JOURNAL OF TIME SERIES ANALYSIS
J. Time. Ser. Anal. **36:** 706–720 (2015)
Published online 3 February 2015 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12119

SPECIAL ISSUE

ON THE TRANSMISSION OF MEMORY IN GARCH-IN-MEAN MODELS

CHRISTIAN CONRAD[a] AND MENELAOS KARANASOS[b,*]

[a] *Alfred-Weber-Institute, Heidelberg University, Heidelberg, Germany*
[b] *Economics and Finance, Brunel University, Uxbridge, West London, UK*

In this article, we show that in times series models with in-mean and level effects, persistence will be transmitted from the conditional variance to the conditional mean and vice versa. Hence, by studying the conditional mean/variance independently, one will obtain a biased estimate of the true degree of persistence. For the specific example of an AR(1)-APARCH(1,1)-in-mean-level process, we derive the autocorrelation function, the impulse response function and the optimal predictor. Under reasonable assumptions, the AR(1)-APARCH(1,1)-in-mean-level process will be observationally equivalent to an autoregressive moving average (ARMA)(2,1) process with the largest autoregressive root being close to one. We illustrate the empirical relevance of our results with applications to S&P 500 return and US inflation data.

Received 25 November 2013; Revised 1 September 2014; Accepted 18 November 2014

Keywords: Conditional heteroscedasticity; GARCH-in-mean; persistence; unit root tests.
JEL. E31; E58; C12; C22; C52.

1. INTRODUCTION

Many economic time series are characterized by an autocorrelation structure that makes it difficult to classify the series as being either stationary I(0) or nonstationary I(1). A primary example for such a series is inflation rates. Conventional wisdom then suggests that we employ unit root tests in order to base the econometric analysis either on the level of such a series or on the first difference. Clearly, the decision whether the series is treated as being I(0) or I(1) has important implications for the subsequent modelling, hypothesis testing, forecasting and the like. As pointed out by Haldrup and Jansson (2006), a frequent criticism of unit root tests concerns the poor power and size properties that many such tests exhibit (e.g. DeJong *et al.*, 1992a, 1992b). Besides the observation that many economic time series are strongly dependent over time, there is the stylized fact that for the same series, one typically finds generalized autoregressive conditional heteroskedasticity (GARCH) effects with highly persistent volatility. Moreover, economic theory often suggests that the level and the second conditional moment of these series should be interrelated. For example, Cukierman and Meltzer (1986) and Holland (1995) argue that inflation uncertainty has either a positive or a negative effect on the level of inflation, while Friedman (1977) and Ball (1992) rationalize an effect of the level of inflation on its second conditional moment. Against this theoretical background, the phenomena of persistence in the level and in the conditional variance are usually analysed and treated independently. For example, standard unit root tests are based either on the assumption that the variance of the series is constant or on the assumption that some type of heteroscedasticity is present (e.g. Kim and Schmidt, 1993; Seo, 1999; Ling and Li, 2003; Ling *et al.*, 2003; Rodrigues and Rubia, 2005; Kourogenis and Pittis, 2008) but ignore the possibility that the volatility has a direct effect on the level.[1]

* Correspondence to: Menelaos Karanasos, Economics and Finance, Brunel University, Uxbridge, West London, UB3 3PH, UK. E-mail: Menelaos.Karanasos@brunel.ac.uk

[1] For the important research issue of testing for unit roots in the presence of permanent volatility shifts, see e.g. Cavaliere and Taylor (2007, 2008a).

In this article, we consider an AR-APARCH-in-mean-level (AR-APARCH-ML) process, that is, a model in which the conditional variance affects the conditional mean and the level affects the conditional variance. This model has been introduced by Engle *et al.* (1987) and applied by, for example, Grier *et al.* (2004), Conrad *et al.* (2010), Conrad and Karanasos (2010, 2014) and Karanasos and Zeng (2013). We provide a new interpretation of the model's properties by arguing that it has an observationally equivalent representation as an ARMA(2,1) process. The largest autoregressive root of the AR part will, under reasonable assumptions, be close to and statistically indistinguishable from one. This means that in empirical applications, the process will appear to be an ARIMA(1, 1, 1). Most importantly, the largest root of the AR part is closely linked to the persistence of the conditional variance of the process. We then show how the in-mean effect leads to a 'transmission of memory' from the conditional variance to the conditional mean and, thereby, affects the persistence properties of the level process. We illustrate this important point by deriving the autocorrelation function (ACF), the impulse response function (IRF), new measures of persistence and the optimal predictor for the level process.[2] We also show that procedures that ignore the transmission mechanism and do not distinguish between the different types of persistence in the two moments lead to biased estimates of the persistence in either the mean or the variance. In particular, using a Monte Carlo study, we show that in the presence of an in-mean effect, conventional unit root tests tend to falsely indicate that the underlying process is I(1). Further, we illustrate the empirical importance of our results by applying our model to S&P 500 return and US inflation data. Finally, it is important to note that our results are closely related to Rahbek and Nielsen (2014) who show that in a multivariate model in which lagged levels enter both the conditional mean and the conditional variance, the multivariate process can be strictly stationary and ergodic although the individual series have unit roots.

The outline of the article is as follows. Section 2 presents the model and its properties. Section 3 provides the empirical applications. Section 4 concludes the article. Additional material, including the optimal predictor, can be found in the Supplementary Information.

2. THE MODEL

The AR(1)-APARCH(1,1)-in-mean-level [AR(1)-APARCH(1,1)-ML] model is given by

$$(1-\phi L)y_t = \varphi + \vartheta h_t^{\frac{\delta}{2}} + \varepsilon_t \tag{1}$$

with $\varepsilon_t = e_t h_t^{\frac{1}{2}}$, where L is the lag operator, $\delta > 0$, $\{e_t\}$ is a sequence of i.i.d. random variables with zero mean and finite variance, $\mathbb{E}\left(e_t^2\right)$, and h_t is the conditional variance of y_t (see, among others, Christensen and Nielsen, 2007). The AR(1) parameter ϕ naturally measures the *intrinsic* persistence in the level of y_t. The power-transformed conditional variance, $h_t^{\frac{\delta}{2}}$, 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 h_t is specified as an asymmetric power ARCH(1,1)-level [APARCH(1,1)-L] process:

$$(1-\beta L)h_t^{\frac{\delta}{2}} = \omega + \alpha f(\varepsilon_{t-1}) + \gamma y_{t-1}, \tag{2}$$

where $f(\varepsilon_{t-1}) = [|\varepsilon_{t-1}| - \varsigma\varepsilon_{t-1}]^{\delta} = f(e_{t-1})h_{t-1}^{\delta/2}$ with $|\varsigma| < 1$ (for details on the power ARCH model, see e.g. Karanasos and Kim, 2006). By including the lagged y_t in the conditional variance equation (the so-called level effect) and $h_t^{\frac{\delta}{2}}$ in the mean equation (the so-called in-mean effect), we allow for simultaneous feedback between the two variables. The following conditions are necessary and sufficient for $h_t > 0$ for all t : $\omega > 0, \alpha, \beta, \gamma \geq 0$ and $y_t \geq 0$ for all t.

[2] For prediction in ARMA models with GARCH-M effects, see also Karanasos (2001).

DOI: 10.1111/jtsa.12119

Hereafter, we will denote $\mu_r = \mathbb{E}\left(h_t^{\frac{\delta r}{2}}\right)$. Expressions for μ_1 and μ_2 are given in Proposition 2 and equations (B.7/B.8) in the Supplementary Information. The APARCH(1,1)-L formulation in equation (2) can readily be interpreted as an ARMA(1,1)-L process for the conditional variance:

$$(1-cL)h_t^{\frac{\delta}{2}} = \omega + \alpha v_{t-1} + \gamma y_{t-1}, \tag{3}$$

where $c = \alpha\kappa^{(1)} + \beta$, with $\kappa^{(r)} = \mathbb{E}\left\{[f(e_t)]^r\right\}$, and $v_t = f(\varepsilon_t) - \mathbb{E}\left[f(\varepsilon_t)|\mathcal{F}_{t-1}\right] = \left(f(e_t) - \kappa^{(1)}\right)h_t^{\frac{\delta}{2}}$ is, by construction, an uncorrelated term with expected value 0 and $\mathbb{E}\left(v_t^2\right) = \sigma_v^2 = \mu_2\widetilde{\kappa}$ with $\widetilde{\kappa} = \left[\kappa^{(2)} - \left(\kappa^{(1)}\right)^2\right]$. While the ε_t are the innovations to the level of y_t, the v_t can be considered as the innovations to the conditional variance of y_t.

Note that the parameter c measures the *intrinsic memory* or *persistence* in the conditional variance. The extreme case in which $c = 1$ is well known as the integrated GARCH model (see Engle and Bollerslev, 1986).[3]

Also, notice that $\mathbb{E}\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2 = \mu_{2/\delta}\mathbb{E}\left(e_t^2\right)$ and $\mathbb{E}\left(\varepsilon_t v_t\right) = \sigma_{\varepsilon v} = \mu_{1+1/\delta}\overline{\kappa}$ with $\overline{\kappa} = \mathbb{E}[e_t f(e_t)]$. In other words, the covariance matrix of the two shocks ε_t and v_t is given by

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

If e_t is standard normal, then $\mathbb{E}\left(e_t^2\right) = 1$, and $\kappa^{(r)}$ and $\overline{\kappa}$ are given by

$$\kappa^{(r)} = \frac{1}{\sqrt{\pi}}\left[(1-\varsigma)^{r\delta} + (1+\varsigma)^{r\delta}\right]2^{\left(\frac{r\delta}{2}-1\right)}\Gamma\left(\frac{r\delta+1}{2}\right), \tag{5}$$

$$\overline{\kappa} = \frac{1}{\sqrt{2\pi}}\left[(1-\varsigma)^{\delta} - (1+\varsigma)^{\delta}\right]2^{(\delta/2)}\Gamma\left(\frac{\delta}{2}+1\right), \tag{6}$$

where $\Gamma(\cdot)$ is the Gamma function. Next, we investigate some special cases under the assumption that e_t is standard normal. First, we are interested in the behaviour of $\kappa^{(1)}$ when $\delta = 2$ or $\delta = 1$. It directly follows from equation (5) that

$$\kappa^{(1)} = \begin{cases} 1+\varsigma^2 & \text{when } \delta = 2 \\ \sqrt{2/\pi} & \text{when } \delta = 1. \end{cases}$$

That is, the persistence parameter $c = \alpha\kappa^{(1)} + \beta$ will increase in $|\varsigma|$ for $\delta = 2$ and will be independent of the asymmetry term for $\delta = 1$. Further, for $\varsigma = 0$, equation (6) shows that $\overline{\kappa} = 0$ and, hence, ε_t and v_t will be uncorrelated for all $\delta > 0$. Finally, for $\delta = 2$ and $\varsigma = 0$, we obtain $\tilde{\kappa} = \mathbb{E}\left[e_t^4\right] - \left(\mathbb{E}\left[e_t^2\right]\right)^2 = 2$, and $\mathbf{\Sigma}$ reduces to

$$\mathbf{\Sigma} = \begin{bmatrix} \mu_1 & 0 \\ 0 & 2\mu_2 \end{bmatrix}.$$

[3] The parameter c determines the rate of decay of the IRF of the GARCH(1,1). For $c = 1$, the IRF is constant (see e.g. Conrad and Karanasos, 2006).

  J. Time. Ser. Anal. **36:** 706–720 (2015)
DOI: 10.1111/jtsa.12119

On the other hand, for $\delta = 1$, equations (5) and (6) reduce to $\kappa^{(2)} = 1 + \varsigma^2$ and $\bar{\kappa} = -\varsigma$, which implies that $\mathbf{\Sigma}$ becomes

$$\mathbf{\Sigma} = \mu_2 \begin{bmatrix} 1 & -\varsigma \\ -\varsigma & 1 + \varsigma^2 - 2/\pi \end{bmatrix}. \tag{7}$$

Figure 1 shows the covariance between ε_t and v_t for ς varying between -0.5 and 0.5 and $\delta \in \{1, 1.5, 2\}$. The other parameters are fixed as described in the caption of Figure 1. The figure shows that ε_t and v_t are negatively (positively) correlated for positive (negative) values of ς. For example, in empirical applications using stock return data, we would expect a positive ς due to the leverage effect and, hence, negative return innovations to coincide with positive volatility innovations. On the contrary, if we think about applications to inflation data, we would expect a negative ς because higher than expected inflation rates are typically associated with positive volatility innovations.

Next, we derive the 'univariate representations' of y_t and $h_t^{\delta/2}$. That is, we write y_t as an ARMA process that no longer depends on $h_t^{\delta/2}$. Similarly, we obtain an ARMA representation for $h_t^{\delta/2}$ that does not involve y_{t-1}.

Proposition 1. The univariate ARMA representations of y_t and $h_t^{\frac{\delta}{2}}$ are given by

$$\left(1 - a_1 L - a_2 L^2\right) y_t = \varphi^* + (1 - cL)\varepsilon_t + \vartheta \alpha v_{t-1}, \tag{8}$$

$$\left(1 - a_1 L - a_2 L^2\right) h_t^{\frac{\delta}{2}} = \omega^* + \gamma \varepsilon_{t-1} + (1 - \phi L)\alpha v_{t-1}, \tag{9}$$

where $a_1 = \phi + c + \vartheta\gamma$, $a_2 = -\phi c$, $\varphi^* = \varphi(1 - c) + \vartheta\omega$ and $\omega^* = \omega(1 - \phi) + \varphi\gamma$.

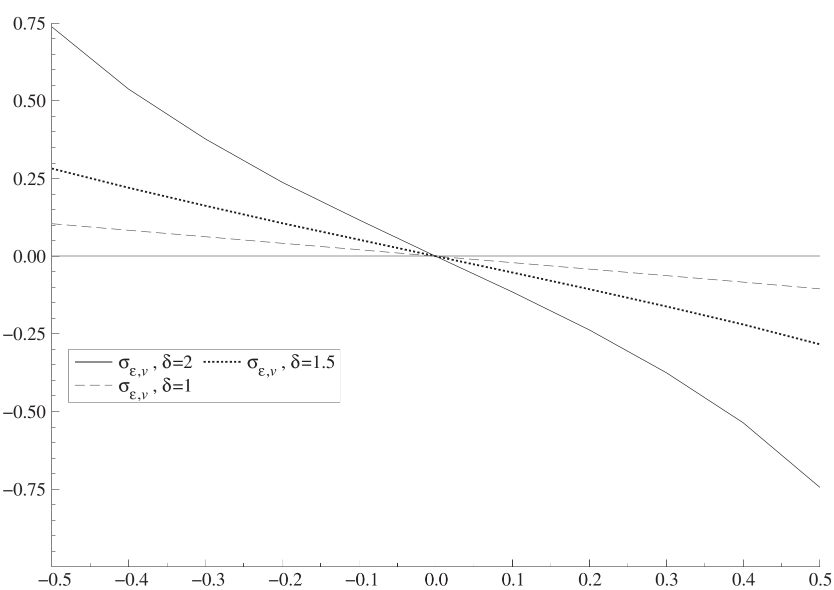

Figure 1. $\sigma_{\varepsilon v}$ as a function of ς. The other parameters are given by $\varphi = 0.05$, $\phi = 0.4$, $\vartheta = 0.5$, $\omega = 0.1$, $\alpha = 0.1$, $\beta = 0.7$, $\gamma = 0.0$ and $\delta \in \{1, 1.5, 2\}$

DOI: 10.1111/jtsa.12119

As Proposition 1 shows, the model given by equations (1) and (2) has a univariate ARMA(2,1) representation in the level and an ARMA(2,2) representation in the conditional variance. Because of the existence of simultaneous feedback, that is, in-mean as well as level effects, both the level and the conditional variance depend on the two shocks ε_t and v_t. Further, note that y_t and $h_t^{\delta/2}$ have the same autoregressive polynomial.

Assumption A1 (Stationarity)**.** The inverse roots λ_1 and λ_2 of $\left(1 - a_1 L - a_2 L^2\right)$ lie inside the unit circle. Moreover, without loss of generality, we assume that $\lambda_1 \neq \lambda_2$.

Assumption (A1) implies that the ARMA(2,1) process given by equation (8) is covariance stationary. It also implies that $a_1 + a_2 < 1$. The representation in equations (8) and (9) illustrates that measures of persistence in the level/conditional variance that are based on a_1 and a_2 will confuse intrinsic persistence with persistence that is due to the in-mean/level transmission channel.[4]

Proposition 2. When Assumption (A1) holds, the unconditional expectation of $h_t^{\frac{\delta}{2}}$ exists if $\omega^* > 0$ and is given by

$$\mu_1 = \frac{\omega^*}{1 - a_1 - a_2}. \tag{10}$$

Clearly, for $\delta = 2$ and $\gamma = \varsigma = 0$, the result in Proposition 2 reduces to the well-known formula for the GARCH(1,1) process: $\mu_1 = \mathbb{E}(h_t) = \mathbb{E}\left(\varepsilon_t^2\right) = \omega/(1 - \alpha - \beta)$. Further, note that the existence of μ_1 guarantees that of $\mu_{2/\delta}$ only if $\delta \geq 2$. Similarly, the existence of the second moment μ_2 guarantees that of $\mu_{1+1/\delta}$ only if $\delta \geq 1$.

Next, we discuss the moving average part of equation (8), which is a function of the mean shocks ε_t and the conditional variance shocks v_t. As discussed earlier, the two shocks will be uncorrelated if $\varsigma = 0$. Since the sum of an MA(1) and a white noise process is again an MA(1), equation (8) can be rewritten as

$$\left(1 - a_1 L - a_2 L^2\right) y_t = \varphi^* + (1 - \theta L)\eta_t, \tag{11}$$

where η_t is an uncorrelated error process with mean zero and variance σ_η^2. The parameters θ and σ_η^2 can be expressed as

$$\theta = \frac{-\sigma_0 \pm \sqrt{\sigma_0^2 - 4\sigma_1^2}}{2\sigma_1} \quad \text{and} \quad \sigma_\eta^2 = \frac{\sigma_1}{-\theta},$$

where $\sigma_0 = (1 + c^2)\sigma_\varepsilon^2 + (\vartheta\alpha)^2\sigma_v^2 - 2c\vartheta\alpha\sigma_{\varepsilon v}$ and $\sigma_1 = -c\sigma_\varepsilon^2 + \vartheta\alpha\sigma_{\varepsilon v}$. Note that (i) θ is real if and only if $\sigma_0^2 > 4\sigma_1^2$ and (ii) $0 \lesseqgtr \theta$ if $\sigma_1 \lesseqgtr 0$, that is, if $\vartheta\alpha\sigma_{\varepsilon v} \lesseqgtr c\sigma_\varepsilon^2$. Equation (11) is interesting because in Section 3, it will allow us to discuss under which parameter constellations our model is able to reproduce the typical time series properties of stock returns and inflation.

2.1. Covariance Structure

Equations (8) and (11) suggest that the ACF of y_t can be very persistent although the intrinsic persistence in the level, that is, ϕ, may be low. Intuitively, this will be the case if there is high persistence in the conditional variance and a sufficiently strong in-mean effect. In this subsection, we formally derive the autocovariance structure of y_t,

[4] Such a measure could be the largest autoregressive root or the sum of the autoregressive coefficients. See Section 2.3.

  J. Time. Ser. Anal. **36:** 706–720 (2015)
DOI: 10.1111/jtsa.12119

$\mathbb{C}ov_k(y_t)$, $k \in \mathbb{N}$, in the presence of the in-mean term $\vartheta \neq 0$.[5] Special attention will be given to the role played by the power transformation, δ. Lemma 1 provides the (co)variance structure of y_t for the case that $\gamma = 0$. The general result including the level effect can be found in Proposition 5 in the Supplementary Information.

Lemma 1. Let $\gamma = 0$. Then, $\lambda_1 = \phi$, $\lambda_2 = c$ and

$$\mathbb{V}(y_t) = \frac{1}{1-\phi^2}\left\{\sigma_\varepsilon^2 + \frac{\vartheta\alpha}{1-\phi c}\left[\frac{\sigma_v^2\vartheta\alpha(1+\phi c)}{1-c^2} + 2\sigma_{\varepsilon v}\phi\right]\right\}, \tag{12}$$

$$\mathbb{C}ov_k(y_t) = \frac{\phi^k}{1-\phi^2}\sigma_\varepsilon^2 + \vartheta^2\alpha^2\sigma_v^2\lambda^{(k)} + \vartheta\alpha\sigma_{\varepsilon v}\phi^{(k)} \quad for \quad k \geq 1, \tag{13}$$

where $\lambda^{(k)}$ and $\phi^{(k)}$ are given by

$$\lambda^{(k)} = \frac{1}{(1-\phi c)(\phi - c)}\left[\frac{\phi^{1+k}}{1-\phi^2} - \frac{c^{1+k}}{1-c^2}\right]$$

$$\phi^{(k)} = \frac{1}{(\phi - c)(1-\phi c)}\left[\frac{\phi^k\left(1+\phi^2-2\phi c\right)}{1-\phi^2} - c^k\right].$$

Lemma 1 shows that the variance as well as the covariances of y_t will be equal to the ones of an AR(1) if there is no in-mean effect. For $\vartheta \neq 0$, both are different from the corresponding expressions for an AR(1) because memory is transmitted from the conditional variance to the level of the series. In particular, it is interesting to discuss the properties of the covariance function. Equation (13) shows that $\mathbb{C}ov_k(y_t)$ is a sum of three terms. The first term is simply the covariance function of an AR(1). The second term is due to the fact that $\vartheta \neq 0$ and $h_t^{\delta/2}$ itself is correlated. If $\phi = 0$, then $\lambda^{(k)} = c^k/(1-c^2)$, and the second term is equal to $\mathbb{C}ov_k\left(h_t^{\delta/2}\right)$ scaled by the squared in-mean term.[6] Although y_t has no intrinsic persistence in this case, it will be correlated because of the transmission of memory from the conditional variance due to the in-mean effect. Finally, even in the presence of an in-mean effect, the third term will be nonzero only if ε_t and v_t are correlated, that is, $\varsigma \neq 0$.[7]

In the following, we will graphically illustrate the consequences of the in-mean term for the dependence structure of y_t. For simplicity, in addition to $\gamma = 0$, we also assume that $\varsigma = 0$. As equation (13) shows, in this case, $\mathbb{C}ov_k(y_t) > \sigma_\varepsilon^2\phi^k/(1-\phi^2)$, that is, greater than the autocovariance of the pure AR(1), whenever $\vartheta \neq 0$. Further, $\mathbb{C}ov_k(y_t)$ is increasing in $|\vartheta|$, that is, in the strength of the in-mean effect. Figure 2 shows the ACFs of y_t for $\phi = 0.1$, that is, a low degree of intrinsic level persistence, and ϑ increasing from 0.0 to 2.0 in steps of 0.5 with $\alpha = 0.1$ (upper panel), $\alpha = 0.15$ (middle panel) and $\alpha = 0.19$ (lower panel) while $\beta = 0.8$. The power term δ is assumed to be either 2 (left panel) or 1 (right panel). The lowest line ($\vartheta = 0$) in each panel corresponds to the ACF of the pure AR(1) model that is given by ϕ^k. The figure clearly shows how increasing the value of ϑ increases the correlation of y_t. When the conditional variance is almost integrated ($c = \alpha + \beta = 0.99$) and $\delta = 2$, the ACF of y_t essentially behaves like one for a nonstationary process. Since there is almost no intrinsic persistence in the

[5] The corresponding results for the correlation structure of the power-transformed conditional variance are presented in the Supplementary Information B.3 together with the cross correlations between y_t and $h_t^{\delta/2}$.

[6] See the expression for $\mathbb{C}ov_k\left(h_t^{\delta/2}\right)$ in Lemma 3 in the Supplementary Information for $\gamma = 0$.

[7] Interestingly, when $\delta = 1$ and $\vartheta = \varsigma(1-c^2)/\left(c\alpha\left(1+\varsigma^2-2/\pi\right)\right)$, $\mathbb{C}ov_k(y_t)$ reduces to $\mu_2(1+\vartheta\alpha\varsigma/c)\phi^k/(1-\phi^2)$. This result is in line with the fact that in this case, $\theta = c$ and the ARMA(2,1) representation reduces to an AR(1) process (see the Supplementary Information B.1).

DOI: 10.1111/jtsa.12119

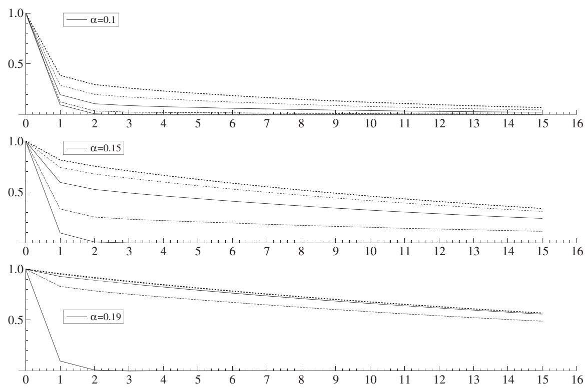

Figure 2. Autocorrelation function (ACF) of y_t for $\phi = 0.1$ and ϑ increasing from 0.0 (lowest ACF) to 2.0 (highest ACF) in steps of 0.5. Left panel: $\delta = 2$, right panel: $\delta = 1$. The remaining parameters are $\varphi = 0$, $\omega = 0.1$, $\beta = 0.8$ and $\gamma = 0$

level of y_t, this behaviour is purely due to the transmission of memory from the conditional variance to the level.[8] Thus, an applied researcher who investigates the ACF of y_t might erroneously come to the conclusion that the process is integrated of order one in the level and, hence, model the first difference of y_t. In the next section, we will show that standard unit root tests will tend to corroborate this false conclusion.

2.2. Unit Root Tests

We investigate the performance of standard unit root tests in the presence of an in-mean effect. A typical empirical example of a very persistent and potentially integrated conditionally heteroscedastic series is inflation rates.[9] We generate y_t series, $t = 1, \ldots, T$, with $T = 1000$, according to our AR(1)-GARCH(1,1)-M model. For simplicity, we set $\delta = 2$ and $\gamma = 0$. The constants in the mean and conditional variance are set to $\varphi = 0$ and $\omega = 0.1$. The innovation e_t is assumed to be standard normal. We then apply the augmented Dickey–Fuller (DF) test (including a constant) and the generalized least squares (GLS)-demeaned DF test (DF-GLS) as suggested by Elliot *et al.* (1996) to the y_t series.[10] In both cases, the lag order is automatically selected by the Bayesian information criterion (BIC). The following tables report the fraction of cases in which the null hypothesis of a unit root is rejected at the 5% significance level. The results are based on 10,000 Monte Carlo replications.

We first consider the case $\phi = 1$; that is, we are under the null hypothesis. Table I shows the rejection rates for $\beta = 0.85$ (low persistence in the conditional variance) and $\beta = 0.88$ (high persistence in the conditional variance) in combination with $\varsigma = 0$ or $\varsigma = 0.3$. The in-mean parameter ϑ is varying from -2 to 2 in steps of 0.5. Clearly, for $\vartheta = 0$, both tests are only slightly oversized. However, for the standard DF test, the size distortion becomes stronger with $|\vartheta|$ increasing. That is, although y_t is I(1), the DF test tends to falsely reject the null hypothesis in the presence of an in-mean effect. For a given value of ϑ, the size distortion is stronger in the high persistence case and in the presence of asymmetry. In sharp contrast, whenever $\vartheta \neq 0$, the DF-GLS test never rejects the null hypothesis, that is the DF-GLS is clearly undersized.

Next, we consider the behaviour of both tests under the alternative. In Table II, we set $\phi = 0.9$; that is, y_t is stationary. When $\vartheta = 0$, both tests reject the null in 100% of the replications. However, the table clearly shows that the power of both tests deteriorates with $|\vartheta|$ increasing. That is, the larger $|\vartheta|$ is, the more persistence is

[8] Figures 5 and 6 in the Supplementary Information illustrate the behaviour of the ACF when y_t itself has moderate ($\boldsymbol{\phi = 0.5}$) and high ($\boldsymbol{\phi = 0.9}$) intrinsic persistence.
[9] Cavaliere and Taylor (2007) and Cavaliere and Taylor (2008b) illustrate the empirical relevance of their unit root testing techniques using US inflation data.
[10] We are grateful to Robinson Kruse for sharing his code for simulating the GLS-demeaned DF test (see Demetrescu and Kruse, 2013).

DOI: 10.1111/jtsa.12119

Table I. Rejection rates under the null hypothesis ($\phi = 1$)

	$\beta = 0.85$				$\beta = 0.88$			
	$\varsigma = 0$		$\varsigma = 0.3$		$\varsigma = 0$		$\varsigma = 0.3$	
ϑ	DF	DF-GLS	DF	DF-GLS	DF	DF-GLS	DF	DF-GLS
−2.00	0.10	0.00	0.18	0.00	0.26	0.00	0.24	0.00
−1.50	0.11	0.00	0.17	0.00	0.25	0.00	0.29	0.00
−1.00	0.09	0.00	0.17	0.00	0.25	0.00	0.26	0.00
−0.50	0.06	0.00	0.14	0.00	0.21	0.00	0.28	0.00
0.00	0.05	0.06	0.06	0.06	0.08	0.06	0.07	0.07
0.50	0.05	0.00	0.08	0.00	0.22	0.00	0.22	0.00
1.00	0.10	0.00	0.15	0.00	0.25	0.00	0.26	0.00
1.50	0.11	0.00	0.16	0.00	0.27	0.00	0.27	0.00
2.00	0.11	0.00	0.16	0.00	0.26	0.00	0.27	0.00

Notes: The table reports the rejection rates of the DF and DF-GLS tests over the 10,000 replications at the 5% nominal level.
DF, Dickey–Fuller; DF-GLS, Dickey–Fuller-generalized least square.

Table II. Rejection rates under the alternative ($\phi = 0.9$)

	$\beta = 0.85$				$\beta = 0.88$			
	$\varsigma = 0$		$\varsigma = 0.3$		$\varsigma = 0$		$\varsigma = 0.3$	
ϑ	DF	DF-GLS	DF	DF-GLS	DF	DF-GLS	DF	DF-GLS
−2.00	0.52	0.00	0.07	0.00	0.04	0.00	0.00	0.00
−1.50	0.64	0.00	0.10	0.00	0.06	0.00	0.00	0.00
−1.00	0.83	0.00	0.29	0.00	0.12	0.00	0.00	0.00
−0.50	0.99	0.83	0.74	0.24	0.55	0.00	0.10	0.00
0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.99	0.85	0.92	0.63	0.55	0.00	0.18	0.00
1.00	0.86	0.00	0.58	0.00	0.11	0.00	0.03	0.00
1.50	0.64	0.00	0.32	0.00	0.07	0.00	0.01	0.00
2.00	0.54	0.00	0.17	0.00	0.04	0.00	0.01	0.00

Notes: The table reports the rejection rates of the DF and DF-GLS tests over the 10,000 replications at the 5% nominal level.
DF, Dickey–Fuller; DF-GLS, Dickey–Fuller-generalized least square.

transmitted from the conditional variance to the level of y_t, and, hence, both tests falsely do not reject the null hypothesis of y_t being I(1). The results clearly suggest that unit root tests will have low power against alternatives that allow for in-mean effects in combination with persistent conditional variances. In this case, a stationary y_t process may be easily confused with a process that is integrated of order one in the level.

2.3. Measures of Persistence

The aforementioned considerations suggest that conventional measures of persistence might result in misleading conclusions regarding the persistence in the level of the y_t process. The most often applied measures are (i) the largest autoregressive root, which we denote by $\lambda^* = \max(\lambda_1, \lambda_2)$ and (ii) the sum of the coefficients in the autoregressive polynomial, that is, $a_1 + a_2$ (see e.g. Pivetta and Reis, 2007).[11] Obviously, both measures would ignore the presence of the in-mean effect and, hence, potentially overestimate the persistence in the mean, which is partly induced by the persistence in the conditional variance.

We follow Fiorentini and Sentana (1998) who argue that any reasonable measure of shock persistence should be based on the IRF. For a univariate process x_t with Wold representation $x_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}$, they define the persistence of a shock u_t on x_t as $P_\infty(x_t|u_t) = \mathbb{V}ar(x_t)/\mathbb{V}ar(u_t) = \sum_{j=0}^{\infty} \psi_j^2$. Clearly, $P_\infty(x_t|u_t)$ will take

[11] If $\vartheta\gamma > 0$, then $a_1 + a_2 = \phi + c(1-\phi) + \vartheta\gamma > \phi + c(1-\phi) > \phi, c$.

DOI: 10.1111/jtsa.12119

its minimum value of one if x_t is white noise and will be infinite for a nonstationary process. A natural measure of the *interim* persistence of the effect of a shock n periods after its occurrence is given by $P_n(x_t|u_t) = \sum_{j=0}^{n} \psi_j^2$. This measure can be calculated for both stationary and nonstationary processes.

In the following, we suggest persistence measures that are able to distinguish between the effects of a *mean shock* and a *volatility shock* on the level and conditional variance respectively. We first obtain the bivariate Wold representation of the univariate processes for y_t and $h_t^{\delta/2}$ given in equations (8) and (9).

Proposition 3. Let Assumption (A1) hold. Then, equations (8) and (9) admit the Wold representation

$$\begin{pmatrix} y_t \\ h_t^{\frac{\delta}{2}} \end{pmatrix} = \begin{pmatrix} y^* \\ \mu_1 \end{pmatrix} + \begin{pmatrix} \psi_{y\varepsilon}(L) & \psi_{yv}(L) \\ \psi_{h\varepsilon}(L) & \psi_{hv}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix}, \tag{14}$$

where $y^* = \varphi^*/(1 - a_1 - a_2)$ and $\psi_{ij}(L) = \sum_{k=0}^{\infty} \psi_{ij}^{(k)} L^k, i = y, h; j = \varepsilon, v$ with

$$\begin{aligned}
\psi_{y\varepsilon}^{(0)} &= 1,\ \psi_{y\varepsilon}^{(k)} = \left[\frac{\lambda_1^k(\lambda_1 - c)}{\lambda_1 - \lambda_2} + \frac{\lambda_2^k(\lambda_2 - c)}{\lambda_2 - \lambda_1}\right],\ k \geq 1,\\
\psi_{yv}^{(0)} &= 0,\ \psi_{yv}^{(k)} = \vartheta\alpha\left(\frac{\lambda_1^k}{\lambda_1 - \lambda_2} + \frac{\lambda_2^k}{\lambda_2 - \lambda_1}\right),\ k \geq 1,\\
\psi_{h\varepsilon}^{(0)} &= 0,\ \psi_{h\varepsilon}^{(k)} = \gamma\left(\frac{\lambda_1^k}{\lambda_1 - \lambda_2} + \frac{\lambda_2^k}{\lambda_2 - \lambda_1}\right),\ k \geq 1,\\
\psi_{hv}^{(0)} &= 0,\ \psi_{hv}^{(1)} = \alpha, \psi_{hv}^{(k)} = \alpha\left[\frac{\lambda_1^{k-1}(\lambda_1 - \phi)}{\lambda_1 - \lambda_2} + \frac{\lambda_2^{k-1}(\lambda_2 - \phi)}{\lambda_2 - \lambda_1}\right],\ k \geq 2.
\end{aligned}$$

If $\sigma_{\varepsilon v} = 0$, then ε_t and v_t can be viewed as 'structural' shocks and the 'IRFs' of a one unit mean or variance shock to the process y_t are directly given by $\psi_{y\varepsilon}^{(k)}$ and $\psi_{yv}^{(k)}$. In general, the shocks ε_t and v_t will be correlated with covariance matrix $\mathbf{\Sigma}$ [see equation (4)]. Next, we define two uncorrelated white noise shocks $\tilde{\varepsilon}_t$ and $\tilde{v}_t$ with variances equal to one. These orthogonal shocks can be obtained from the original shocks via the transformation

$$\begin{pmatrix} \tilde{\varepsilon}_t \\ \tilde{v}_t \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon & 0 \\ \rho_{\varepsilon v}\sigma_v & \sigma_v\sqrt{1 - \rho_{\varepsilon v}^2} \end{pmatrix}^{-1} \begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} = \widetilde{\mathbf{\Sigma}}^{-1} \begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix}$$

with $\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{\Sigma}}' = \mathbf{\Sigma}$. Rewriting the Wold representation in terms of the orthogonal innovations yields

$$\begin{pmatrix} y_t \\ h_t^{\frac{\delta}{2}} \end{pmatrix} = \begin{pmatrix} y^* \\ \mu_1 \end{pmatrix} + \begin{pmatrix} \psi_{y\varepsilon}(L) & \psi_{yv}(L) \\ \psi_{h\varepsilon}(L) & \psi_{hv}(L) \end{pmatrix} \begin{pmatrix} \sigma_\varepsilon & 0 \\ \rho_{\varepsilon v}\sigma_v & \sigma_v\sqrt{1 - \rho_{\varepsilon v}^2} \end{pmatrix} \begin{pmatrix} \tilde{\varepsilon}_t \\ \tilde{v}_t \end{pmatrix}.$$

Now, the variance of y_t can be decomposed into two parts:[12]

$$\mathbb{V}(y_t) = P_\infty(y_t|\tilde{\varepsilon}_t) + P_\infty(y_t|\tilde{v}_t), \tag{15}$$

[12] A similar decomposition of the variance of $h_t^{\delta/2}$ is provided in the Supplementary Information B.5.

  J. Time. Ser. Anal. **36:** 706–720 (2015)
DOI: 10.1111/jtsa.12119

where

$$P_\infty\left(y_t \mid \tilde{\varepsilon}_t\right) = \sigma_\varepsilon^2 P_\infty(y_t|\varepsilon_t) + \rho_{\varepsilon v}^2 \sigma_v^2 P_\infty(y_t|v_t) + 2\sigma_{\varepsilon v} P_\infty\left(y_t|\sqrt{\varepsilon_t v_t}\right) \tag{16}$$

and

$$P_\infty\left(y_t|\tilde{v}_t\right) = \sigma_v^2\left(1-\rho_{\varepsilon v}^2\right) P_\infty(y_t|v_t) \tag{17}$$

with individual components $P_\infty\left(y_t|\varepsilon_t\right) = \sum_{k=0}^{\infty}\left(\psi_{y\varepsilon}^{(k)}\right)^2$, $P_\infty(y_t|v_t) = \sum_{k=0}^{\infty}\left(\psi_{yv}^{(k)}\right)^2$ and $P_\infty\left(y_t|\sqrt{\varepsilon_t v_t}\right) = \sum_{k=0}^{\infty}\psi_{y\varepsilon}^{(k)}\psi_{yv}^{(k)}$.

Following Fiorentini and Sentana (1998), we define the persistence of a *standardized mean shock* as $P_\infty\left(y_t|\tilde{\varepsilon}_t\right)$, that is, the part of the variance of y_t that is due to $\tilde{\varepsilon}_t$ innovations, and the persistence of a *standardized volatility shock* as $P_\infty\left(y_t|\tilde{v}_t\right)$, that is, the part of the variance of y_t that is due to $\tilde{v}_t$ innovations. Note that if there is no asymmetry, that is, $\varsigma = 0$, then $\sigma_{\varepsilon v} = 0$ and the two persistence measures $P_\infty\left(y_t|\tilde{\varepsilon}_t\right)$ and $P_\infty\left(y_t|\tilde{v}_t\right)$ are equal to the persistence measures with respect to the original shocks ε_t and v_t scaled by the corresponding variances. In the following proposition, we derive the expressions for the individual components of the two persistence measures.

Proposition 4. If $0 < \boldsymbol{\Sigma} < \infty$, then $P_\infty\left(y_t|\tilde{\varepsilon}_t\right)$ and $P_\infty\left(y_t|\tilde{v}_t\right)$ are given by equations (16) and (17) respectively where

$$P_\infty(y_t|\varepsilon_t) = 1 + \frac{1}{(\lambda_1-\lambda_2)^2}\left[\frac{\lambda_1^2(\lambda_1-c)^2}{1-\lambda_1^2} + \frac{\lambda_2^2(\lambda_2-c)^2}{1-\lambda_2^2} - \frac{2\lambda_1\lambda_2(\lambda_1-c)(\lambda_2-c)}{1-\lambda_1\lambda_2}\right], \tag{18}$$

$$P_\infty(y_t|v_t) = \frac{(\vartheta\alpha)^2\left(1+\lambda_1\lambda_2\right)}{\left(1-\lambda_1^2\right)\left(1-\lambda_2^2\right)\left(1-\lambda_1\lambda_2\right)}, \tag{19}$$

$$P_\infty\left(y_t|\sqrt{\varepsilon_t v_t}\right) = \frac{\vartheta\alpha\left[\lambda_2 - c + \lambda_1(1-c\lambda_2)\right]}{\left(1-\lambda_1^2\right)\left(1-\lambda_2^2\right)\left(1-\lambda_1\lambda_2\right)}. \tag{20}$$

An interesting case is again the situation in which we have no level effects.

Lemma 2. If $\gamma = 0$, then $\lambda_1 = \phi, \lambda_2 = c$ and $P_\infty(y_t|\varepsilon_t) = 1/(1-\phi^2)$,

$$P_\infty(y_t|v_t) = \frac{(\vartheta\alpha)^2(1+\phi c)}{(1-\phi^2)(1-c^2)(1-\phi c)} \quad \text{and} \quad P_\infty\left(y_t|\sqrt{\varepsilon_t v_t}\right) = \frac{\vartheta\alpha\phi}{(1-\phi c)(1-\phi^2)}.$$

Lemma 2 shows that without a level effect, the expression for $P_\infty(y_t|\varepsilon_t)$ is the one for a simple AR(1) process (see Fiorentini and Sentana, 1998). Unsurprisingly, if there is no in-mean effect, that is, $\vartheta = 0$, then $P_\infty(y_t|v_t) = P_\infty\left(y_t|\sqrt{\varepsilon_t v_t}\right) = 0$. That is, $P_\infty\left(y_t|\tilde{\varepsilon}_t\right) = \sigma_\varepsilon^2 P_\infty\left(y_t|\varepsilon_t\right)$. On the other hand, if there is no intrinsic persistence in y_t, that is, $\phi = 0$, then $P_\infty\left(y_t|\varepsilon_t\right) = 1$; that is, the persistence measure takes its minimum value. Further, $\phi = 0$ also implies that $P_\infty(y_t|v_t) = \vartheta^2\alpha^2/(1-c^2) = \vartheta^2 P_\infty\left(h_t^{\delta/2}|v_t\right)$; that is, the persistence of a v_t shock on y_t is the same as on $h_t^{\delta/2}$ scaled by the squared in-mean term (see the Supplementary Information B.5). Finally, the restriction $\phi = 0$ implies $P_\infty\left(y_t|\sqrt{\varepsilon_t v_t}\right) = 0$.

In the Supplementary Information B.6, we derive an expression for the n-step ahead predictor $\mathbb{E}\left(y_{t+n}|\mathcal{F}_t\right)$ of y_{t+n}, given the information available at time t.[13] In particular, we show that the variance of the forecast error is given by

$$\mathbb{V}\left[\mathbb{FE}\left(y_{t+n}|\mathcal{F}_t\right)\right] = P_n\left(y_t|\tilde{\varepsilon}_t\right) + P_n\left(y_t|\tilde{v}_t\right),$$

where $P_n\left(y_t|\tilde{\varepsilon}_t\right)$ and $P_n\left(y_t|\tilde{v}_t\right)$ are defined analogously to $P_\infty\left(y_t|\tilde{\varepsilon}_t\right)$ and $P_\infty\left(y_t|\tilde{v}_t\right)$. Clearly, for covariance stationary processes, $\lim_{n\to\infty}\mathbb{V}\left[\mathbb{FE}\left(y_{t+n}|\mathcal{F}_t\right)\right] = P_\infty\left(y_t|\tilde{\varepsilon}_t\right) + P_\infty\left(y_t|\tilde{v}_t\right)$.

3. EMPIRICAL APPLICATIONS

3.1. S&P 500 Returns

First, suppose that y_t denotes the stock returns. Then, standard empirical results would suggest that y_t is basically white noise, while its conditional variance is highly persistent; that is, $\phi = 0$, and c is close to one. On the other hand, the risk-return trade-off implied by Merton's (1973) intertemporal capital asset pricing model (ICAPM) suggests that $\vartheta > 0$. For simplicity, we assume that $\gamma = 0$; that is, there is no level effect. In this case, equation (11) reduces to an ARMA(1,1) with c being the AR(1) parameter. Thus, under this parameter constellation, y_t can be only white noise if $c = \theta$. However, for $\delta = 2$ and $\vartheta \neq 0$, it is straightforward to show that the moving average parameter is always less than c. The two parameters are identical only in the trivial case in which there is no in-mean effect, that is, $\vartheta = 0$. This property of the model may explain why the in-mean parameter is typically estimated to be close to and not significantly different from zero (see e.g. French *et al.*, 1987). Thus, within the GARCH(1,1)-M model with $\delta = 2$, the possibility of having time-varying expected returns and at the same time uncorrelated returns is ruled out by construction. In sharp contrast, for $\delta = 1$, we can show that $\theta = c$ if and only if either $\vartheta = 0$ or $\vartheta = \varsigma(1-c^2)/\left(c\alpha\left(1+\varsigma^2-\frac{2}{\pi}\right)\right)$. That is, in the model with $\delta = 1$ and asymmetry, we can have time-varying expected returns and at the same time white noise returns.[14] Figure 3 plots c and θ with asymmetry term $\varsigma = 0.5$ and the remaining parameters fixed as described in the caption of the figure.

We estimate the AR(1)-APARCH(1,1)-M model for the daily continuously compounded returns on the S&P 500 during the 03 January 1975 to 20 August 2014 period. The data were obtained from *YahooFinance!*. In line with the aforementioned considerations, we choose $\delta = 1$. As panel A of Table III shows, the in-mean parameter is positive and significant, which confirms the risk-return trade-off suggested by the ICAPM. As expected, the asymmetry parameter is positive and highly significant. The estimated parameters imply a correlation of -0.74 between ε_t and v_t. In addition, since the implied c and θ are basically the same, the ARMA(2,1) reduces to an AR(1) with a small but highly significant ϕ parameter. Despite the presence of the in-mean effect, the variance of the returns is almost entirely explained by the mean shocks.[15]

3.2. US Inflation

In this example, we link our specification to the discussion on how to model US inflation. Stock and Watson (2007) find that an IMA(1,1) model is a good approximation for the quarterly US inflation rate [as measured by the GDP

[13] For a more general treatment of optimal predictors in ARMA models with GARCH-M effects, see Karanasos (2001).
[14] See the Supplementary Information B.1 for details.
[15] When we estimate the unrestricted model, the power term is equal to 1.335 (0.150), while the other parameters hardly change. In sharp contrast, when we impose the restriction $\delta = 2$, the in-mean parameter ϑ becomes insignificant (results not reported).

DOI: 10.1111/jtsa.12119

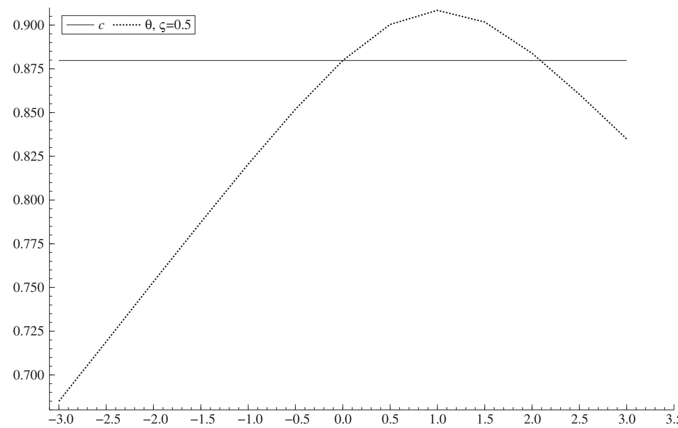

Figure 3. θ as a function of ϑ between -3.0 and 3.0. The other parameters are given by $\varphi = 0.05$, $\phi = 0.0$, $\omega = 0.1$, $\alpha = 0.1$, $\beta = 0.8$, $\gamma = 0.0$, $\delta = 1$ and $\varsigma = 0.5$

Table III. Parameter estimates for stock returns and inflation

Panel A: S&P 500 return data (03 January 1975 to 20 August 2014)							
φ	ϕ	ϑ	ω	α	ς	β	δ
-0.033 (0.030)	0.036 (0.010)	0.068 (0.037)	0.021 (0.004)	0.073 (0.012)	0.653 (0.071)	0.922 (0.011)	1 (–)
$\mathbb{C}orr(\varepsilon_t, v_t)$	c	θ		$P_\infty(y_t \mid \tilde{\varepsilon}_t)$		$P_\infty(y_t \mid \tilde{v}_t)$	
-0.735	0.981	0.984		1.333		0.0003	
Panel B: US inflation (1960/Q1–2014/Q2)							
φ	ϕ	ϑ	ω	α	ς	β	δ
0.250 (0.122)	0.779 (0.047)	0.384 (0.168)	0.029 (0.027)	0.094 (0.033)	–	0.884 (0.050)	2 (–)
$\mathbb{C}orr(\varepsilon_t, v_t)$	c	θ		$P_\infty(y_t \mid \tilde{\varepsilon}_t)$		$P_\infty(y_t \mid \tilde{v}_t)$	
0	0.978	0.924		3.338		3.193	

Notes: The numbers in parentheses are Bollerslev–Wooldridge robust standard errors.

(gross domestic product) deflator]. Since AR-GARCH-M models are often used to model the US inflation rate (see e.g. Grier and Perry, 2000), it is interesting to investigate how this specification links to the IMA(1,1). For illustrative purposes, consider again the case without level effect. In this situation, the inverse roots are given by $\lambda_1 = \phi$ and $\lambda_2 = c$. Suppose that y_t denotes the inflation rate. Given the empirical evidence on the persistence in the conditional variance of the inflation series, it is reasonable to assume that $c = 1$. According to equation (11), the reduced form process for y_t is then given by

$$(1 - \phi L)(1 - L)y_t = \varphi^* + (1 - \theta L)\eta_t,$$

that is, by an ARIMA(1,1,1). Clearly, if $\phi = 0$, the reduced form representation of the AR(1)-APARCH(1,1)-M specification coincides with the IMA(1,1) model proposed by Stock and Watson (2007).[16]

Further, Stock and Watson (2007) argue that the IMA(1,1) process is observationally equivalent to an unobserved components model with the transitory component being white noise and the permanent component a

[16] Alternatively, the IMA(1,1) model could be obtained by assuming that y_t denotes the log price level with $c = 1$ and $\phi = 1$.

DOI: 10.1111/jtsa.12119

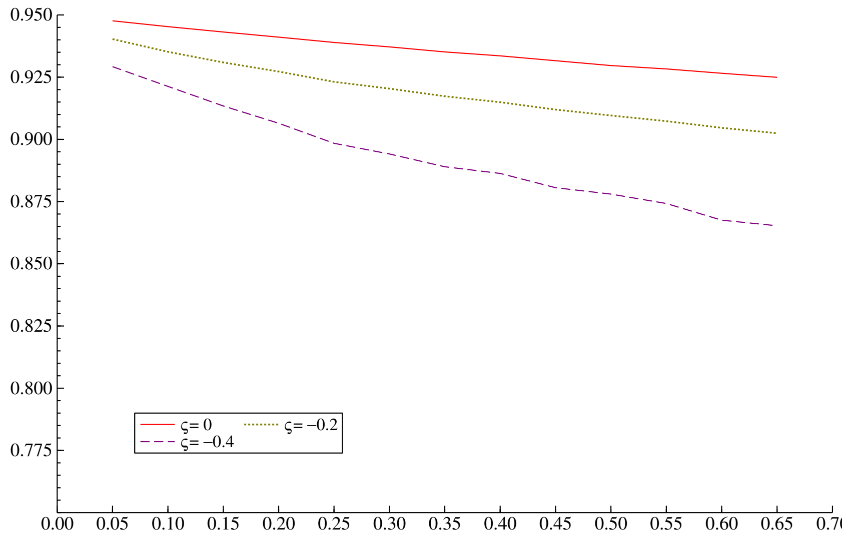

Figure 4. θ as a function of ω. The other parameters are given by $\varphi = 0.05$, $\phi = 0.0$, $\vartheta = 0.1$, $\alpha = 0.1$, $\gamma = 0$, $\delta = 2$, $\varsigma = 0$ (solid), $\varsigma = -0.2$ (dotted) and $\varsigma = -0.4$ (dashed). β is chosen such that $c = 0.95$

random walk with drift. The MA(1) parameter is then inversely related to the ratio of the variances of the innovations in the permanent and the transitory component. In particular, they argue that the MA(1) parameter has increased during the 'Great Moderation' because of a decrease in the variance of the permanent component during that period. Figure 4 shows θ as a function of ω and for different values of ς while holding the other parameters fixed as described in the caption. The figure clearly shows that in the APARCH-M specification, a decrease in the unconditional variance, that is, a decrease in ω, has the same effect as the one described by Stock and Watson (2007); that is, it will increase the MA(1) parameter.

Next, we apply our model to US inflation data. Seasonally adjusted quarterly GDP deflator data were obtained from the Federal Reserve Bank of St. Louis for the period 1960/Q1–2014/Q2. The inflation rate was calculated as $y_t = 400 \times [\ln(P_t) - \ln(P_{t-1})]$, where P_t denotes the price level in quarter t. Both the DF and DF-GLS tests reject the null hypothesis of a unit root in the inflation rate at the 1% significance level (results not reported). Panel B of Table III presents the estimation results. In line with the previous literature on the link between the level of inflation and its conditional variance, we set $\delta = 2$. As expected, the conditional variance is highly persistent with $c = 0.978$, that is, close to integrated behaviour.[17] In agreement with the so-called Friedman (1977) hypothesis, the in-mean parameter is positive and significant; that is, the higher the conditional variance, the higher the level of inflation. Since $\phi = 0.779$ and the implied MA(1) parameter is $\theta = 0.924$, our results suggest that the reduced form process is (close to) an ARIMA(1,1,1). Finally, the two persistence measures show that both mean and variance shocks contribute almost equally to the variation of the inflation rate.

4. CONCLUSIONS

We discuss the persistence properties of the AR(1)-APARCH(1,1)-ML model. This model allows for an in-mean effect as well as a level effect. Both effects are in line with economic theory, which, for example, suggests that inflation uncertainty should have an effect on the level of inflation and vice versa. Our main result is that the

[17] When using quarterly GDP deflator data, we do not find evidence for asymmetry in the conditional variance. However, such evidence can be found in monthly consumer price index (CPI) inflation series.

  J. Time. Ser. Anal. 36: 706–720 (2015)
DOI: 10.1111/jtsa.12119

commonly observed persistence in the mean/conditional variance of many economic times series may be at least partly a result of a transmission mechanism. If this mechanism is ignored, then conventional procedures for estimating the persistence in the mean/variance may lead to biased estimates. In particular, unit root tests might falsely indicate a unit root and, hence, suggest the modelling of the differenced series rather than the level series.

ACKNOWLEDGEMENTS

We thank Simone Elmer who was a co-author in an early version of this article. We are grateful to the Editor, Robert Taylor, a Guest Editor, and two anonymous referees for their comments that greatly improved our paper. We also thank Nikolaus Hautsch, Robinson Kruse, Dieter Nautz, Winfried Pohlmeier, Melanie Schienle, Timo Teräsvirta, Jürgen Wolters and the participants of the European meeting of the Econometric Society (Barcelona, 2009), the Economic Risk Seminar at the Humboldt University Berlin (2009), the annual meeting of the Society for Nonlinear Dynamics and Econometrics (Novara, 2010), the annual meeting of the Swiss Society of Economics and Statistics (Fribourg, 2010), the Forschungskolloquium at the University of Konstanz (2010) and the departmental seminar at the Free University Berlin (2011) for helpful comments and suggestions.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's website.

REFERENCES

Ball L. 1992. Why does high inflation raise inflation uncertainty? *Journal of Monetary Economics* **29**: 371–388.
Cavaliere G, Taylor RAM. 2007. Testing for unit roots in time series with non-stationary volatility. *Journal of Econometrics* **140**: 919–947.
Cavaliere G, Taylor RAM. 2008a. Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory* **24**: 43–71.
Cavaliere G, Taylor RAM. 2008b. Testing for a change in persistence in the presence of non-stationary volatility. *Journal of Econometrics* **147**: 84–98.
Christensen BJ, Nielsen MØ. 2007. The effect of long memory in volatility on stock market fluctuations. *Review of Economics and Statistics* **89**: 684–700.
Conrad C, Karanasos M. 2006. The impulse response function of the long memory GARCH process. *Economics Letters* **90**: 34–41.
Conrad C, Karanasos M. 2010. Negative volatility spillovers in the unretsricted ECCC-GARCH model. *Econometric Theory* **26**: 838–862.
Conrad C, Karanasos M. 2014. Modeling the link between US inflation and output: the importance of the uncertainty channel. *Scottish Journal of Political Economy,* forthcoming.
Conrad C, Karanasos M, Zeng N. 2010. The link between macroeconomic performance and variability in the UK. *Economics Letters* **106**: 154–157.
Cukierman A, Meltzer AH. 1986. A theory of ambiguity, credibility, and inflation under discretion and asymmetric information. *Econometrica* **54**: 1099–1128.
DeJong DN, Nankervis JC, Savin NE, Whiteman CH. 1992a. Integration versus trend-stationarity in macroeconomic time series. *Econometrica* **60**: 423–34.
DeJong DN, Nankervis JC, Savin NE, Whiteman CH. 1992b. The power problems of unit root test in time series with autoregressive errors. *Journal of Econometrics* **53**: 323–343.
Demetrescu M, Kruse R. 2013. The power of unit root tests against nonlinear local alternatives. *Journal of Time Series Analysis* **34**: 40–61.
Elliot G, Rothenberg TJ, Stock JH. 1996. Efficient tests for an autoregressive unit root. *Econometrica* **64**: 813–836.
Engle RF, Bollerslev T. 1986. Modelling the persistence of conditional variances. *Econometric Reviews* **5**: 1–50.
Engle RF, Lilien DM, Robins RP. 1987. Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica* **55**: 391–407.

DOI: 10.1111/jtsa.12119

Fiorentini G, Sentana E. 1998. Conditional means of time series processes and time series processes for conditional means. *International Economic Review* **39**: 1101–1118.
French KR, Schwert GW, Stambaugh RF. 1987. Expected stock returns and volatility. *Journal of Financial Economics* **19**: 3–29.
Friedman M. 1977. Nobel lecture: inflation and unemployment. *Journal of Political Economy* **85**: 451–472.
Grier KB, Perry M. 2000. The effects of real and nominal uncertainty on inflation and output growth: some GARCH-M evidence. *Journal of Applied Econometrics* **15**: 45–58.
Grier KB, Henry OT, Olekalns N, Shields K. 2004. The asymmetric effects of uncertainty on inflation and output growth. *Journal of Applied Econometrics* **19**: 551–565.
Haldrup N, Jansson M. 2006. *Improving size and power in the unit root testing*. In *Palgrave Handbook of Econometrics*, Mills TC, Patterson K (eds.), Vol. 1, pp. 252–277.
Holland S. 1995. Inflation and uncertainty: tests for temporal ordering. *Journal of Money, Credit, and Banking* **27**: 827–837.
Karanasos M. 2001. Prediction in ARMA models with GARCH in mean effects. *Journal of Time Series Analysis* **22**: 555–576.
Karanasos M. 2007. The covariance structure of some financial time series models. *Quantitative and Qualitative Analysis in Social Sciences* **1**: 71–87.
Karanasos M, Kim J. 2006. A re-examination of the asymmetric power ARCH model. *Journal of Empirical Finance* **13**: 113–128.
Karanasos M, Zeng N. 2013. *Conditional heteroskedasticity in macroeconomics data: UK inflation, output growth and their uncertainties*. In *Handbook of Research Methods and Applications in Empirical Macroeconomics*, Hashimzade N, Thornton MA (eds.) Edward Elgar Publishing, Inc.
Kim K, Schmidt P. 1993. Unit root tests with conditional heteroskedasticity. *Journal of Econometrics* **59**: 287–300.
Kourogenis N, Pittis N. 2008. Testing for a unit root under errors with just barely infinite variance. *Journal of Time Series Analysis* **29**: 1066–1087.
Ling S, Li WK. 2003. Asymptotic inference for unit root processes with GARCH(1,1) errors. *Econometric Theory* **19**: 541–564.
Ling S, McAleer S, Li WK. 2003. Estimation and testing for unit root processes with GARCH(1,1) errors: theory and Monte Carlo evidence. *Econometric Reviews* **22**: 179–202.
Merton RC. 1973. An intertemporal capital asset pricing model. *Econometrica* **41**: 867–886.
Pivetta F, Reis R. 2007. The persistence of inflation in the United States. *Journal of Economic Dynamics and Control* **31**: 1326–1358.
Rahbek A, Nielsen HB. 2014. Unit root vector autoregression with volatility induced stationarity. *Journal of Empirical Finance* **29**: 144–167.
Rodrigues PM, Rubia A. 2005. The performance of unit root tests under level-dependent heteroskedasticity. *Economics Letters* **89**: 262–268.
Seo B. 1999. Distribution theory for unit root tests with conditional heteroscedasticity. *Journal of Econometrics* **91**: 113–144.
Stock JH, Watson MW. 2007. Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking* **39**: 3–33.

APPENDIX A

A1. Proofs

Proof of Proposition 1

Multiplying equation (1) by $(1-cL)$ and substituting (3) into equation (1) give equation (8). Similarly, multiplying equation (3) by $(1-\phi L)$ and substituting (1) into equation (3) give equation (9). □

Proof of Proposition 2

Taking expectations from both sides of equation (9) yields equation (10). □

Proof of Proposition 3

On account of equations (8) and (9) and equation (A.1) in Karanasos (2007), we obtain equation (14) by straightforward manipulation. □

Proof of Proposition 4

The desired result is obtained straightforwardly from Proposition 3. □

DOI: 10.1111/jtsa.12119