0.027	0.161	0.166

Note: The DGP is $y_t = 1 + 0.8y_{t-1} + \varsigma(t)\sigma_t + \varepsilon_t$ and $\sigma_t = 0.2 + 0.1|\varepsilon_{t-1}\sigma_{t-1}| + 0.7\sigma_{t-1}$, where $\varsigma(\tau) = \varsigma_1$ if $\tau > t - \kappa_1$ or $\tau < t - \kappa_2$, and $\varsigma(\tau) = \varsigma_2 = \varsigma_1 + \Delta_\varsigma$ otherwise with $\varsigma_1 = 0.3$, $k = 1,000$, $k_1 = (k - k_2) \in \{100, 333, 450\}$ or $\Delta_k \in \{800, 333, 100\}$

Looking now at the results, from the top panel of Table 2 it appears that the forecasting accuracy deteriorates when the forecasting horizon under consideration increases, as all three performance criteria considered are considerably larger for the 10-steps ahead period. However, comparing the top and bottom part of Table 2 it is clear that the location of the breaks does affect the forecasting performance of the model. Similarly, comparing the benchmark case of $\Delta_\varsigma = 0.00$ in the top panel of Table 2 with $\Delta_\varsigma = 0.50$ it appears that, when the forecasting horizon increases, the greater the magnitude of the break the worse the forecasting accuracy.

Having investigated the size and power properties of unit root tests in the presence of GARCH-M effects and unknown structural breaks in the in-mean parameter, next we will derive an explicit formula for the general solution of the DAB-BVARMA$(1, 1; 2)$ representation.

4 VAR General Solution

In this section we provide the generating solution of the DAB-BVARMA$(1, 1; 2)$ representation, which generates explicit formulas for the optimal predictors and the bidimensional time varying covariance matrix of $\{\mathbf{y}_\tau\}$, $\tau = t + r$, $r \in \mathbb{Z}_{\geq 0}$ (the set of nonnegative integers).

First, let $\lambda_{\max}(\mathbf{X})$ denote the modulus of the largest eigenvalue of $\mathbf{X}$. The following theorem holds (the proof is presented in the Appendix).

Theorem 1 *The general solution of the bivariate system in eq. (6), subject to the initial condition* $\mathbf{y}_{\tau-k}$*, for* $k \geq k_2 + r$*, is given by*

$$\mathbf{y}_{\tau,k} = \mathbf{E}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k}) + \mathbf{FE}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k}), \tag{8}$$

where

$$\begin{aligned}
\mathbf{E}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k}) &= \varphi_k(\tau) + \mathbf{\Phi}_1^{k_1+r}\mathbf{\Phi}_2^{k_2-k_1}\mathbf{\Phi}^{k-k_2-1}(\mathbf{\Phi}\mathbf{y}_{\tau-k} + \mathbf{Z}\varepsilon_{\tau-k}),\\
\mathbf{FE}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k}) &= \mathbf{J}\varepsilon_\tau + \sum_{\ell=1}^{k_1+r}\mathbf{\Phi}_1^{\ell-1}(\mathbf{\Phi}_1\mathbf{J} + \mathbf{Z}_1)\varepsilon_{\tau-\ell} + \mathbf{\Phi}_1^{k_1+r}\{\sum_{\ell=1}^{k_2-k_1}\mathbf{\Phi}_2^{\ell-1}(\mathbf{\Phi}_2\mathbf{J} + \mathbf{Z}_2)\varepsilon_{t-k_1-\ell}\\
&\quad + \mathbf{\Phi}_2^{k_2-k_1}[\sum_{\ell=1}^{k-k_2-1}\mathbf{\Phi}^{\ell-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})\varepsilon_{t-k_2-\ell}]\},
\end{aligned}$$

and if $\lambda_{\max}(\mathbf{\Phi}_n) \neq 1$*,* $n = 1, 2, 3$*, then*

$$\varphi_k(\tau) = (\mathbf{I} - \mathbf{\Phi}_1^{k_1+r})(\mathbf{I} - \mathbf{\Phi}_1)^{-1}\varphi_1 + \mathbf{\Phi}_1^{k_1+r}[(\mathbf{I} - \mathbf{\Phi}_2^{k_2-k_1})(\mathbf{I} - \mathbf{\Phi}_2)^{-1}\varphi_2 + \mathbf{\Phi}_2^{k_2-k_1}(\mathbf{I} - \mathbf{\Phi}^{(k-k_2)})(\mathbf{I} - \mathbf{\Phi})^{-1}\varphi].$$

In the above expression if $\lambda_{\max}(\mathbf{\Phi}_n) = 1$, then $(\mathbf{I} - \mathbf{\Phi}_n^{k_n-k_{n-1}})(\mathbf{I} - \mathbf{\Phi}_n)^{-1}$, with $k_0 = -r$ and $k_3 = k$, should be replaced by $\sum_{\ell=0}^{k_n-k_{n-1}-1}\mathbf{\Phi}_n^\ell$ (a similar argument holds for any of the analogous cases that follow).

The above theorem expresses the general solution, $\mathbf{y}_{\tau,k}$, in terms of the $(k + r)$-step ahead optimal in (L_2 sense) linear predictor, $\mathbf{E}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k})$, and the associated forecast error, $\mathbf{FE}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k})$. Clearly, if $k_2 = k_1$ eq. (8) gives the solution in the case of one break, whereas if $k_2 = k_1 = k$, it gives the general solution when there is no time variation. For example, for the time invariant case, since $\mathbf{\Phi}_1 = \mathbf{\Phi}_2 = \mathbf{\Phi}$ and $\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}$, the forecast error in eq. (8) reduces to

$$\mathbf{FE}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k}) = \mathbf{J}\boldsymbol{\varepsilon}_\tau + \sum_{\ell=1}^{k+r} \mathbf{\Phi}^{\ell-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})\boldsymbol{\varepsilon}_{\tau-\ell}. \tag{9}$$

The general solutions when $k \leq k_1 + r$ and $k_1 + r < k < k_2 + r$ can be obtained along the lines of Theorem 1 and are equivalent to the time invariant case and the case when there is one break, respectively. In this section, in the context of the DAB-AR-M model, we show the importance of taking into account abrupt breaks for the in-sample forecasting.

Having found an explicit formula for the general solution of the DAB-BVARMA$(1,1;2)$ representation, in the next section we will derive an explicit formula for the bidimensional time varying covariance matrix of $\{\mathbf{y}_t\}$, which, as noted above, is needed in order to obtain a time varying measure of second-order persistence

4.1 Second Moment Structure

In this section we will examine the second moment structure of the DAB-BVARMA representation in eq. (6). First we will introduce some further notation.

Let $\mathbf{X}^{\otimes 2} = \mathbf{X} \otimes \mathbf{X}$ where $\otimes$ is the Kronecker product. In addition, let $vec(\mathbf{X})$ be a vector in which the columns of matrix $\mathbf{X}$ are stacked one underneath the other, and $\mathbf{s} = vec(\mathbf{\Sigma})$. Finally, let $\mathbf{\Gamma}_\tau$ denote the zero order bidimensional time varying covariance matrix of $\{\mathbf{y}_\tau\}$ and $\boldsymbol{\gamma}_\tau = vec(\mathbf{\Gamma}_\tau)$, that is $\boldsymbol{\gamma}_\tau = (\mathbb{V}ar(y_\tau)$, $\mathbb{C}ov(y_\tau, \sigma^\delta_\tau)$, $\mathbb{C}ov(y_\tau, \sigma^\delta_\tau)$, $\mathbb{V}ar(\sigma^\delta_\tau))'$.

Assumption 1 (Second-Order): We assume that $\lambda_{\max}(\mathbf{\Phi}_n)^{\otimes 2} < 1$, $n = 1, 3$.

Assumption 1 implies that the DAB-BVARMA$(1,1;2)$ representation is second-order. The equivalent Assumption for this representation to be first-order is: $\lambda_{\max}(\mathbf{\Phi}_n) < 1$, $n = 1, 3$. Clearly, this condition is sufficient for the condition in Assumption 1 to hold. Due to space considerations the first moment structure of the above process and its Wold-Crámer decomposition are not reported but are available upon request.

The following theorem states expressions for the $\mathbf{\Gamma}_\tau$ (the proof is presented in the Appendix; in the interest of brevity the results for higher order time varying covariances are not reported but are available upon request).

Theorem 2 *Consider the general model in eq. (6). Then under Assumption* 1 $\boldsymbol{\gamma}_\tau$ *is given by*

$$\boldsymbol{\gamma}_\tau = \mathbf{G}(\tau)\mathbf{s}, \tag{10}$$

where

$$\mathbf{G}(\tau)=[g_{ij}(\tau)] = \mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_1)^{\otimes 2}]^{-1}(\mathbf{\Phi}_1\mathbf{J} + \mathbf{Z}_1)^{\otimes 2}$$
$$+ (\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}\{[\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2^{k_2-k_1})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2)^{\otimes 2}]^{-1}(\mathbf{\Phi}_2\mathbf{J} + \mathbf{Z}_2)^{\otimes 2}$$
$$+ (\mathbf{\Phi}_2^{k_2-k_1})^{\otimes 2}(\mathbf{I}^{\otimes 2} - \mathbf{\Phi}^{\otimes 2})^{-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})^{\otimes 2}\},$$

and thus $\mathbf{G}_1 = [g_{ij,1}] = \lim_{r\to\infty} \mathbf{G}(\tau)$ *is given by*

$$\mathbf{G}_1 = \mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_1)^{\otimes 2}]^{-1}(\mathbf{\Phi}_1\mathbf{J} + \mathbf{Z}_1)^{\otimes 2}. \tag{11}$$

Clearly, if we set $k_1 = k_2$, and therefore $\mathbf{\Phi}_1 = \mathbf{\Phi}_2$ and $\mathbf{Z}_1 = \mathbf{Z}_2$ (the case of one break), then we obtain the results for the simpler case, where we have only one abrupt break, at time $t - k_2$. In this case the form for $\mathbf{G}(\tau)$ simplifies to

$$\mathbf{G}(\tau) = \mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2^{k_2+r})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2)^{\otimes 2}]^{-1}(\mathbf{\Phi}_2\mathbf{J} + \mathbf{Z}_2)^{\otimes 2} + (\mathbf{\Phi}_2^{k_2+r})^{\otimes 2}(\mathbf{I}^{\otimes 2} - \mathbf{\Phi}^{\otimes 2})^{-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})^{\otimes 2},$$

and thus $\mathbf{G}_2 = [g_{ij,2}] = \lim_{r\to\infty} \mathbf{G}(\tau)$ is given by

$$\mathbf{G}_2 = \mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2)^{\otimes 2}]^{-1}(\mathbf{\Phi}_2\mathbf{J} + \mathbf{Z}_2)^{\otimes 2}. \tag{12}$$

Further, if in addition $k_2 = k$, then $\mathbf{G}(\tau)$, since $\mathbf{\Phi}_2 = \mathbf{\Phi}$ and $\mathbf{Z}_2 = \mathbf{Z}$, reduces to the well known formula for the time invariant model

$$\mathbf{G} = \mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - \mathbf{\Phi}^{\otimes 2}]^{-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})^{\otimes 2}, \tag{13}$$

which is the result obtained in Conrad and Karanasos (2015a), but it is expressed in a more compact way. It follows directly from the above theorem that the first element of the time varying covariance vector, $\boldsymbol{\gamma}_\tau$, which is the time dependent variance of y_τ, is given by

$$\mathbb{V}ar(y_\tau) = \sigma_\varepsilon g_{11}(\tau) + \sigma_{\varepsilon v} 2 g_{12}(\tau) + \sigma_v g_{14}(\tau). \tag{14}$$

Notice that the three time invariant variances for each of the three periods, denoted by $\mathbb{V}ar_n(y_\tau)$, $n = 1, 2, 3$, are obtained from the above expression by replacing $g_{1j}(\tau)$ with $g_{1j,n}$ (see eqs. (11)-(13)). In addition, since the matrices $\mathbf{\Phi}_n$ are upper triangular, the $\mathbf{G}(\tau)$ matrix is also upper triangular and its $(4, 4)$ time invariant element is $g_{44} = \frac{\alpha^2}{1-c^2}$. Thus, the fourth element of $\boldsymbol{\gamma}_\tau$, which is the time invariant unconditional variance of σ_τ^δ, is given by

$$\mathbb{V}ar(\sigma_\tau^\delta) = \mu_2 - \mu_1^2 = \frac{\alpha^2}{1 - c^2}\sigma_v.$$

Since $\sigma_v = \mu_2 \kappa$ (see eq. 4) and using $\mu_1 = \frac{\omega}{1-c}$ we obtain (if and only if $c^2 - \alpha^2\kappa < 1$) by straightforward manipulation:

$$\mu_2 = \frac{(1 + c)\omega^2}{(1 - c)(1 - c^2 - \alpha^2\kappa)}, \tag{15}$$

which is a standard result (see, e.g., Karanasos, 1999, He and Teräsvirta, 1999, and Karanasos and Kim, 2006).

In the next section we will show how the above results can be used to derive a time varying second-order measure of persistence.

4.2 Time Varying Persistence

The most often applied time invariant measures of first-order (or mean) persistence are the LAR, and the SAR. As pointed out by Pivetta and Reis (2007) in relation to the issue of recidivism by monetary policy its occurrence depends very much on the model used to test the natural rate hypothesis, i.e., the hypothesis that the SAR or the LAR for inflation data is equal to one. Obviously, both measures would ignore the presence of breaks and in-mean effects and, hence, potentially under or over estimate the persistence in the levels, which is partly induced by the persistence in the conditional variance.

The LAR has been used to measure persistence in the context of testing for the presence of unit roots (see, for details, Pivetta and Reis, 2007). The authors find no evidence pointing to a rejection of a unit root in inflation. However, as we show in Section 3 if the in-mean mechanism together with the possible presence of breaks in the in-mean parameter are ignored, then conventional procedures (such as unit root tests) for estimating the persistence in the mean may lead to biased estimates. In particular, they might falsely indicate a unit root, and, hence, suggest the modeling of the differenced series rather than their levels.

In the following, we suggest a time varying second-order (or variance) persistence measure that is able to take into account the presence of breaks and to distinguish between the effects of a *mean shock* and a *volatility shock* on the level and conditional variance respectively. Fiorentini and Sentana (1998) argue that any reasonable measure of shock persistence should be based on the IRFs. For a univariate process x_t with *i.i.d* errors, e_t, they define the persistence of a shock e_t on x_t as $P(x_t\,|e_t) = \mathbb{V}ar(x_t)/\mathbb{V}ar(e_t)$. Clearly $P(x_t\,|e_t)$ will take its minimum value of one if x_t is white noise and it will not exist (will be infinite) for an $I(1)$, process.

Regarding the DAB-AR-(APGARCH) M model, if $\sigma_{\epsilon v} = 0$, that is there are no asymmetries (see eq. (4)), then ε_t and v_t can be viewed as 'structural' shocks. Thus it follows directly from eq. (14) that:

$$\mathbb{V}ar(y_\tau) = \sigma_\varepsilon g_{11}(\tau) + \sigma_v g_{14}(\tau). \tag{16}$$

Correlated Shocks

In general, the two shocks will be correlated with covariance matrix $\mathbf{\Sigma}$. In this case we will define two uncorrelated shocks with variances equal to one.

The new orthogonal shocks, $\widetilde{\varepsilon}_t$ and $\widetilde{v}_t$, can be obtained from the original shocks via the transformation:

$$\widetilde{\varepsilon}_\tau = \frac{\varepsilon_\tau}{\sqrt{\sigma_\epsilon}},\ \ \widetilde{v}_\tau = \frac{1}{\sqrt{1-\rho^2_{\varepsilon v}}}(-\rho_{\varepsilon v}\widetilde{\varepsilon}_\tau + \frac{v_\tau}{\sqrt{\sigma_v}}),$$

where $\rho_{\varepsilon v}$ is the correlation between ε_t and v_t.

Now, the persistence of the two shocks, $\widetilde{\varepsilon}_t$ and $\widetilde{v}_t$, for the variance of y_t, can be decomposed as follows:

$$\mathbb{V}ar(y_\tau) = P(y_\tau\,|\widetilde{\varepsilon}) + \; P(y_\tau\,|\widetilde{v}), \tag{17}$$

where

$$P(y_\tau\,|\widetilde{\varepsilon}) = \sigma_\epsilon g_{11}(\tau) + \rho^2_{\varepsilon v}\sigma_v g_{14}(\tau) + 2\sigma_{\epsilon v} g_{12}(\tau), \;\; P(y_\tau\,|\widetilde{v}) = \sigma_v(1-\rho^2_{\varepsilon v})g_{14}(\tau).$$

Clearly, if we have the symmetric case, that is $\gamma = 0$ and, therefore, $\sigma_{\epsilon v} = 0$ the above expression reduces to the one in eq. (16). To save space the equivalent persistence measures for the power transformed conditional variance and also for the product $y_t\sigma_t^\delta$ are not reported but are available upon request. Notice that when $\delta = 1$ the above expressions (since $\sigma_\epsilon = \mu_2$, $\sigma_v = \mu_2(1+\gamma^2 - \frac{2}{\pi})$ and $\sigma_{\epsilon v} = -\mu_2\gamma$ [see Corollary 1]) reduce to

$$P(y_\tau\,|\widetilde{\varepsilon}) = \mu_2[g_{11}(\tau) + \gamma^2 g_{14}(\tau) - 2\gamma g_{12}(\tau)], \;\; P(y_\tau\,|\widetilde{v}) = \mu_2(1-\frac{2}{\pi})g_{14}(\tau), \tag{18}$$

where μ_2 is given in eq. (15) with $c = \alpha\sqrt{\frac{2}{\pi}} + \beta$ (see eq. (3) and Corollary 1).

If Assumption 1 is violated then conditional measures of second-order persistence can be constructed using the variance of the forecast error (see eq. A.3 in the Appendix) instead of the unconditional variance (results not reported but are available upon request).

We have derived explicit formulas for time varying second-order (or variance) persistence measures.[6] In the next section we show the empirical relevance of these results using U.S. inflation data.

5 Inflation Data

In our empirical application we consider log-differences of quarterly data of Personal Consumption Expenditure (CPE) in the United States from 1947Q1 to 2016Q3. The CPE index is used by the Federal Reserve as an inflation proxy when reviewing economic conditions and charting a course of action designed to impact on the real economy. It is therefore crucially important to be able to make appropriate modeling, inference and forecasting in order to avoid unwanted effects of monetary policy actions.

The first step in the estimation procedure is to identify possible points of parameter changes. In order to do so the Bai and Perron (2003) breakpoint estimation technique on inflation rates is used to identify possible breaks during the sample period.[7] Using the Bai and Perron procedure two significant breaks were identified. The first break took place in the 1960's, when an expansion of social programs

[6] Cogley and Sargent (2001) measured persistence by the spectrum at frequency zero, S_0. As an example, for the time invariant AR(2) model this will be given by: $S_0 = \frac{\sigma^2_\varepsilon}{2\pi(1-\phi_1-\phi_2)^2}$.

[7] Since the seminal paper by Perron (1989) a great deal of research has been directed to the detection and estimation of breaks, and forecasting in the presence of breaks (see, e.g., Andrews, 1993; Andrews and Ploberger, 1994; Bai and Perron, 1998).

was undertaken by the U.S. administration in the aftermath of a contraction period when unemployment and inflation reached high levels. The second break occurred in 2008 at the height of the financial crisis and during a spike of the oil price. In particular, the breaks were identified in 1966Q4 and 2008Q3.[8]

Accordingly, below we estimate the DAB-AR$(1;2)$-M model in eqs. (1)-(2), (with $\delta = 1$),[9] allowing for both the *intrinsic* persistence of inflation (as captured by the autoregressive coefficient, $\phi(t)$) and the in-mean coefficient, $\varsigma(t)$, to switch across breakpoints.[10] This should allow us to determine whether changes in the structure of the conditional mean of inflation recently observed in the U.S. derive from changes in either $\phi(t)$ or $\varsigma(t)$, or possibly both. To capture these changes we use dummy variables that take the value zero in the period before each break and the value one after the break.[11]

5.1 Estimation Results

As far as the QML estimation results are concerned Table 3 reports the estimated parameters for each of the three models and the relative misspecification tests. In particular, the top part of Table 3 reports the estimated parameters for the conditional mean, whereas the ones for the conditional variance are given in Panel B.

[8] Kim et al. (2004) found evidence for a structural break in inflation in late 1979, resulting in lower persistence. The Bai and Perron methodology also identified a third break in 1977Q4. However, in the estimation of the DAB-AR-M model the corresponding dummy variable was insignificant. Pivetta and Reis (2007) point out that extra data points in their sample (1965-2001) might show a break in 1991. The Bai and Perron methodology also identified a fourth break in 1991Q3. However, in the estimation of our model the corresponding dummy variable was insignificant.

Another line of work has identified changes in the way monetary policy is conducted in the United States. Therefore we also add a dummy variable for the period 1981–1983, which was an anomalous period in the data for inflation, commonly referred to as the Volcker disinflation. The dummy was insignificant.

[9] Karanasos and Schurer (2008) show that it is optimal to model the conditional standard deviation of inflation instead of the conditional variance. So far the relevant empirical literature has ignored this important characteristic of the inflation data.

[10] The asymptotic theory for the QML estimator of the parametric GARCH-M model has recently been developed by Conrad and Mammen (2016). However, this theory does not yet treat all standard specifications.

[11] An alternative approach to account for structural breaks would be to estimate similar models for subperiods. Due to the limited number of quarterly observations we do not consider this option.

Table 3. Estimated DAB-AR$(1;2)$-M model using U.S. inflation data

	Model 1	Model 2	Model 3
	Panel A: Conditional Mean		
φ	0.0009	0.0004	0.0004
	(0.0006)	(0.0005)	(0.0005)
ϕ_3	0.405^{*}	0.545^{*}	0.377^{*}
	(0.093)	(0.059)	(0.088)
$\Delta\phi_3$	-0.318^{*}	–	-0.28^{*}
$(\phi_2 = \phi_3 - \Delta\phi_3)$	(0.085)		(0.153)
$\Delta\phi_2$	0.263^{**}	–	–
$(\phi_1 = \phi_3 - \Delta\phi_3 - \Delta\phi_2)$	(0.142)		
ς_3	0.481^{**}	0.432^{**}	0.643^{**}
	(0.217)	(0.188)	(0.191)
$\Delta\varsigma_3$	–	-0.613^{**}	–
$(\varsigma_2 = \varsigma_3 - \Delta\varsigma_3)$		(0.153)	
$\Delta\varsigma_2$	–	0.376^{***}	0.181
$(\varsigma_1 = \varsigma_3 - \Delta\varsigma_3 - \Delta\varsigma_2)$		(0.227)	(0.230)
	Panel B: Conditional Variance		
ω	0.0001^{**}	0.0002^{**}	0.0001
	(0.000)	(0.000)	(0.000)
α	0.092^{**}	0.103^{**}	0.103^{**}
	(0.044)	(0.052)	(0.048)
γ	-0.803^{**}	-0.791^{***}	-0.671^{**}
	(0.370)	(0.358)	(0.305)
β	0.871^{*}	0.857^{*}	0.868^{*}
	(0.044)	(0.055)	(0.048)
$c = \alpha\kappa_1 - \beta$	0.944	0.939	0.950
R^2	0.625	0.626	0.616
	Panel C: Q-Statistics and Information Criteria		
Q-Statistics (4)	1.702	0.852	0.335
	[0.199]	[0.310]	[0.562]
Akaike	−8.403	−8.004	−8.432
Schwarz	−8.269	−7.789	−8.376
	Panel D: Forecasting		
	Conditional Mean		
MSE	0.000	0.000	0.000
MAE	0.005	0.005	0.004
RMSE	0.004	0.006	0.006
	Conditional Variance		
MSE	0.000	0.000	0.000
MAE	0.006	0.005	0.008
RMSE	0.007	0.002	0.005

Note: *, **, *** indicate statistical significance at 1%, 5% and 10%, respectively. The numbers in parentheses are standard errors. The numbers in brackets are p-values. $-\Delta\phi_3$, $-\Delta\varsigma_3$ and $-\Delta\phi_2$, $-\Delta\varsigma_2$ are the estimated parameters for the first and second dummy, respectively. Accordingly, in Model 1: $\phi_2 = 0.723$ and $\phi_1 = 0.460$; In Model 2, $\varsigma_2 = 1.045$ and $\varsigma_1 = 0.669$; In Model 3: $\phi_2 = 0.377$ and $\varsigma_1 = 0.462$.

In order to investigate whether changes in inflation were due to breaks in the *intrinsic* persistence coefficient or the in-mean coefficient the model in eqs. (1)-(2) was estimated with no dummy variables for the latter (i.e., with $\varsigma_2 = \varsigma_1 = 0$ in Table 3) and the resulting specification is labeled as Model 1. Similarly, to investigate if breaks in the in-mean coefficient did affect inflation, the model was estimated with no dummies capturing breaks in the *intrinsic* persistence (i.e., with the parameters $\phi_2 = \phi_1 = 0$) and the resulting specification is referred to as Model 2. Finally, to investigate the joint effects of the two types of breaks various attempts were made to estimate a model with breaks in both coefficients. The best (based on information criteria and likelihood ratio tests) resulting specification is labeled as Model 3.[12]

Looking now at the estimated parameters, according to the estimates in Model 1 until 1966 the inflation process was well approximated by a first-order autoregression with low *intrinsic* persistence ($\phi_3 = 0.40$), but from 1967 onwards the autoregression coefficient increased considerably ($\phi_2 = \phi_3 - \Delta\phi_3 = 0.72$). It was only after 2008 that the *intrinsic* persistence level went back to roughly its previous regime ($\phi_1 = \phi_2 - \Delta\phi_2 = 0.46$).

According to Model 2 it appears that inflation variability imposed upward pressure on inflation ($\varsigma_3 = 0.43$). In general, higher inflation uncertainty increases long term risk premia, inducing extra hedging costs due to inflation risks therefore shifting upward inflation levels, as predicted by Cukierman and Meltzer. Interestingly, the in-mean effect becomes stronger (it more than doubles in size) after 1966 ($\varsigma_2 = \varsigma_3 - \Delta\varsigma_3 = 1.04$). However, in the post-(global financial) crisis it decreases ($\varsigma_1 = \varsigma_2 - \Delta\varsigma_2 = 0.67$) although it does not return to the pre-1967 level. Finally, Model 3 confirms that changes in inflation dynamics can be explained by both changes in the *intrinsic* persistence and the in-mean coefficient. In all models the estimated intrinsic variance persistence is high (c is either 0.94 or 0.95) and the asymmetry coefficient, γ, is negative, indicating that negative shocks have a higher impact on the volatility than positive shocks.

Looking now at the specification tests in Panel C the Q-Statistics do not reject the null hypothesis of no serial correlation, therefore indicating that the models do not suffer from misspecification. Also, from the information criteria it appears that Model 2 offers the best specification for the inflation process. Finally, the bottom part reports the 5-step ahead forecasting criteria for the models under consideration. It appears that all three models have relatively good forecasting properties.

To summarize our results, we find evidence that the parameters in the models capturing *intrinsic* persistence and in-mean effects change over time. Therefore, not allowing for time varying coefficients in the estimation procedure would result in a less accurate modeling of the inflation process. This, in the light of the simulation results in Section 3.3, may lead to poor forecasting.

Next, we will investigate whether inflation and its variability are highly persistent.

5.2 Inflation Persistence

Pivetta and Reis (2007) employ different estimation methods and measures of persistence. Estimating the persistence of inflation over time using different measures and procedures is beyond the scope of

[12] There is also the possibility of a change in the drift of inflation. Such a shift can be interpreted as a change in the long-run inflation target of the Federal Reserve (see, for details, Pivetta and Reis, 2007 and the references therein). Our DAB-AR-M model also allows for changes in the intercept. However, and in spite of allowing for a time varying drift, we find that the second-order persistence is unchanged (see the analysis below), a result which is in agreement with the one in Pivetta and Reis.

this paper.[13] In this section we depart from their study in an important way, that is we contribute to the measurement over time of inflation persistence by taking a different approach to the problem and estimating a model of inflation dynamics grounded in economic (rather than statistical) theory. In particular, we employ the DAB AR-(APGARCH) M model and we compute an alternative measure of persistence, that is, the second-order persistence (using the methodology in Sections 4.1 and 4.2), which not only distinguishes between changes in the dynamics of inflation and its volatility (and their persistence) but also allows for feedback from volatility (inflation uncertainty) to the level of the process (inflation).

As pointed out by Pivetta and Reis (2007) estimates of the persistence of inflation affect the tests of the natural hypothesis neutrality. Therefore detecting whether persistence has recently fallen is key in assessing the likelihood of recidivism by the central bank. In addition, if the central bank feels encouraged to exploit an illusory inflation-output trade off, the result could be high inflation without any accompanying output gains. Furthermore, research on dynamic price adjustment has emphasized the need for theories that generate inflation persistence.

Table 4 presents the time invariant (within each period) second-order measures of persistence for the three models under consideration. The first three columns report the persistence for the mean shock ($\widetilde{\varepsilon}$), the next three columns the persistence for the volatility shock ($\widetilde{v}$), and the last three columns the sum of the two shocks (see eq. (18)). Model 2, which is the preferred one, generates the highest persistence. Interestingly, for this model the persistence is very similar in all three periods (approximately 11). Model 1 is the one with the lowest persistence. For this model the persistence doubles in the second period, but after the global financial crisis it almost returns to the pre-1967 levels. In model 3 the persistence increases by 75% in the period 1967-2008 and decreases by 28% in the post-crisis period.

Table 4. 'Second-order persistence' for each of the three periods and models.

	$P(y_t \mid \widetilde{\varepsilon}) \times 10^6$			$P(y_t \mid \widetilde{v}) \times 10^6$			$[P(y_t \mid \widetilde{\varepsilon}) + P(y_t \mid \widetilde{v})] \times 10^6$		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
1947Q1-1966Q4	4.51	10.88	5.64	0.062	0.32	0.18	4.57	11.20	5.82
1967Q1-2008Q2	9.01	10.98	9.67	0.26	0.41	0.55	9.26	11.40	10.22
2008Q3-2016Q3	4.84	10.92	6.81	0.07	0.37	0.53	4.92	11.29	7.34

Note: We use eq. (18) to calculate the (within each period time invariant) second-order persistence for the three models. For each period, $n = 1, 2, 3$, we obtained the $g_{1j,n}$, using eqs. (11)-(13).

For comparison Table A.1 in the Appendix presents the (time invariant within each period) 'first-order persistence' for all three models. The pre-1967 period exhibits the lowest persistence whereas in

[13] Pivetta and Reis (2007) applied a Bayesian approach, which explicitly treats the autoregressive parameters as being stochastically varying and it provides their posterior densities at all points in time. From these, they obtained posterior densities for the measures of inflation persistence. Such estimates of persistence are forward-looking, since they are meant to capture the perspective of a policy maker who at a point in time is trying to forsee what the persistence of inflation will be. They also estimated backward-looking measures of persistence that the applied economist forms at a point in time, given all the sample until then.

Pivetta and Reis (2007) also used an alternative set of estimation techniques for persistence. They assumed time invariant autoregressive parameters and re-estimated their AR model on different sub-samples of the data, obtaining median unbiased estimates of persistence for each regression. Finally, Pivetta and Reis also employed rolling and recursive unit root tests.

the second period the persistence is the highest. In particular, for the second model from 1967 onwards it increases by 60%. For all models in the post global financial crisis period the persistence falls, but it remains higher than the pre-1967 levels except perhaps for model 1, where it more than doubles in the second period and then it almost halves in the post-crisis period.

Therefore our findings regarding the first-order persistence are in line with the findings of i) Barsky (1987), who found very low inflation persistence for the pre-war period but that inflation is very persistent since the 1960s, and ii) Pivetta and Reis (2007), who by computing alternative statistical measures of persistence came to the conclusion that inflation persistence in the United States is best described as unchanged over their sample, which was 1965-2001.

In sum our main conclusion is that for our chosen specification (model 2) the preferred measure of persistence, that is the second-order persistence, remained relatively unchanged throughout the whole period 1947-2016. These results are in line with Pivetta and Reis (2007) and Stock (2001), who also found no evidence of a change in inflation persistence in the United States.[14]

6 Conclusions

In recent years economists have placed significant increasing emphasis on investigating structural shifts in the dynamics of the inflation process in the United States. A number of detailed and rigorous empirical studies regarding changes in inflation persistence have, however, reached diverging conclusions. Several studies find evidence of little or no change of inflation persistence over the past four decades, whereas others conclude that there has been a pronounced decline over the same period.

In this paper we have attempted to reconcile different strands of the literature by showing that seemingly conflicting results regarding changes in inflation persistence actually constitute two sides of the same problem. We started our analysis by showing that if the estimated model is misspecified with respect to the data generating process, commonly used test statistics to detect persistence would deliver spurious results. In particular, using Monte Carlo simulation experiments we have shown that if the in-mean mechanism (together with the possible presence of breaks in the in-mean parameter) is ignored, conventional unit root tests might falsely indicate inflation as being a nonstationary rather then a stationary process. The obvious consequence is that commonly used inference procedures would suggest the modeling of inflation processes in their first differences rather then in their levels. We then proceeded by using the general solution of a DAB AR-(APGARCH) M model to compute a time varying second-order persistence measure that is able to take into account the presence of breaks and to distinguish between the effects of a mean shock and a volatility shock on the level and conditional variance respectively.

[14] Stock and Watson (2002) also found no evidence of a change in persistence in U.S. inflation. Therefore their results are in agreement with ours. However, they found strong evidence of a fall in volatility. We also checked for possible changes in the unconditional volatility by adding dummy variables in the drift of the conditional variance. However, they are all insignificant.

Our paper adds to the literature that has challenged the empirical relevance of the Lucas critique on reduced-form models. In this respect, from the empirical point of view our measure of persistence constitutes an important breakthrough in the literature since we allow for feedback from volatility (inflation uncertainty) to the level of the process (inflation). Our estimation results lead to the conclusion that U.S. inflation persistence has been high and approximately constant over time, a finding which agrees with those of Stock and Watson (2002, 2001) and Pivetta and Reis (2007).

References

[1] ANDREWS, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821-856.

[2] ANDREWS, D. W. K. AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383-1414.

[3] BAI, J. AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47-78.

[4] BAI, J. AND P. PERRON (2003): "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics*, 18, 1-22.

[5] BAILLIE, R., C. CHUNG, AND M. TIESLAU (1996): "Analyzing Inflation by the Fractionally Integrated ARFIMA-GARCH Model," *Journal of Applied Econometrics*, 11, 23-40.

[6] BALL, L. AND S. G. CECCHETTI (1990): "Inflation and Uncertainty at Short and Long Horizons," *Brookings Papers on Economic Activity*, 1, 215-254.

[7] BARSKY, R. (1987): "The Fisher Hypothesis and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics*, 19, 3–24.

[8] BATINI, N. (2006): "Euro Area Inflation Persistence," *Empirical Economics,* 31, 977-1002.

[9] BEECHEY, M., AND P. OSTERHOLM (2009): "Time-Varying Inflation Persistence in the Euro Area", *Economic Modelling*, 26, 532-535.

[10] BERAN, J. (2009): "On Parameter Estimation for Locally Stationary Long-Memory Processes," *Journal of Statistical Planning and Inference*, 143, 201-220.

[11] BERUMENT H., Z. KILINC AND U. OZLALE (2005): "The Missing Link Between Inflation Uncertainty and Interest Rates," *Scottish Journal of Political Economy*, 52, 222-241.

[12] BLANCHARD, O. "*Macroeconomics*," 2nd edition, Prentice-Hall.

[13] BLANCHARD, O. AND J. GALI (2007): "Real Wage Rigidities and the New Keynesian Model," *Journal of Money, Credit, and Banking*, Supplement to 39, 35-66.

[14] BRUNNER, A. D. AND G. D. HESS (1993): "Are Higher Levels of Inflation Less Predictable? a State-Dependent Conditional Heteroscedasticity Approach," *Journal of Business & Economic Statistics*, 11, 187-197.

[15] CAPORALE, G. M. AND A. KONTONIKAS (2009): "The Euro and Inflation Uncertainty in the European Monetary Union," *Journal of International Money and Finance*, 28, 954-97.

[16] CAPORALE, G. M., L. ONORANTE, P. PAESANI (2010); "Inflation and Inflation Uncertainty in the Euro Area," *ECB Working Papers Series* No 1229.

[17] CHANDLER, G. AND W. POLONIK (2006): "Discrimination of Locally Stationary Time Series Based on the Excess Mass Functional," *Journal of the American Statistical Association,* 101, 240-253.

[18] CHRISTENSEN, B. J. AND M. Ø. NIELSEN (2007): "The Effect of Long Memory on Stock Market Fluctuations," *The Review of Economics and Statistics*, 89, 684-700.

[19] GLOSTEN, L. A., R. JAGANNATHAN, AND D. E. RUNKLE (1993): "On the Relation Between the Expected Value and the Volatility of Nominal Excess Return on Stocks," *Journal of Finance*, 46, 1779-1801.

[20] COGLEY, T. AND T. J. SARGENT (2001): " Evolving Post-World War II U.S. Inflation Dynamics," In Getler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual, pp. 331-373. MIT Press, Cambridge.

[21] COGLEY, T. AND T. J. SARGENT (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US," *Review of Economic Dynamics*, 8, 262-302.

[22] COGLEY, T. and M. SBORDONE (2008): "Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve", *American Economic Review*, 98, 2101-2126.

[23] CONRAD, C. AND M. KARANASOS (2015a): "On the Transmission of Memory in GARCH-in-mean Models," *Journal of Time Series Analysis*, 36, 706-720.

[24] CONRAD, C. AND M. KARANASOS (2015b): "Modeling the Link Between US Inflation and Output: the Importance of the Uncertainty Channel," *Scottish Journal of Political Economy*, 2015, 62, 431-453.

[25] CONRAD, C. AND E. MAMMEN (2016): "Asymptotics for Parametric GARCH-in-mean Models," *Journal of Econometrics*, 194, 319-329.

[26] CUKIERMAN, A. AND A. H. MELTZER (1986): "A Theory of Ambiguity, Credibility, and Inflation Under Discretion and Asymmetric Information," *Econometrica*, 54, 1099–1128.

[27] DICKEY, D. A. AND W. A. FULLER (1981): "Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root," *Econometrica*, 49, 1057-1072 .

[28] DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83-106.

[29] ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 75, 813-836.

[30] ELLIOTT, G. AND A. TIMMERMANN (2008): "Economic Forecasting," *Journal of Economic Literature*, 46, 3-56.

[31] ENGLE, R. F., D. M. LILIEN, AND R. P. ROBINS (1987): "Estimating Time Varying Risk Premia in the Term Structure: the ARCH-M Model," *Econometrica*, 55, 391-407.

[32] Evans, M. (1991): "Discovering the Link between Inflation Rates and Inflation Uncertainty." *Journal of Money, Credit, and Banking*, 23, 169-84.

[33] FIORENTINI, G. AND E. SENTANA (1998): "Conditional Means of Time Series Processes and Time Series Processes for Conditional Means," *International Economic Review*, 39, 1101-1018.

[34] FRIEDMAN, M. (1977): "Nobel Lecture: Inflation and Unemployment," *Journal of Political Economy*, 85, 451–472.

[35] FOUNTAS, S. M. KARANASOS, AND J. KIM (2006): "Inflation Uncertainty, Output Growth Uncertainty, and Macroeconomic Performance," *Oxford Bulletin of Economics and Statistics*, 68, 319-343.

[36] FOUNTAS, S. and M. KARANASOS, (2007): "Inflation, Output Growth, and Nominal and Real Uncertainty: Empirical Evidence for the G7," *Journal of International Money and Finance*, 26, p. 229-250.

[37] HE, C. AND T. TERÄSVIRTA (1999): "Fourth Moment Structure of the GARCH(p, q) Process," *Econometric Theory*, 15, 824-846.

[38] KARANASOS, M. (1999): "The Second Moment and the Autocovariance Function of the Squared Errors of the GARCH Model," *Journal of Econometrics,* 90, 63-76.

[39] KARANASOS, M. AND J. KIM (2006): "A Re-examination of the Asymmetric Power ARCH Model," *Journal of Empirical Finance*, 13, 113-128.

[40] KARANASOS, M. AND S. SCHURER (2008): "Is the Relationship Between Inflation and its Uncertainty Linear?," *German Economic Review*, 9, 265-286.

[41] KIM, C. J., C. R. NELSON, AND J. PIGER (2004): "The Less Volatile U.S. Economy: a Bayesian Investigation of Timing, Breadth, and Potential Explanations," *Journal of Business & Economics Statistics* 22, 80–93.

[42] KOOP, G. AND S. M. POTTER (2007): "Estimation and Forecasting in Models with Multiple Breaks," *The Review of Economic Studies*, 74, 763-789.

[43] LEVIN, A. T. AND J. PIGER (2003): "Is Inflation Persistence Intrinsic in Industrial Economies?" *Working Paper 2002-023B*, Federal Reserve Bank of St. Louis, February.

[44] PALMA, W. AND OLEA, R. (2010): "An Efficient Estimator for Locally Stationary Gaussian Long-Memory Process," *The Annals of Statistics*, 38, 2958-2997.

[45] PERRON, P. (1989): "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57, 1361-1401.

[46] PERRON, P. AND S. NG (1996): "Useful Modifications to some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties," *The Review of Economic Studies*, 63, 435–463.

[47] PESARAN, M. H., D. PETTENUZZO, AND A. TIMMERMANN (2006): "Forecasting Time Series Subject to Multiple Structural Breaks," *The Review of Economic Studies*, 73, 1057-1084.

[48] PESARAN, M. H. AND A. TIMMERMANN (2005): "Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks," *Journal of Econometrics*, 129, 183-217.

[49] PIVETTA, F. AND R. REIS (2007): "The Persistence of Inflation in the United States," *Journal of Economic Dynamics and Control*, 31, 1326-1358.

[50] ROSE, A. K. (1988): "Is the Real Interest Rate Stable?" *Journal of Finance,* 43, 1095-1112.

[51] SIMS, C. A., J. H. STOCK, AND M. W. WATSON (1990): "Inference in Linear Time Series Models with some Unit Roots," *Econometrica*, 58, 113-144.

[52] STOCK, J. H. (2001) "Comment on: Evolving Post-World War II U.S. Inflation Dynamics," In: Gertler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual, pp. 379-387. MIT Press, Cambridge.

[53] STOCK, J. H. AND M. W. WATSON (2002): "Has the Business Cycle Changed and Why?" In: Gertler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual, 159-218. MIT Press, Cambridge.

[54] STOCK, J. H. AND M. W. WATSON (2007): "Why has U.S. inflation Become Harder to Forecast?" *Journal of Money, Credit and Banking*, 39, 3–33.

[55] STOCK, J.H., and M.W. WATSON (2009): "Phillips Curve Inflation Forecasts," Ch. 3 in Fuhrer, J., Y., Kodrzycki, J., Little, Olivei, J. (Eds.), Understanding Inflation and the Implications for Monetary Policy, pp. 101-186, Cambridge: MIT Press (with M. W. Watson).

[56] TAYLOR, J. B. (2000): "Low Inflation Pass-Through, and Pricing Power of Firms," *European Economic Review,* 44, 1389-1408.

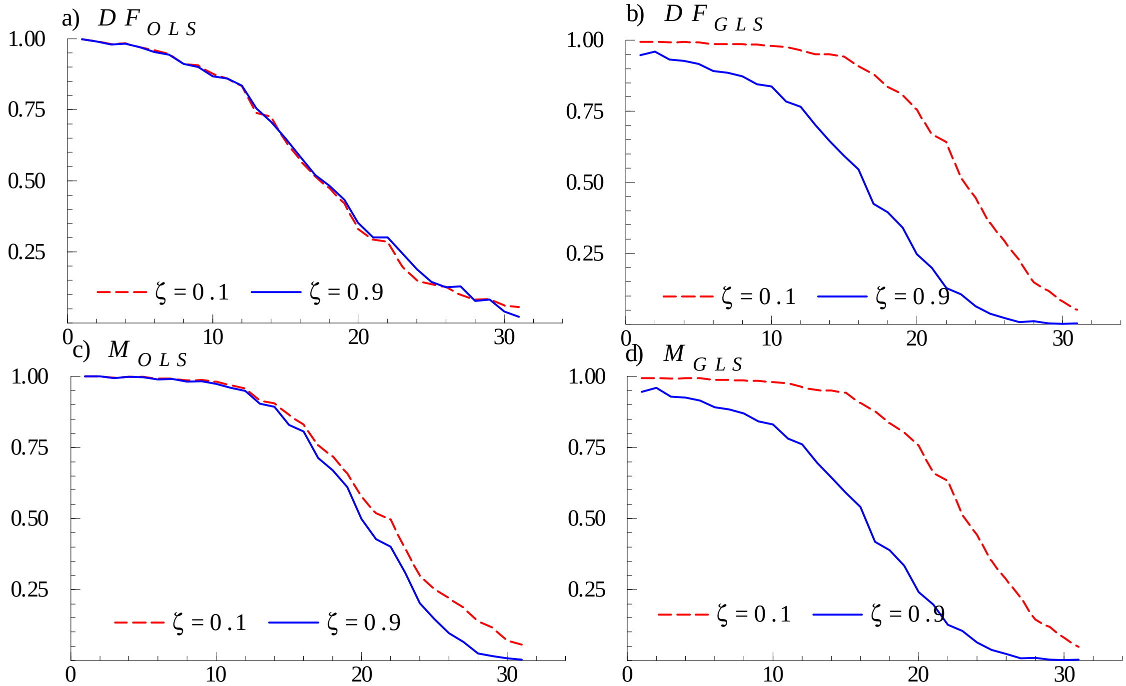
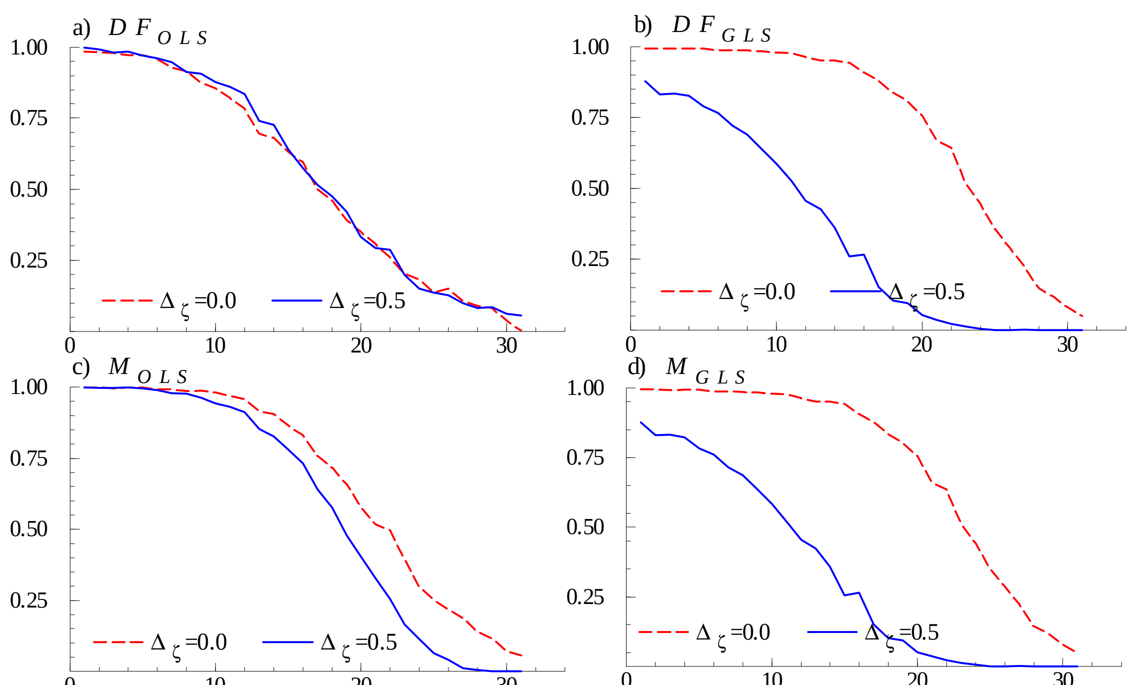


Figure 1. Power of DF and M tests. The DGP is $y_t = 1 + y_{t-1} + \varsigma\sigma_t + \varepsilon_t$ and

$$\sigma_t = 0.2 + 0.1\left|\varepsilon_{t-1}\sigma_{t-1}\right| + 0.7\sigma_{t-1}.$$

$_TAU01$



$_2.pdf$

Figure 2: Power of DF and M tests. The DGP is $y_t = 1 + y_{t-1} + \varsigma(t)\,\sigma_t + \varepsilon_t$ and

$\sigma_t = 0.2 + 0.1\left|\varepsilon_{t-1}\sigma_{t-1}\right| + 0.7\sigma_{t-1}$, where $\varsigma(\tau) = \varsigma_1$ if $\tau > t - k_1$ or $\tau < t - k_2$, and $\varsigma(\tau) = \varsigma_2 = \varsigma_1 + \Delta_\varsigma$

otherwise with $\varsigma_1 = 0.9$, $k = 1,000$, $k_1 = (k - k_2) = 100$ or $\Delta_k = 800$.

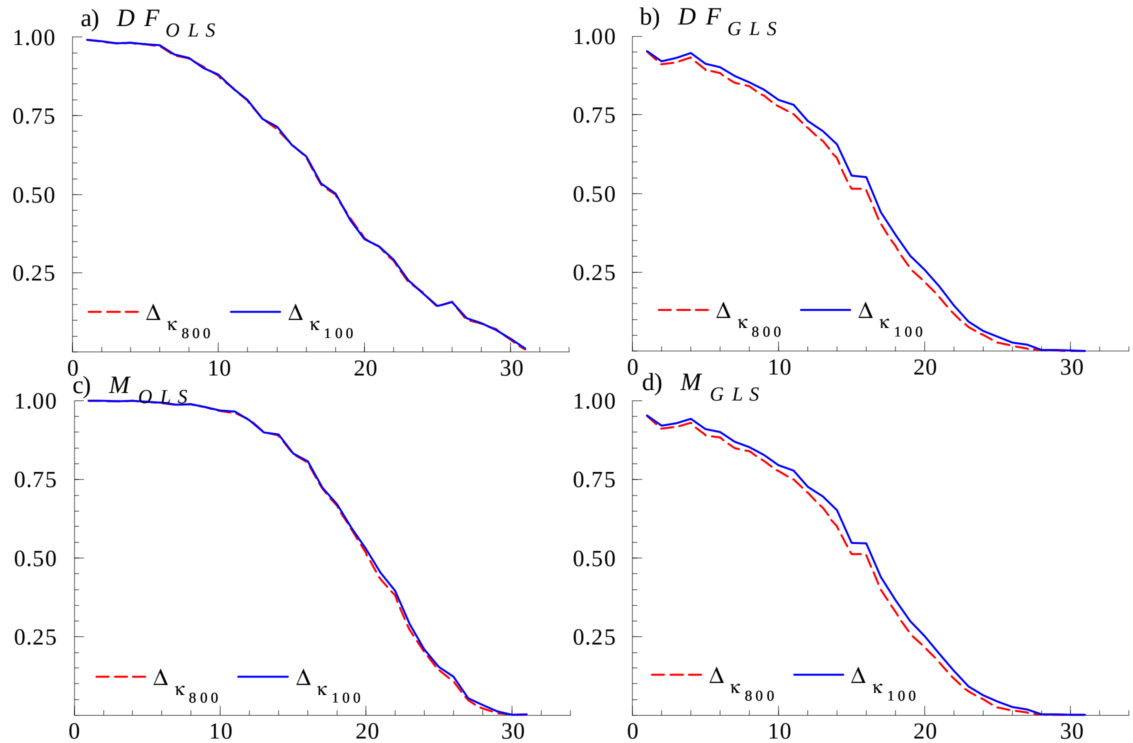


Figure 3. Power of DF and M tests. The DGP is $y_t = 1 + y_{t-1} + \varsigma(t)\,\sigma_t + \varepsilon_t$ and $\sigma_t = 0.2 + 0.1\,|\varepsilon_{t-1}\sigma_{t-1}| + 0.7\sigma_{t-1}$, where $\varsigma(\tau) = \varsigma_1$ if $\tau > t - k_1$ or $\tau < t - k_2$, and $\varsigma(\tau) = \varsigma_2 = \varsigma_1 + \Delta_\varsigma$ otherwise with $\varsigma_1 = 0.0$, $\Delta_\varsigma = 0.07$, $k = 1,000$, $k_1 = (k - k_2) \in \{100, 450\}$ or $\Delta_k \in \{800, 100\}$

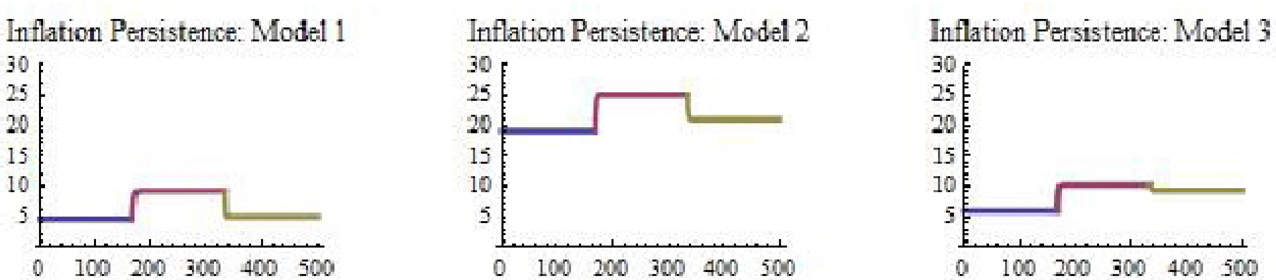


Figure 4. Time Varying Inflation Persistence

Appendix

In this appendix we present the proofs of Theorems 1 and 2. We will prove Theorem 1 by induction with respect to k.

Proof. (**of Theorem** 1). Clearly, it holds for $k+r=1$: In eq. (8) setting $k=1$, and thus $r=k_1=k_2=0$ we obtain eq. (6) since $\mathbf{E}(\mathbf{y}_\tau \left| \mathcal{F}_{\tau-1} \right.) = \boldsymbol{\varphi} + \boldsymbol{\Phi}\mathbf{y}_{\tau-1} + \mathbf{Z}\boldsymbol{\varepsilon}_{\tau-1}$ and $\mathbf{FE}(\mathbf{y}_\tau \left| \mathcal{F}_{\tau-k} \right.) = \mathbf{J}\boldsymbol{\varepsilon}_\tau$.
Next if we assume that it holds for k, then it will suffice to prove that it also holds for $k+1$. First, rewrite eq. (6) as of time $\boldsymbol{\tau}-k$:

$$\mathbf{y}_{\tau-k} = \boldsymbol{\varphi}(\tau-k) + \boldsymbol{\Phi}(\tau-k)\mathbf{y}_{\tau-(k+1)} + \mathbf{J}\boldsymbol{\varepsilon}_{\tau-k} + \mathbf{Z}(\tau-k)\boldsymbol{\varepsilon}_{\tau-(k+1)}.$$

Substituting the above equation into eq. (8) using straightforward algebra shows that

$$\mathbf{y}_{\tau,k+1} = \mathbf{E}(\mathbf{y}_\tau \left| \mathcal{F}_{\tau-(k+1)} \right.) + \mathbf{FE}(\mathbf{y}_\tau \left| \mathcal{F}_{\tau-(k+1)} \right.)$$

as claimed. ■
Proof. (**of Theorem** 2). Theorem 1 implies that

$$\boldsymbol{\Gamma}_\tau = \mathbf{V}ar[\mathbf{E}(\mathbf{y}_\tau \left| \mathcal{F}_{\tau-k} \right.)] + \mathbf{V}ar[\mathbf{FE}(\mathbf{y}_\tau \left| \mathcal{F}_{\tau-k} \right.)]. \tag{A.1}$$

The variance of the optimal predictor is given by

$$\mathbf{V}ar[\mathbf{E}(\mathbf{y}_\tau \left| \mathcal{F}_{\tau-k} \right.)] = \boldsymbol{\Phi}_1^{k_1+r}\boldsymbol{\Phi}_2^{k_2-k_1}\boldsymbol{\Phi}^{k-k_2-1}\mathbf{V}ar(\boldsymbol{\Phi}\mathbf{y}_{\tau-k} + \mathbf{Z}\boldsymbol{\varepsilon}_{\tau-k})(\boldsymbol{\Phi}_1^{k_1+r}\boldsymbol{\Phi}_2^{k_2-k_1}\boldsymbol{\Phi}^{k-k_2-1})', \tag{A.2}$$

where, under Assumption 1,

$$\mathbf{V}ar(\boldsymbol{\Phi}\mathbf{y}_{\boldsymbol{\tau}-k} + \mathbf{Z}\boldsymbol{\varepsilon}_{\boldsymbol{\tau}-k}) = \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}^\ell(\boldsymbol{\Phi}\mathbf{J} + \mathbf{Z})\boldsymbol{\Sigma}[\boldsymbol{\Phi}^\ell(\boldsymbol{\Phi}\mathbf{J} + \mathbf{Z})]'.$$

Similarly, the variance of the forecast error is given by

$$\begin{aligned} \mathbf{V}ar[\mathbf{FE}(\mathbf{y}_\tau \left| \mathcal{F}_{\tau-k} \right.)] \ &= \ \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' + \sum_{\ell=1}^{k_1+r} \boldsymbol{\Phi}_1^{\ell-1}\mathbf{G}_1\boldsymbol{\Sigma}(\boldsymbol{\Phi}_1^{\ell-1}\mathbf{G}_1)' \qquad \text{(A.3)} \\ + \boldsymbol{\Phi}_1^{k_1+r} \sum_{\ell=1}^{k_2-k_1} \boldsymbol{\Phi}_2^{\ell-1}\mathbf{G}_2\boldsymbol{\Sigma}(\boldsymbol{\Phi}_\mathbf{1}^{\mathbf{k_1+r}}\boldsymbol{\Phi}_2^{\ell-1}\mathbf{G}_2)' &+ \boldsymbol{\Phi}_1^{k_1+r}\boldsymbol{\Phi}_2^{k_2-k_1} \sum_{\ell=1}^{k-k_2-1} \boldsymbol{\Phi}^{\ell-1}\mathbf{G}_3\boldsymbol{\Sigma}(\boldsymbol{\Phi}_\mathbf{1}^{\mathbf{k_1+r}}\boldsymbol{\Phi}_\mathbf{2}^{\mathbf{k_2-k_1}}\boldsymbol{\Phi}^{\ell-1}\mathbf{G}_3)', \end{aligned}$$

where now

$$\mathbf{G}_n = (\boldsymbol{\Phi}_n\mathbf{J} + \mathbf{Z}_n),\ \ n = 1,2,3.$$

Since $(\mathbf{XY})^{\otimes 2} = \mathbf{X}^{\otimes 2}\mathbf{Y}^{\otimes 2}$ and $(\mathbf{X}^r)^{\otimes 2} = (\mathbf{X}^{\otimes 2})^r$, using

$$\begin{aligned} \sum_{\ell=1}^{k_n-k_{n-1}} [(\boldsymbol{\Phi}_n^{\ell-1})(\boldsymbol{\Phi}_n\mathbf{J} + \mathbf{Z}_n)]^{\otimes 2} \ &= \ \sum_{\ell=1}^{k_n-k_{n-1}} [(\boldsymbol{\Phi}_n)^{\otimes 2}]^{\ell-1}(\boldsymbol{\Phi}_n\mathbf{J} + \mathbf{Z}_n)^{\otimes 2} \\ = [\mathbf{I}^{\otimes 2} - (\boldsymbol{\Phi}_n^{k_n-k_{n-1}})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\boldsymbol{\Phi}_n)^{\otimes 2}]^{-1}&(\boldsymbol{\Phi}_n\mathbf{J} + \mathbf{Z}_n)^{\otimes 2},\ \ n = 1,2,3 \end{aligned}$$

(we recall that $k_0 = -r$ and $k_3 = k$) the vec form of the right hand side of eq. (A.3) gives

$$\{\mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_1)^{\otimes 2}]^{-1}(\mathbf{\Phi}_1\mathbf{J} + \mathbf{Z}_1)^{\otimes 2}$$
$$+ (\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}[\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2^{k_2-k_1})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2)^{\otimes 2}]^{-1}(\mathbf{\Phi}_2\mathbf{J} + \mathbf{Z}_2)^{\otimes 2}$$
$$+ (\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}(\mathbf{\Phi}_2^{k_2-k_1})^{\otimes 2}[\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}^{k-k_2-1})^{\otimes 2}](\mathbf{I}^{\otimes 2} - \mathbf{\Phi}^{\otimes 2})^{-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})^{\otimes 2}\}\mathbf{s}. \quad (A.4)$$

Similarly, the vec form of the right hand side of eq. (A.2) is given by

$$(\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}(\mathbf{\Phi}_2^{k_2-k_1})^{\otimes 2}(\mathbf{\Phi}^{k-k_2-1})^{\otimes 2}(\mathbf{I}^{\otimes 2} - \mathbf{\Phi}^{\otimes 2})^{-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})^{\otimes 2}\mathbf{s}. \quad (A.5)$$

Finally, taking the vec form of both sides of eq. (A.1), and using eqs. (A.4) and (A.5), we obtain eq. (10) as claimed.

■

Table A.1. 'First-order persistence' for each of the three periods and models.

	$P(y_t\,\vert\varepsilon)$			$P(y_t\,\vert v)$			$P(y_t\,\vert\varepsilon) + P(y_t\,\vert v)$		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
1947Q1-1966Q4	1.68	2.20	1.60	1.34	1.610	2.13	3.02	3.81	3.74
1967Q1-2008Q2	3.61	2.20	2.91	2.87	3.89	3.88	6.48	6.09	6.79
2008Q3-2016Q3	1.85	2.20	2.91	1.47	2.49	2.78	3.33	4.69	5.70

Note: We used the limit as $k \to \infty$ of eq. (9) with unit mean and volatility shocks, that is $\overline{\mathbf{\Phi}}_n \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ with $\overline{\mathbf{\Phi}}_n = \mathbf{J} + (\mathbf{I} - \mathbf{\Phi}_n)^{-1}(\mathbf{\Phi}_n\mathbf{J} + \mathbf{Z}_n)$, $n = 1, 2, 3,$ to calculate the corresponding first-order persistence. For each model and period $P(y_t\,|\varepsilon)$ and $P(y_t\,|v)$ are the $(1,1)$ and $(1,2)$ elements of $\overline{\mathbf{\Phi}}_n$.