On the Transmission of Memory in GARCH-in-mean models

Christian Conrad†, and Menelaos Karanasos‡
†University of Heidelberg, Heidelberg, Germany
‡Brunel University, West London, UK

November 2013

Abstract

In this paper 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)-GARCH(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)-GARCH(1,1)-in-mean-level process will be observationally equivalent to an ARMA(2,1) process with the largest autoregressive root being close to one. In particular, unit root tests for the level will not be able to reject the null hypothesis of the process being integrated of order one. We argue that the commonly observed decrease in U.S. inflation persistence and inflation variability can be well explained by our model in combination with a change in monetary policy in the early 1980’s.

Keywords: Conditional heteroscedasticity, GARCH-in-Mean, persistence, unit root tests.

JEL Classification: E31, E58, C12, C22, C52.

*We would like to thank Simone Elmer who was a co-author in an early version of this paper. We would also like to thank Nikolaus Hautsch, 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), 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.

†Christian Conrad, University of Heidelberg, Alfred-Weber-Institute, Grabengasse 14, 69117 Heidelberg, Germany; email: christian.conrad@awi.uni-heidelberg.de, tel: +49 (0)6221 54 3173, fax: +49 (0)6221 54 3578.

‡Corresponding author: Menelaos Karanasos, Economics and Finance, Brunel University, Uxbridge, West London, UB3 3PH, UK; email: menelaos.karanasos@brunel.ac.uk, tel: +44 (0)1895 265 284, fax: +44 (0)1895 269 770.
1 Introduction

Many economic time series are characterized by an autocorrelation structure which makes it difficult to classify the series as being either stationary $I(0)$ or non-stationary $I(1)$. A primary example for such a series are inflation rates. Conventional wisdom then suggests to 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 modeling, hypothesis testing, forecasting and the like. As pointed out by Haldrup and Jansson (2005), 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 typically GARCH effects with highly persistent volatility are found. 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 analyzed and treated independently. For example, standard unit root tests are based either on the assumption that the variance of the series is constant or that some type of heteroscedasticity is at place (e.g. Kim and Schmidt, 1993; Seo, 1999; Ling and Li, 2003; Ling et al., 2003; Kourogenis and Pittis, 2008) but ignore the possibility that the volatility has a direct effect on the level.

In this paper we consider an AR-GARCH-in-mean-level (AR-GARCH-ML) process, i.e., a model in which the conditional variance affects the level of the dependent variable and vice versa. This model has been introduced by Engle et al. (1987) and applied in, e.g. Grier et al. (2004), Conrad et al. (2010), Conrad and Karanasos (2010, 2013) and Karana-
sos 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 highest autoregressive root of the AR part will, under reasonable assumptions, be close to and statistically indistinguishably 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. Using a Monte Carlo study, we show that in the presence of volatility spillovers conventional unit root tests fail to reject the null that the underlying process is I(1). In the presence of volatility spillovers, the persistence of the conditional variance is transmitted to the level of the process and procedures which do not distinguish between the different degrees of persistence in the two moments tend to overestimate the persistence either in the mean or variance. Rodrigues and Rubia (2005) examine the performance of unit root tests under level-dependent heteroscedasticity (for the important research issue of testing for unit roots in the presence of permanent volatility shifts see, for example, Cavaliere and Taylor, 2007, 2008a). We also illustrate this important point both by deriving the autocorrelation function, the impulse response function and the optimal predictor for the level process (for prediction in ARMA models with GARCH-in-mean effects see also Karanasos, 2001). Recently Rahbek and Nielsen (2012) formulate a multivariate AR model which on the one hand allows for unit-roots in the conditional mean, while at the same time, allows levels to appear in the multivariate conditional ARCH part, which, possibly, induce stationarity.

An empirical application to U.S. inflation data shows that the model accurately explains the changes in inflation persistence in the mid-1980’s and the accompanying decrease in volatility which is commonly referred to as the “Great Moderation”. Cavaliere and Taylor (2008b) develop tests for a change in persistence in the presence of non-stationary volatility and illustrate the empirical relevance of these tests using US inflation data. Our main result is the observation that the persistence in the level and the conditional variance of inflation are directly linked and a meaningful analysis has to take into
account both properties jointly. Our findings concur with Stock and Watson (2007) who suggest that the U.S. rate of price inflation can reasonably well be approximated by an IMA(1, 1) process with a change in both the MA(1) parameter and the variance of the error term in the mid-1980’s (Cavaliere and Taylor, 2007, illustrate the empirical relevance of their unit root testing techniques using producer price inflation series from the Stock-Watson database). However, we show that the increase in the MA(1) parameter and the decrease in the error variance are ultimately linked and driven by the same sources: a decrease in inflation persistence and a change in the sign of the effect of inflation uncertainty on the level of inflation. Both can be attributed to a change in U.S. monetary policy in the post 1984 era and is line with the predictions made in Clarida and Waldman (2008).

The outline of the paper is as follows. Section 2 presents the model and its properties, including the autocorrelation structure, measures of persistence and optimal predictors. In Section 3 the model is applied to U.S. inflation data. Finally, Section 4 concludes. All proofs are deferred to the appendix.

2 The Model

The AR(1)-GARCH(1, 1)-ML model is given by

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

with

\[\varepsilon_t = e_t h_t^{\frac{1}{2}}\]

where \(\delta > 0\), \(\{e_t\}\) is a sequence of independent, identically distributed random variables with zero mean and finite variance, \(\mathbb{E}(e_t^2)\), and \(h_t\) is the conditional variance of \(y_t\) (see, among others, Christensen and Nielsen, 2007). The AR(1) parameter \(\phi\) naturally
measures the intrinsic persistence in the level of \( y_t \). The power transformed conditional variance, \( h_t^{\delta/2} \), 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 APARCH(1,1)-L process:

\[
(1 - \beta L)h_t^{\delta/2} = \omega + \alpha f(\varepsilon_{t-1}) + \gamma y_{t-1}
\]  

(2)

with

\[
f(\varepsilon_{t-1}) = [\varepsilon_{t-1}]^{\delta} - \varepsilon_{t-1}^{\delta}
\]

where \(|\varsigma| < 1\) (for the power ARCH model see, for example, Karanasos and Kim, 2006). By including the lagged \( y_t \) in the conditional variance equation and \( h_t^{\delta/2} \) in the mean equation, 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 \( t \). Hereafter, we will denote \( \mu_r = \mathbb{E}(h_t^r) \). Notice that when \( \delta \neq 0.5, 1, 2 \) then both \( \mu_{2/\delta} \) and \( \mu_{1+1/\delta} \) are fractional moments and have to be calculated numerically. In addition, \( \mu_1 \) and \( \mu_2 \) are given below (see Proposition 2 below and equation (28) in the Appendix).

The APARCH(1,1)-L formulation in equation (2) can readily be interpreted as an ARMA(1,1)-L process for the conditional variance:

\[
(1 - cL)h_t^{\delta/2} = \omega + \alpha v_{t-1} + \gamma y_{t-1}
\]  

(3)

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

Note, that the parameter \( c \) measures the 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).
Notice also that $\mathbb{E}(\varepsilon_t^2) = \sigma_\varepsilon^2 = \mu_{2/\delta}\mathbb{E}(e_t^2)$ and $\mathbb{E}(\varepsilon_t v_t) = \sigma_{\varepsilon v} = \mu_{1+1/\delta}\overline{\kappa}$ with $\overline{\kappa} = \mathbb{E}[e_t f(e_t)]$. In other words, the covariance matrix of the two ‘shocks’ $\varepsilon_t$ and $v_t$ is given by

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

(4)

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

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

(5)

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

(6)

It is now interesting to investigate some special cases (still under the assumption that $e_t$ is standard normal). First, we are interested in the behavior of $\kappa^{(1)}$ when $\delta = 2$ or $\delta = 1$. It directly follows from equation (5) that

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

This 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, $\varepsilon_t$ and $v_t$ will be uncorrelated for all $\delta > 0$. Hence, for $\delta = 2$ and $\varsigma = (\gamma =) 0$ the variance covariance matrix $\Sigma$ reduces to

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

where $\mu_1$ is given in Proposition 2 below and $\mu_2$ by equation (30) in the Appendix. On
the other hand, for $\delta = 1$ equation (6) reduces to $\bar{\kappa} = -\varsigma$ which implies that $\Sigma$ becomes

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

Figure 1 shows the covariance between $\varepsilon_t$ and $v_t$ for $\varsigma$ 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 $\varepsilon_t$ and $v_t$ are negatively (positively) correlated for positive (negative) values of $\varsigma$. For example, in empirical applications using stock return data, we would expect a positive $\varsigma$ due to the leverage effect and, hence, negative return innovations coincide with positive volatility innovations. On the contrary, if we think about applications to inflation data, we would expect a negative $\varsigma$ because higher than expected inflation rates typically lead to more inflation variability.

Next, we derive the univariate representations of $y_t$ and $h_t^{\delta/2}$. That is, we will write $y_t$ as an ARMA process that does not longer depend on $h_t^{\delta/2}$. Similarly, we obtain an ARMA representation for $h_t^{\delta/2}$ which does not involve $y_{t-1}$.

**Proposition 1** The univariate ARMA$(2,1)$ representation of the process $y_t$ and the corresponding ARMA$(2,2)$ representation of the power transformed conditional variance $h_t^{\delta/2}$ are given by

$$(1 - a_1 L - a_2 L^2)y_t = \phi^* + (1 - c)\varepsilon_t + \vartheta \alpha v_{t-1}, \quad (8)$$

$$(1 - a_1 L - a_2 L^2)h_t^{\delta/2} = \omega^* + \gamma \varepsilon_{t-1} + (1 - \phi)\alpha v_{t-1} \quad (9)$$

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

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

**Assumption A1 (Stationarity)** The inverse roots \( \lambda_1 \) and \( \lambda_2 \) of \( (1 - a_1 L - a_2 L^2) \) 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. Note that it also implies that \( a_1 + a_2 < 1 \). The representation in equation (8) illustrates that measures of persistence in the level/conditional variance which are based on \( a_1 \) and \( a_2 \) will confuse intrinsic persistence with persistence that is due to the in-mean/level transmission channel.

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

\[
\mu_1 = \frac{\omega^*}{1 - a_1 - a_2}.
\]

(10)

For \( \delta = 2 \) and \( \gamma = \varsigma = 0 \) the result in Proposition 2 reduces to the well known formula for the GARCH(1,1) process: \( \mu_1 = \mathbb{E}(h_t) = \mathbb{E}(\varepsilon_t^2) = \omega/(1 - \alpha - \beta) \). Further, note that the existence of \( \mu_1 \) guarantees that of \( \mu_{2/\delta} \) only if \( \delta \geq 2 \). Similarly the existence of the second moment \( \mu_2 \) (see below) 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 \( \varepsilon_t \) and the conditional variance shocks \( v_t \). As discussed above, the two shocks will be uncorrelated if \( \varsigma = 0 \). Since the sum of a MA(1) and a white noise process is again a MA(1), equation (8) can be rewritten as

\[
(1 - a_1 L - a_2 L^2)y_t = \varphi^* + (1 - \theta L)\eta_t
\]

(11)

where \( \eta_t \) is an uncorrelated error process with mean zero and variance \( \sigma_\eta^2 \). The parameters
and $\sigma^2_\eta$ can be expressed as

$$
\theta = \frac{-\sigma_0 \pm \sqrt{\sigma_0^2 - 4\sigma_1^2}}{2\sigma_1},
$$

$$
\sigma^2_\eta = \frac{\sigma_1}{-\theta},
$$

where $\sigma_0 = (1 + c^2)\sigma^2_\varepsilon + (\vartheta\alpha)^2\sigma^2_\varepsilon - 2c\vartheta\alpha\sigma_\varepsilon\varepsilon$, and $\sigma_1 = -c\sigma^2_\varepsilon + \vartheta\alpha\sigma_\varepsilon\varepsilon$. Note that (i) $\theta$ is real if and only if $\sigma_0^2 > 4\sigma_1^2$, and (ii) $0 \leq \theta$ if $\sigma_1 \leq 0$, that is if $\vartheta\alpha\sigma_\varepsilon\varepsilon \leq c\sigma^2_\varepsilon$.

We now focus on the situation in which $\gamma = 0$, i.e. there is no level effect. In this case the two inverse roots are given by $\lambda_1 = \phi$ and $\lambda_2 = c$. That is, the inverse roots are equal to the two measures of intrinsic persistence in the level and conditional variance.

We are interested in the question whether in the presence of an in-mean effect, i.e. $\vartheta \neq 0$, there are parameter combinations for which $c = \theta$ and, hence, $y_t$ reduces to an AR(1). For $\delta = 2$, we can show that the moving average parameter is always less or equal to $c$.

The two parameters are identical only in the trivial case in which there is no in-mean effect, i.e. $\vartheta = 0$. The situation is different for $\delta = 1$. In this case $\sigma_0$ and $\sigma_1$ become:

$$
\sigma_0 = \mu_2[1 + c^2 + (\vartheta\alpha)^2(1 + \zeta^2 - \frac{2}{\pi}) + 2c\vartheta\alpha\zeta]\quad \text{and} \quad \sigma_1 = -\mu_2(c + \vartheta\alpha\zeta)\quad \text{respectively.}
$$

It follows that $\theta = c$ if and only if either $\vartheta = 0$ or $\vartheta = (\zeta(1 - c^2))/(c\alpha(1 + \zeta^2 - \frac{2}{\pi}))$, and the ARMA(2,1) representation reduces to an AR(1) process. This implies that when $c = 1$, that is there is a unit root in the conditional standard deviation, then $\theta = c = 1$ only if there are no in-mean effects: $\vartheta = 0$. Moreover, $\theta < c$, if and only if either $\vartheta\zeta < 0$, or $|\vartheta| > \frac{|\vartheta|(1-c^2)}{c\alpha(1+c^2-\frac{2}{\pi})}$. In this case as i) $|\vartheta|$ increases, $\theta$ decreases, ii) $\alpha$ increases $c - \theta$ increases and iii) $\zeta$ increases $\theta$ decreases if and only if either $\vartheta < 0$, $\zeta > 0$ or $\vartheta > 0$ and $\zeta > (1 - c^2)/(2\vartheta\alpha\zeta)$ (see Appendix A.1). Note also, that when $\delta = 1$ then $\theta$ does not depend on $\omega$. Figure 2 shows $\theta$ as a function of $\alpha$, $\delta = 1, 2$ and for different values of the asymmetry parameter $\zeta$ (for combinations of $\alpha$ and $\zeta$ for which $\mu_1$ and $\mu_2$ exist). In the lower panel of Figure 2 the parameter $c$ is kept constant by adjusting $\beta$ when changing $\alpha$. Clearly, the MA(1) parameter $\theta$ increases as $\alpha$ decrease Figure 3 shows that $\theta$ is an
increasing in $\gamma$.

Example 1 (Application to US inflation)

In this example we link our specification to the discussion on potential changes in inflation persistence. For illustrative purposes consider the case that $\gamma = 0$, i.e. there is no level effect. In this situation, the inverse roots are given by $\lambda_1 = \phi$ and $\lambda_2 = c$. Suppose that $y_t$ is the log of the price level. Then, we choose $\phi = 1$ and $\pi_t = (1 - L)y_t$ denotes the inflation rate. According to equation (11), the reduced form process for $\pi_t$ is then given by

$$(1 - cL)\pi_t = \varphi^* + (1 - \theta L)\eta_t$$

i.e. by an ARMA(1, 1). In many empirical applications of GARCH models to inflation the parameter $c$ is found to be close to and statistically indistinguishable from one (see, e.g., Engle and Bollerslev, 1986). Imposing this restriction the model further reduces to

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

which is an IMA(1, 1). Interestingly, the IMA(1, 1) model has been recently suggested by Stock and Watson (2007) as a good approximation for the quarterly US inflation rate. Stock and Watson (2007) consider the period 1960:Q1–2004:Q4 and the two subperiods 1960:Q1–1983:Q4 and 1984:Q1–2004:Q4. They find evidence for a unit root in both subperiods and, hence, argue that in this respect inflation persistence has not changed. On the other hand, they find that the MA(1) parameter $\theta$ is considerably higher in the second subperiod than in the first subperiod. To rationalize this result, 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 random walk with drift. The MA(1) parameter is then inversely related to the ratio of the variance of the innovations in the permanent and the transitory component. According to their estimates, the variance of the transitory component is basically identical in both
subperiods, while the variance of the permanent component decreased significantly from the first to the second subperiod. Thus, the decrease in the variance of the innovation in the permanent component – which may be due to the Great Moderation – explains the observed increase in the MA(1) parameter. Although the unobserved components model is quite different from our AR-GARCH-in-mean specification both models lead to the same reduced form representation of the inflation process. Because the change in the MA(1) parameter between the two subperiods is driven by a change in the variance, we can check whether this stylized fact of the data can also be replicated by our model. For this we need to change the unconditional variance of the process without affecting the parameter $c$. The easiest way to do this is by changing the constant $\omega$ in the conditional variance equation. Figure 4 plots $\theta$ as a function of $\omega$ and for the different values of $\zeta$ while holding the other parameters fixed. The figure shows that an increase in the unconditional variance in the in-mean specification has the same effect as the one described in Stock and Watson (2007), i.e. it will decrease the MA(1) parameter.

**Example 2 (Application to risk-return relationship)**

French et al. (1987) use a GARCH-in-mean specification for testing Merton’s (1973) intertemporal CAPM (ICAPM). They investigate relations of the form

$$\mathbb{E}[r_{M,t} - r_{f,t} | \mathcal{F}_{t-1}] = \vartheta(V[r_{M,t} - r_{f,t} | \mathcal{F}_{t-1}])^{1/2}$$

where $r_{M,t}$ is the market return, $r_{f,t}$ the risk free rate and, according to the ICAPM, $\vartheta$ is the representative agent’s coefficient of relative risk aversion.

Defining $y_t = r_{M,t} - r_{f,t}$ the French et al. (1987) model is nested in our specification under the restrictions that $\phi = 0$, $\gamma = 0$ and $\zeta = 0$.\(^1\) For this case the model can be

\(^1\)More precisely, French et al. (1987) estimate a GARCH(1,2) specification for the conditional variance.
written as the bivariate VAR

$$\begin{pmatrix} y_t \\ h_t^{\delta/2} \end{pmatrix} = \begin{pmatrix} 0 & \vartheta \\ 0 & c \end{pmatrix} \begin{pmatrix} y_{t-1} \\ h_{t-1}^{\delta/2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \alpha v_t \end{pmatrix}. \quad (12)$$

Hence, combining equations (11) and (12) the univariate representations of the excess returns and their conditional expectation are given by

$$(1 - cL) y_t = (1 - \theta L) \eta_t,$$

$$(1 - cL) \mathbb{E}[y_t | \mathcal{F}_{t-1}] = \vartheta \alpha v_{t-1}.$$ 

That is, $y_t$ follows an ARMA(1,1) process while its conditional expectation follows an AR(1) process. Empirically, it is again clear that the conditional variance will be highly persistent, i.e. $c$ is close to one. This is in line with observation that expected returns are highly persistent. On the other hand, we would like to replicate the fact that the observed excess returns $y_t$ are almost uncorrelated. In the above specification $y_t$ will be white noise only if $c = \theta$.

First, consider the case that $\delta = 2$. In Figure 5 we plot $\theta$ as a function of the in-mean parameter $\vartheta$ (the other parameters are fixed as described in the caption of Figure 5). Note that $\theta$ decreases symmetrically as $|\vartheta|$ increases. The figure also shows the corresponding value of $c$ which does not depend on $\vartheta$. Clearly, $c = \theta$ if and only if $\vartheta = 0$. That is, the case in which $y_t$ behaves like white noise requires that there is no in-mean effect, i.e. the conditional expectation is constant. These considerations may explain why in empirical applications such as in French et al. (1987) the in-mean parameter is typically estimated to be close to and not significantly different from zero. Within the GARCH-in-mean (with parameters as above) the possibility to have time-varying expected returns and at the same time uncorrelated ex-post returns is ruled out by assumption. Figure 6 shows again a plot of $\theta$ as a function of $\vartheta$ but now allows for asymmetry $\varsigma \in \{-0.5, 0.5\}$ (while still $\delta = 2$).
In this case $\theta$ still decreases in $|\vartheta|$ but no longer symmetrically. Note that the asymmetry parameter affects the value of the persistence because we have $c = \alpha(1+\zeta^2) + \beta$. However, our main conclusion that the combination of $y_t$ being white noise and its conditional expectation persistent but time varying is still ruled out.

Finally, consider the case that $\delta = 1$. If there is no asymmetry, i.e. $\zeta = 0$, we can show that again $\theta \leq c$ for all values of $\vartheta$ (for details, see Appendix A.1). As before, $\theta = c$ requires that $\vartheta = 0$. Figure 7 plots $c$ and $\theta$ with asymmetry term $\zeta = 0.5$ and the remaining parameters fixed at the same values as before. Now, $\theta = c$ not only for $\vartheta = 0$, but there exists a second value $\vartheta > 0$ for which $\theta = c$. That is, in the model with $\delta = 1$ and asymmetry we can have time-varying expected returns and at the same time white noise ex-post returns.

### 2.1 Covariance Structure

In this subsection we derive the autocovariance structure of $y_t$, $\text{Cov}_k(y_t)$, $k \in \mathbb{N}$, in the presence of the in-mean term $\vartheta \neq 0$. We are particularly interested in the transmission of memory from the variance to the level of the series. Special attention will be given to the role played by the asymmetry term, $\zeta$, and the power transformation, $\delta$. The following lemma provides the (co)variance structure of $y_t$ for the case that $\gamma = 0$, i.e. the are no level effects. The general result including the level effect can be found in Proposition 6 in Appendix A.2.

**Lemma 1** Let $\gamma = 0$, then $\lambda_1 = \phi$, $\lambda_2 = c$ and we have

\begin{align}
\mathbb{V}(y_t) &= \frac{1}{1 - \phi^2} \left\{ \sigma_z^2 + \frac{\vartheta \alpha}{1 - \phi} \left[ \frac{\sigma_z^2 \vartheta \alpha (1 + \phi c)}{1 - \phi^2} + 2\sigma_v \phi \right] \right\} , \\
\text{Cov}_k(y_t) &= \frac{\phi^k}{1 - \phi^2} \sigma_z^2 + \vartheta^2 \alpha^2 \lambda (k) \sigma_v^2 \\
&\quad + \frac{\vartheta \sigma_v \alpha}{(\phi - c)(1 - \phi c)} \left[ \frac{\phi^k (1 + \phi^2 - 2\phi c)}{1 - \phi^2} - \phi^k \right] \quad \text{for } k \geq 1
\end{align}
where \( \lambda^{(k)} \) is defined in Appendix A.2.

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, i.e. \( \vartheta = 0 \). For \( \vartheta \neq 0 \) both are affected because memory is transmitted from the conditional variance to the level of the series. In particular, it interesting to discuss the properties of the covariance function. Equation (14) shows that \( \text{Cov}_k(y_t) \) 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 \( h_t^{k/2} \) itself is correlated. If \( \phi = 0 \), i.e. \( y_t \) has no intrinsic persistence, then \( \lambda^{(k)} = \phi^k/(1 - \phi^2) \) and, hence, the second term is equal to \( \text{Cov}_k(h_t^{k/2}) \) scaled by the squared in-mean term (see Appendix A.2). Finally, even in the presence of an in-mean effect, the third term will be nonzero only if \( \varepsilon_t \) and \( v_t \) are correlated, that is \( \varsigma \neq 0 \).

Denote \( \phi^{(k)} = \frac{1}{(\phi - \kappa)(1 - \phi^2)} \left[ \frac{\phi^k(1 + \phi^2 - 2\phi\kappa)}{1 - \phi^2} - \phi^k \right] > 0 \). Then \( \text{Cov}_k(y_t) > \sigma_\varepsilon^2 \phi^{(k)} \) if either \( \vartheta < 0 \), or \( |\vartheta| > \frac{|\alpha \lambda^{(k)}| \phi^{(k)}}{\sigma_v^2} \). If this is the case then as \( |\vartheta| \) increases, \( \text{Cov}_k(y_t) \) increases. When \( \vartheta \) is either 0 or \( \frac{\sigma_v^2 \phi^{(k)}}{\alpha \lambda^{(k)} \sigma_v^2} \) then \( \text{Cov}_k(y_t) = \frac{\phi^k}{1 - \phi^2} \sigma_\varepsilon^2 \). In general, if and only if \( \vartheta > \frac{\sigma_v^2 \phi^{(k)}}{2\alpha \lambda^{(k)} \sigma_v^2} \) then as \( \vartheta \) increases, \( \text{Cov}_k(y_t) \) increases.

In the following we will graphically illustrate the consequences of the in-mean term for the dependence structure in \( y_t \). Figure 8 shows the autocorrelation functions of \( y_t \) with \( \phi = 0.1 \), i.e. a low degree of intrinsic level persistence, and \( \vartheta \) 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 \( \delta \) is assumed to be either 2 (left Panel) or 1 (right Panel) and there is no asymmetry, \( \varsigma = 0 \). The figure clearly shows that increasing the value of \( \vartheta \) increases the correlation of \( y_t \). When the variance is almost integrated \( (c = \alpha + \beta = 0.99) \) and \( \delta = 2 \) the autocorrelation function of \( y_t \) behaves like the one of a non-stationary process. Since there is almost no intrinsic persistence in the level of \( y_t \), this behavior is due to the transmission of memory from the conditional variance to the level. Figures 9 and 10 illustrate the behavior of the ACF when \( y_t \) itself has moderate (\( \phi = 0.5 \)) and high (\( \phi = 0.9 \)) intrinsic persistence. The combination of higher
values of \( \phi \), nonzero in-mean effects and a high \( c \) clearly leads to almost linearly decaying autocorrelation functions. Figures 11-13 show the corresponding autocorrelation functions of \( y_t \) for \( \beta = 0.80 \), \( \beta = 0.85 \) and \( \beta = 0.89 \) while \( \alpha = 0.1 \). Thus, any applied researcher who investigates the autocorrelation function of \( y_t \) might come to the conclusion that the process is integrated and, hence, would model the first difference of \( y_t \). Later on, we will discuss implications of these findings for the application of unit root tests to the series \( y_t \).

If we further assume that \( \delta = 1 \), then \( \Sigma \) is given by equation (7) and the (co)variances of \( y_t \) are given by

\[
\begin{align*}
\text{Var}(y_t) & = \frac{\mu_2}{1-\sigma^2} \left\{ 1 + \frac{\vartheta \alpha}{1-\phi c} \left[ \frac{(1+\xi^2-\frac{2}{\pi})\vartheta \alpha(1+\phi c)}{1-c^2} - 2\xi \phi \right] \right\}, \\
\text{Cov}_k(y_t) & = \mu_2 \left\{ \frac{\varphi^k}{1-\sigma^2} + \varphi^2 \alpha^2 \lambda^{(k)}(1+\xi^2-\frac{2}{\pi}) \right. \\
& \quad - \left. \frac{\vartheta \xi \alpha}{(\phi-c)(1-\phi c)} \left[ \frac{\varphi^k(1+\phi^2-2\phi c)}{1-\phi^2} - c^k \right] \right\}, \quad k \geq 1.
\end{align*}
\]

From the above equation it follows that \( \text{Cov}_k(y_t) > \frac{\mu_2 \varphi^k}{1-\sigma^2} \) if either \( \vartheta \xi < 0 \), or \( |\vartheta| > \frac{\alpha \lambda^{(k)}(1+\xi^2-\frac{2}{\pi})}{\sigma} \). If this is the case then as \( |\vartheta| \) increases, \( \text{Cov}_k(y_t) \) increases. When \( \vartheta \) is either 0 or \( \frac{\alpha \lambda^{(k)}(1+\xi^2-\frac{2}{\pi})}{\sigma} \), then \( \text{Cov}_k(y_t) = \frac{\mu_2 \varphi^k}{1-\sigma^2} \). Moreover, \( \text{Cov}_k(y_t) \) increases when i) \( \vartheta \) increases and \( \vartheta > \frac{\alpha \lambda^{(k)}(1+\xi^2-\frac{2}{\pi})}{\sigma} \), ii) \( \alpha \) increases and \( \alpha > \frac{\alpha \lambda^{(k)}(1+\xi^2-\frac{2}{\pi})}{\sigma} \), and iii) \( \xi \) increases and \( \xi > \frac{\alpha \lambda^{(k)}(1+\xi^2-\frac{2}{\pi})}{\sigma} \). Interestingly, when \( \vartheta = \frac{\xi(1-c^2)}{\alpha(1+\xi^2-\frac{2}{\pi})} \), \( \text{Cov}_k(y_t) \) reduces to \( \frac{\mu_2 \varphi^k}{1-\sigma^2} (1 + \frac{2\alpha}{c}) \).

The latter 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.

In summary, our results on the covariance structure of \( y_t \) suggest that a process with persistent conditional variance and in-mean effect may easily be confused with a process that is integrated of order one in the level. The corresponding results for the correlation structure of the powered conditional variance are presented in Appendix A.3 together with the cross correlations between the process \( y_t \) and its powered conditional variance.
2.2 Measures of Persistence

The above 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

a) the largest autoregressive root (LAR), which we denote by $\lambda^* = \max(\lambda_1, \lambda_2)$ and

b) the sum of the coefficients (SUM) in the autoregressive process, that is $a_1 + a_2$ (see, e.g., Pivetta and Reis, 2007).\(^2\)

Obviously, both measures would ignore the presence of the in-mean effect and, hence, ‘overestimate’ the persistence in the mean which is partly induced by the persistence in the conditional variance.

Hence, we follow Fiorentini and Sentana (1998) who argue that any reasonable measure of persistence of shocks must be based on the impulse response function (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 $P_\infty(x_t|u_t) = \text{Var}(x_t)/\text{Var}(u_t) = \sum_{j=0}^{\infty} \psi_j^2$. Clearly, $P_\infty(x_t|u_t)$ will take its minimum value of one if $x_t$ is white noise and will be infinite for a non-stationary process. Note that 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 non-stationary processes.

In the following we will suggest persistence measures which are able to distinguish between the effect of a \textit{mean shock} and a \textit{volatility shock} on the level and conditional variance respectively. We begin by obtaining the bivariate Wold representation of the univariate processes for $y_t$ and $h_t^{5/2}$ given in equations (8) and (9).

\section*{Proposition 3} Let Assumption (A1) hold. Then, equations (8) and (9) admit the Wold

\footnote{If $\vartheta_\gamma > 0$ then $a_1 + a_2 = \phi + c(1 - \phi) + \vartheta_\gamma > \phi + c(1 - \phi) > \phi, c$.}
representation

\[
\begin{pmatrix}
y_t \\
\frac{\delta}{h_t^2}
\end{pmatrix} = \begin{pmatrix}
y^* \\
\mu_1
\end{pmatrix} + \begin{pmatrix}
\psi_{ye}(L) & \psi_{yv}(L) \\
\psi_{he}(L) & \psi_{hv}(L)
\end{pmatrix} \begin{pmatrix}
\varepsilon_t \\
v_t
\end{pmatrix},
\]

(15)

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

\[
\begin{align*}
\psi_{ye}^{(0)} &= 1, & \psi_{ye}^{(k)} &= \left[ \frac{\lambda_1^k (\lambda_1 - c)}{\lambda_1 - \lambda_2} + \frac{\lambda_2^k (\lambda_2 - c)}{\lambda_2 - \lambda_1} \right], & k \geq 1, \\
\psi_{ye}^{(0)} &= 0, & \psi_{yv}^{(k)} &= \partial \alpha \left( \frac{\lambda_1^k}{\lambda_1 - \lambda_2} + \frac{\lambda_2^k}{\lambda_2 - \lambda_1} \right), & k \geq 1, \\
\psi_{he}^{(0)} &= 0, & \psi_{he}^{(k)} &= \gamma \left( \frac{\lambda_1^k}{\lambda_1 - \lambda_2} + \frac{\lambda_2^k}{\lambda_2 - \lambda_1} \right), & k \geq 1, \\
\psi_{hv}^{(0)} &= 0, & \psi_{hv}^{(1)} &= \alpha, & \psi_{hv}^{(k)} &= \alpha \left[ \frac{\lambda_1^{k-1} (\lambda_1 - \phi)}{\lambda_1 - \lambda_2} + \frac{\lambda_2^{k-1} (\lambda_2 - \phi)}{\lambda_2 - \lambda_1} \right], & k \geq 2.
\end{align*}
\]

**Remark 2** If \( \sigma_{ey} = 0 \), then \( \psi_{ye}^{(k)} \) and \( \psi_{yv}^{(k)} \) are the ‘impulse response functions’ of a one unit mean shock \( \varepsilon_t \) or variance shock \( v_t \) to the process \( y_t \). The expression for \( \psi_{yv}^{(k)} \) shows that a positive shock to the conditional variance can either imply a positive or a negative response, depending on the sign of \( \delta \). To the contrary, a positive shock to the mean always implies a positive response. We now consider the situation where \( \gamma = 0 \) and, hence, \( \lambda_1 = \phi \) and \( \lambda_2 = c \). For this specific example, the above expressions reduce to:

\[
\psi_{ye}^{(k)} = \phi^k \quad \text{and} \quad \psi_{yv}^{(k)} = \partial \alpha \left( \frac{\phi^k - c^k}{\phi - c} \right).
\]

Recall that in general the shocks \( \varepsilon_t \) and \( v_t \) will be correlated with covariance matrix \( \Sigma \) (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 obtained from the original shocks
via the transformation

\[
\begin{pmatrix}
\tilde{\varepsilon}_t \\
\tilde{v}_t
\end{pmatrix} =
\begin{pmatrix}
\sigma_\varepsilon & 0 \\
\rho_{ev}\sigma_v & \sigma_v\sqrt{1 - \rho_{ev}^2}
\end{pmatrix}^{-1}
\begin{pmatrix}
\varepsilon_t \\
v_t
\end{pmatrix} = \Sigma^{-1}_1 \begin{pmatrix}
\varepsilon_t \\
v_t
\end{pmatrix}
\]

where it is easy to check that $\Sigma \Sigma' = \Sigma$. We can now rewrite the Wold representation in terms of the orthogonal innovations

\[
\begin{pmatrix}
y_t \\
h_t^{\frac{3}{2}} \\
\mu_t
\end{pmatrix} =
\begin{pmatrix}
y^* \\
\psi_{\varepsilon e}(L) & \psi_{\varepsilon v}(L) \\
\psi_{he}(L) & \psi_{hv}(L)
\end{pmatrix}
\begin{pmatrix}
\sigma_\varepsilon & 0 \\
\rho_{ev}\sigma_v & \sigma_v\sqrt{1 - \rho_{ev}^2}
\end{pmatrix}
\begin{pmatrix}
\tilde{\varepsilon}_t \\
\tilde{v}_t
\end{pmatrix}.
\]

Now, the variance of $y_t$ can be decomposed in two parts as

\[
\mathbb{V}(y_t) = P_\infty(y_t|\tilde{\varepsilon}_t) + P_\infty(y_t|\tilde{v}_t) \tag{16}
\]

where

\[
P_\infty(y_t|\tilde{\varepsilon}_t) = \sigma_\varepsilon^2 P_\infty(y_t|\varepsilon_t) + \rho_{ev}^2 \sigma_v^2 P_\infty(y_t|v_t) + 2\sigma_{ev} P_\infty(y_t|\sqrt{\varepsilon_tv_t}) \tag{17}
\]

and

\[
P_\infty(y_t|\tilde{v}_t) = \sigma_v^2 (1 - \rho_{ev}^2) P_\infty(y_t|v_t) \tag{18}
\]

with individual components $P_\infty(y_t|\varepsilon_t) = \sum_{k=0}^{\infty} (\psi_{\varepsilon e}^{(k)})^2$, $P_\infty(y_t|v_t) = \sum_{k=0}^{\infty} (\psi_{\varepsilon v}^{(k)})^2$ and $P_\infty(y_t|\sqrt{\varepsilon_tv_t}) = \sum_{k=0}^{\infty} \psi_{\varepsilon e}^{(k)} \psi_{\varepsilon v}^{(k)}$.

Following Fiorentini and Sentana (1998) we define the persistence of a standardized mean shock as $P_\infty(y_t|\tilde{\varepsilon}_t)$, i.e. the part of the variance of $y_t$ which is due to $\tilde{\varepsilon}_t$ innovations, and the persistence of a standardized volatility shock as $P_\infty(y_t|\tilde{v}_t)$, i.e. the part of the variance of $y_t$ which is due to $\tilde{v}_t$ innovations. Note that if there is no asymmetry, i.e. $\varsigma = 0$, then $\sigma_{ev} = 0$ and the two persistence measures $P_\infty(y_t|\tilde{\varepsilon}_t)$ and $P_\infty(y_t|\tilde{v}_t)$ are equal to the persistence measures with respect to the original shocks $\varepsilon_t$ and $v_t$ scaled by the

\footnote{A similar decomposition of the variance of $h_t^{3/2}$ is provided in Appendix A.4.}
corresponding variance.

In the following proposition we derive the expression for the individual components of the two persistence measures.

**Proposition 4** If $0 < \Sigma < \infty$ then $P_\infty(y_t|\tilde{\varepsilon}_t)$ and $P_\infty(y_t|\tilde{v}_t)$ are given by equations (17) and (18) respectively where

\[
P_\infty(y_t|\tilde{\varepsilon}_t) = 1 + \frac{1}{(\lambda_1 - \lambda_2)^2} \left[ \frac{\lambda_1^2(\lambda_1 - c)^2}{1 - \lambda_1^2} + \frac{\lambda_2^2(\lambda_2 - c)^2}{1 - \lambda_2^2} - \frac{2\lambda_1\lambda_2(\lambda_1 - c)(\lambda_2 - c)}{1 - \lambda_1\lambda_2} \right],
\]

(19)

\[
P_\infty(y_t|\tilde{v}_t) = \frac{(\dot{\vartheta}\alpha)^2(1 + \lambda_1\lambda_2)}{(1 - \lambda_1^2)(1 - \lambda_2^2)(1 - \lambda_1\lambda_2)},
\]

(20)

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

(21)

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

\[
P_\infty(y_t|\tilde{\varepsilon}_t) = \frac{1}{1 - \phi^2},
\]

\[
P_\infty(y_t|\tilde{v}_t) = \frac{(\dot{\vartheta}\alpha)^2(1 + \phi c)}{(1 - \phi^2)(1 - c^2)(1 - \phi c)},
\]

\[
P_\infty(y_t|\sqrt{\varepsilon_t v_t}) = \frac{\dot{\vartheta}\alpha\phi}{(1 - \phi c)(1 - \phi^2)}.
\]

Lemma 2 shows that without a level effect the expression for $P_\infty(y_t|\tilde{\varepsilon}_t)$ is the one for an AR(1) process (see Fiorentini and Sentana, 1998). Further, notice that if there is no in-mean effect, i.e. $\vartheta = 0$, then $P_\infty(y_t|v_t) = P_\infty(y_t|\sqrt{\varepsilon_t v_t}) = 0$. That is $P_\infty(y_t|\tilde{\varepsilon}_t)$ is again equal to $\sigma^2 \sigma^2 P_\infty(y_t|\varepsilon_t)$. If there is no intrinsic persistence in $y_t$, $\phi = 0$, then $P_\infty(y_t|\varepsilon_t) = 1$, that is the persistence measure takes its minimum value. Secondly, $\phi = 0$ also implies that $P_\infty(y_t|v_t) = \vartheta^2 \alpha^2/(1 - c^2) = \vartheta^2 P_\infty(h_t^{\delta/2}|v_t)$, i.e. 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 Appendix A.4). Thirdly, the restriction $\phi = 0$ also implies $P_\infty(y_t|\sqrt{\varepsilon_t v_t}) = 0$. 19
If we further assume that $\delta = 1$, then $\Sigma$ is given by equation (7) and $P_\infty(y_t|\bar{z}_t)$ and $P_\infty(y_t|\bar{v}_t)$ are given by

$$P_\infty(y_t|\bar{z}_t) = \frac{\mu_2}{1 - \phi^2} \left\{ 1 + \frac{\varsigma \vartheta \alpha}{(1 - \phi c)} \left[ \frac{\varsigma \vartheta \alpha (1 + \phi c)}{1 - c^2} - 2\phi \right] \right\},$$

$$P_\infty(y_t|\bar{v}_t) = \mu_2 \left( 1 - \frac{2}{\pi} \right) \frac{(\vartheta \alpha)^2(1 + \phi c)}{(1 - \phi^2)(1 - c^2)(1 - \phi c)}.$$

For a positive $\phi$, $P_\infty(y_t|\bar{z}_t) > \frac{\mu_2}{1 - \phi^2}$, if either $\varsigma \vartheta < 0$ or $\varsigma \vartheta > 0$ and $|\vartheta| > \frac{2\phi(1-c^2)}{c(1+\phi c)}$. Moreover, $P_\infty(y_t|\bar{z}_t)$ increases when i) $\vartheta$ increases and $\vartheta > \frac{\phi(1-c^2)}{c(1+\phi c)}$, ii) $\varsigma$ increases and $\varsigma > \frac{\phi(1-c^2)}{\vartheta \alpha(1+\phi c)}$, and iii) $\alpha$ increases and $\alpha > \frac{c(1-c^2)}{\vartheta \alpha(1+\phi c)}$. Finally, recall that $V(y_t) = P_\infty(y_t|\bar{z}_t) + P_\infty(y_t|\bar{v}_t)$. Thus if $\delta = 1$ then

$$V(y_t) = \frac{\mu_2}{1 - \phi^2} \left\{ 1 + \frac{\vartheta \alpha}{1 - \phi c} \left[ \frac{(1 + \varsigma^2 - \frac{2}{\pi}) \vartheta \alpha (1 + \phi c)}{1 - c^2} - 2\varsigma \phi \right] \right\}.$$

From the above equation it follows that, for positive $\phi$, $V(y_t) > \frac{\mu_2}{1 - \phi^2}$ if either $\vartheta \varsigma < 0$, or $|\vartheta| > \frac{2\phi(1-c^2)}{(1+\varsigma^2-\frac{2}{\pi})\alpha(1+\phi c)}$. If this is the case then as $|\vartheta|$ increases, $V(y_t)$ increases. When $\vartheta$ is either 0 or $\frac{2\phi(1-c^2)}{(1+\varsigma^2-\frac{2}{\pi})\alpha(1+\phi c)}$ then $V(y_t) = \frac{\mu_2}{1 - \phi^2}$. Moreover, $V(y_t)$ increases when i) $\vartheta$ increases and $\vartheta > \frac{\phi(1-c^2)}{(1+\varsigma^2-\frac{2}{\pi})\alpha(1+\phi c)}$, ii) $\alpha$ increases and $\alpha > \frac{\phi(1-c^2)}{(1+\varsigma^2-\frac{2}{\pi})\vartheta(1+\phi c)}$, and iii) $\varsigma$ increases and $\varsigma > \frac{\phi(1-c^2)}{\vartheta \alpha(1+\phi c)}$. Interestingly, when $\vartheta = \frac{\varsigma(1-c^2)}{\alpha(1+\varsigma^2-\frac{2}{\pi})}$ $V(y_t)$ reduces to $\frac{\mu_2}{1 - \phi^2}(1 + \frac{\vartheta \alpha}{c})$. The latter result is in line with the fact that in this case $\theta = c$ and the ARMA(2,1) representation (11) reduces to an AR(1) process.

### 2.3 Optimal predictors

In this section we provide a formula for the $n$-step ahead predictor $E(y_{t+n}|\mathcal{F}_t)$ of $y_{t+n}$ given the information available at time $t$.

**Proposition 5** Under Assumption (A1) the $n$-step ahead predictor of $y_{t+n}$ is readily seen
to be

\[ \mathbb{E}(y_{t+n} | \mathcal{F}_t) = \frac{1}{\lambda_1 - \lambda_2} \left\{ \tilde{\phi} + [(\lambda_1^{n+1} - \lambda_2^{n+1})y_t - (\lambda_1^{n+1}\lambda_2 - \lambda_2^{n+1}\lambda_1)y_{t-1}] + (\lambda_1^n - \lambda_2^n)(\vartheta\alpha v_t - c\varepsilon_t) \right\} \]

where

\[ \tilde{\phi} = \rho^* \frac{\lambda_1[1 - \lambda_1^n(1 - \lambda_2)] + \lambda_2[1 - \lambda_2^n(1 - \lambda_1)]}{(1 - \lambda_1)(1 - \lambda_2)}. \]

The associated forecast error, \( \mathbb{E}(y_{t+n} | \mathcal{F}_t) = y_{t+n} - \mathbb{E}(y_{t+n} | \mathcal{F}_t) \) is given by

\[ \mathbb{E}(\mathbb{E}(y_{t+n} | \mathcal{F}_t)) = \vartheta\alpha \sum_{l=0}^{n-2} \frac{1}{\lambda_1 - \lambda_2} \left[ \lambda_1^{l+1}L^l - \lambda_2^{l+1}L^l \right] v_{1+n-1} + \left\{ 1 + \frac{1}{\lambda_1 - \lambda_2} \left[ \lambda_1(\lambda_1 - c) \sum_{l=1}^{n-1} \lambda_1^{l-1}L^l - \lambda_2(\lambda_2 - c) \sum_{l=1}^{n-1} \lambda_2^{l-1}L^l \right] \right\} \varepsilon_{t+n}. \]

If \( 0 < \Sigma < \infty \), then the variance of the forecast error is given by

\[ \nabla \{ \mathbb{E}(y_{t+n} | \mathcal{F}_t) \} = P_n(y_t|\bar{\varepsilon}_t) + P_n(y_t|\bar{\nu}_t) \]

where \( P_n(y_t|\bar{\varepsilon}_t) \) and \( P_n(y_t|\bar{\nu}_t) \) are defined analogously to \( P_\infty(y_t|\bar{\varepsilon}_t) \) and \( P_\infty(y_t|\bar{\nu}_t) \) with

\[ P_n(y_t|\bar{\varepsilon}_t) = P_\infty(y_t|\bar{\varepsilon}_t) - \frac{1}{(\lambda_1 - \lambda_2)^2} \left[ \frac{\lambda_1^{2(n+1)}(\lambda_1 - c)^2}{1 - \lambda_1^2} - \frac{\lambda_2^{2(n+1)}(\lambda_2 - c)^2}{1 - \lambda_2^2} - \frac{2\lambda_1\lambda_2^{n+1}(\lambda_1 - c)(\lambda_2 - c)}{1 - \lambda_1\lambda_2} \right], \]

\[ P_n(y_t|\bar{\nu}_t) = P_\infty(y_t|\bar{\nu}_t) - (\vartheta\alpha)^2 \left[ \frac{\lambda_1^{2n}}{1 - \lambda_1^2} - \frac{\lambda_2^{2n}}{1 - \lambda_2^2} - \frac{2\lambda_1\lambda_2^n}{1 - \lambda_1\lambda_2} \right], \]

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

Obviously, for covariance stationary processes \( \lim_{n \to \infty} P_n(y_t|\bar{\varepsilon}_t) = P_\infty(y_t|\bar{\varepsilon}_t) \) and \( \lim_{n \to \infty} P_n(y_t|\bar{\nu}_t) = P_\infty(y_t|\bar{\nu}_t) \).
2.4 Unit Root Tests

In order to obtain some further insights into the persistence properties of $y_t$, we generated series $y_t$, $t = 1, \ldots, T$, with $T = 1000$ according to our AR(1)-GARCH(1,1)-in-mean model. For simplicity, we set $\delta = 2$ and $\gamma = 0$, i.e. we consider a GARCH model with no level effect. The innovation $e_t$ was chosen to be standard normal. We then applied the standard Dickey-Fuller test to the $y_t$ series (results for the Phillips-Perron and the KPSS tests are available upon request). The following figures show the fraction of cases in which the null hypothesis of the Dickey-Fuller test is rejected at the 5% significance level. The results are based on 10,000 Monte-Carlo replications.

We first consider the case $\phi = 1$, i.e. we are under the null hypothesis. Figure 14 shows rejection rates when $\zeta = 0$. Clearly, for $\vartheta = 0$ the Dickey-Fuller test is only slightly oversized. However, the size distortion becomes stronger with $|\vartheta|$ increasing. That is, although $y_t$ is I(1), the Dickey-Fuller test will reject the null hypothesis in the presence of an in-mean effect. For a given value of $\vartheta$ and $c$, the size distortion will be the stronger the larger is $\alpha$. Figure 15 shows similar results for $\zeta \neq 0$.

Next, we consider the behavior of the Dickey-Fuller test under the alternative. In Figure 16, we set $\zeta = 0$ and $\phi = 0.9$, i.e. $y_t$ is stationary. The figure clearly shows that the power of the test deteriorates with $|\vartheta|$ increasing. That is, the larger $|\vartheta|$ the more persistence is transmitted from the conditional variance to the level of $y_t$ and, hence, the Dickey-Fuller test tends to accept the null hypothesis of $y_t$ being I(1). The results clearly suggest that unit root tests will have low power against alternatives which allow for in-mean effects in combination with persistent conditional variances and, hence, a stationary $y_t$ process may easily be confused with a process that is integrated of order one in the level.
In this section we apply our model to U.S. inflation data. Seasonally adjusted monthly consumer price index data was obtained from the Federal Reserve Bank of St. Louis for the period 01/1947 - 12/2008. The CPI inflation rate was calculated as $\pi_t = 1200 \times [\ln(P_t) - \ln(P_{t-1})]$, where $P_t$ denotes the price level in month $t$.

Table 1 presents the estimation results for the full sample and the two subsamples 01/1947 - 12/1983 and 01/1984 - 12/2008. For the full sample, the AR(1)-AGARCH(1, 1)-in-mean implies a positive and significant in-mean term, i.e. more inflation uncertainty leads to higher rates of inflation. The AR(1) coefficient is 0.47 and, as expected, highly significant. The asymmetry term $\zeta$ is found to be negative and significant, suggesting that positive inflation shocks create more uncertainty than negative ones of the same size. The parameter values for $\alpha$ and $\beta$ are rather typical for inflation data and the implied persistence parameter $c = \lambda_2 = 0.96$ suggests strong persistence in the conditional variance but still ensures the existence of the unconditional second moment of $y_t$. For the estimated ARMA(2, 1) model the AR(1) coefficient is larger than one and the AR(2) coefficient is negative. Both observations are in line with our AR(1)-GARCH(1, 1)-in-mean specification. Also the two estimated roots $\lambda_1$ and $\lambda_2$ from the ARMA(2, 1) model are close to their counterparts from the AR(1)-GARCH(1, 1)-M.

The results from the first subsample are quite similar to the ones from the full sample. However, note that the in-mean term is considerable larger and highly significant, implying that the effect of inflation uncertainty on the level of inflation was stronger in the first subsample. To the contrary, our estimates for the second subsample suggest that the in-mean was negative from 1984 onwards, reflecting a change in monetary policy in the early 1980s (the Volcker disinflation). The asymmetry term is now much smaller and no longer significant. Although also the $\alpha$ and $\beta$ parameters change from the first to the second subsample, the overall persistence in the conditional variance – as measure by $\lambda_2$ – remains roughly the same. Nevertheless, the estimated parameter constellations
Table 1: Model Estimates.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AR(1)-GARCH(1,1)-M</strong></td>
<td><strong>ARMA(2,1)</strong></td>
<td><strong>AR(1)-GARCH(1,1)-M</strong></td>
<td><strong>ARIMA(1,1,1)</strong></td>
</tr>
<tr>
<td><strong>Mean Equation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>const.</td>
<td>2.97 (0.20)</td>
<td>3.26 (0.72)</td>
<td>2.92 (0.42)</td>
</tr>
<tr>
<td></td>
<td>2.01 (0.20)</td>
<td>3.26 (0.72)</td>
<td>2.92 (0.42)</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.47 (0.04)</td>
<td>1.23 (0.10)</td>
<td>0.43 (0.07)</td>
</tr>
<tr>
<td></td>
<td>0.18 (0.07)</td>
<td>0.34 (0.10)</td>
<td>0.32 (0.07)</td>
</tr>
<tr>
<td>AR(2)</td>
<td>-0.26 (0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.82 (0.07)</td>
<td>0.80 (0.06)</td>
<td>0.99 (0.01)</td>
</tr>
<tr>
<td></td>
<td>0.80 (0.06)</td>
<td>0.99 (0.01)</td>
<td></td>
</tr>
<tr>
<td>$\vartheta$</td>
<td>0.05 (0.02)</td>
<td>0.11 (0.03)</td>
<td>-0.06 (0.03)</td>
</tr>
<tr>
<td></td>
<td>0.11 (0.03)</td>
<td>-0.06 (0.03)</td>
<td></td>
</tr>
<tr>
<td><strong>Variance Equation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.27 (0.14)</td>
<td>0.31 (0.14)</td>
<td>0.60 (0.30)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.12 (0.05)</td>
<td>0.06 (0.04)</td>
<td>0.27 (0.10)</td>
</tr>
<tr>
<td>$\varsigma$</td>
<td>-0.17 (0.09)</td>
<td>-0.55 (0.32)</td>
<td>-0.15 (0.13)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.84 (0.05)</td>
<td>0.88 (0.04)</td>
<td>0.67 (0.10)</td>
</tr>
<tr>
<td></td>
<td>0.88 (0.04)</td>
<td>0.67 (0.10)</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>5.014</td>
<td>5.251</td>
<td>5.415</td>
</tr>
<tr>
<td></td>
<td>5.415</td>
<td>4.630</td>
<td>4.328</td>
</tr>
<tr>
<td>SIC</td>
<td>5.058</td>
<td>5.276</td>
<td>5.311</td>
</tr>
<tr>
<td></td>
<td>5.442</td>
<td>4.717</td>
<td>4.370</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.47</td>
<td>0.27</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>0.18</td>
<td>0.34</td>
<td>0.32</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.96</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** The numbers in parenthesis are Bollerslev-Wooldrige robust standard errors.

suggest that the implied MA(1) parameter should be higher in the second than in the first subsample.

Considering the corresponding ARIMA(1,1,1) models, the parameter estimates are indeed line with our AR(1)-GARCH(1,1)-in-mean specifications. The MA(1) coefficient is bigger in the second subsample, an observation also made in Stock and Watson (2007).

4 Conclusions

We discuss the persistence properties of the AR(1)-GARCH(1,1)-in-mean-level model. This model allows for an in-mean effect as well as a level effect. Both effects are in
line with economic theory which, e.g., suggests that inflation uncertainty should have an
effect on the level of inflation and vice versa. Our main result is that the commonly
observed persistence in the mean/conditional variance of many economic times series
may be a result of a transmission mechanism. If this mechanism is ignored, conventional
procedures for estimating the persistence in the mean/variance may lead to upward biased
estimates. In particular, unit root tests will falsely indicate a unit root and, hence,
suggest the modeling of the differenced series rather than the level series. Our primary
empirical example are inflation rates and we argue that the decrease in both U.S. inflation
persistence and inflation variability from the mid 1980’s onwards can be well explained
by our specification.

References

etary Economics 29, 371–388.


the presence of non-stationary volatility. Journal of Econometrics 147, 84-98.


exchange rate? in J. Y. Campbell (ed.) Asset Prices and Monetary Policy, University
of Chicago Press.


A Appendix

A.1 Moving Average Parameter $\theta$

As we mentioned in Section 2, when $\delta = 1$ then $\sigma_0 = \mu_2\{(1+c^2)+(\vartheta \alpha)^2(1+\varsigma^2-\frac{2}{\pi})+2c\vartheta \alpha \varsigma\}$, and $\sigma_1 = -\mu_2[c + \vartheta \alpha \varsigma]$. Therefore the moving average parameter $\theta$ in the ARMA(2,1)
representation of $y_t$ reduces to

$$
\theta = \frac{-[(1 + c^2 + 2c\vartheta\alpha) + (\vartheta\alpha)^2(1 + \zeta^2 - \frac{2}{\pi})]}{-2(c + \vartheta\alpha)} + \frac{\sqrt{[(1 + c^2 + 2c\vartheta\alpha) + (\vartheta\alpha)^2(1 + \zeta^2 - \frac{2}{\pi})]^2 - 4(c + \vartheta\alpha)^2}}{2(c + \vartheta\alpha)} - \frac{(1 + c^2 + 2c\vartheta\alpha) + (\vartheta\alpha)^2(1 + \zeta^2 - \frac{2}{\pi})}{2(c + \vartheta\alpha)}
$$

In the above equation, $\theta < c$ if either $\vartheta\alpha < 0$, or $|\vartheta| \alpha(1 + \zeta^2 - \frac{2}{\pi}) > |\zeta| (1 - c^2)$ (in the last expression, see the term inside the brackets). If this is the case then as $|\vartheta|$ increases, $\theta$ decreases, and as $\alpha$ increases $c - \theta$ also increases. When either $\vartheta = 0$ or $\vartheta = \frac{\zeta(1-c)^2}{\alpha(1+\zeta^2-\frac{2}{\pi})}$ then $\theta = c$ and the ARMA(2, 1) representation reduces to an AR(1) process. If $c = 1$, that is there is a unit root in the conditional standard deviation, then $\theta = c = 1$ only if there are no in-mean effects: $\vartheta = 0$.

Moreover, when there are no asymmetries, that is $\zeta = 0$, the above equation reduces to

$$
\theta = \frac{[(1 + c^2) + (\vartheta\alpha)^2(1 - \frac{2}{\pi})] - \sqrt{[(1 - c^2) + (\vartheta\alpha)^2(1 - \frac{2}{\pi})]^2 + 4(c\vartheta\alpha)^2(1 - \frac{2}{\pi})}}{2c}.
$$

In the above equation, since $4(c\vartheta\alpha)^2(1 - \frac{2}{\pi}) \geq 0$, $\theta \leq c$. That is, $\theta = c$ only if there are no in-mean effects ($\vartheta = 0$). In other words, as $|\vartheta|$ increases $\theta$ decreases.

A.2 Correlation Structure of $y_t$ and Cross-correlations

In order to simplify the description of our analysis we will introduce the following notation:
\[
\lambda = \frac{1}{(1 - \lambda_1 \lambda_2)(\lambda_1 - \lambda_2)},
\]
\[
\lambda^{(k)} = \lambda \left[ \frac{\lambda_1^{1+k}}{1 - \lambda_1^2} - \frac{\lambda_2^{1+k}}{1 - \lambda_2^2} \right],
\]
\[
\lambda^{(k)}_g = \begin{cases} \lambda \left[ \frac{\lambda_1^{1+k}(\lambda_1 - g)}{1 - \lambda_1^2} - \frac{\lambda_2^{1+k}(\lambda_2 - g)}{1 - \lambda_2^2} \right] & \text{if } k = 0, \\
\lambda \left[ \frac{\lambda_1^{(k)}(1 - g\lambda_1)}{1 - \lambda_1^2} - \frac{\lambda_2^{(k)}(1 - g\lambda_2)}{1 - \lambda_2^2} \right] & \text{if } k \geq 1
\end{cases}, \quad g \in \{\phi, c\}.
\]

Note that \(\lambda^{(0)} = \frac{(1 + \lambda_1 \lambda_2)}{(1 - \lambda_1 \lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)}\) and \(\lambda^{(1)} = \frac{(\lambda_1 + \lambda_2)}{(1 - \lambda_1 \lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)}\).

Further, for \(k \geq 0\) and \(g = \phi, c\), we define
\[
\lambda^{(k)}_g = \lambda \left[ \frac{\lambda_1^{1+k}(\lambda_1 - g)}{1 - \lambda_1^2} - \frac{\lambda_2^{1+k}(\lambda_2 - g)}{1 - \lambda_2^2} \right],
\]
\[
\lambda^{(k)}_g = \lambda \left[ \frac{\lambda_1^{1+k}(1 - g\lambda_1)}{1 - \lambda_1^2} - \frac{\lambda_2^{1+k}(1 - g\lambda_2)}{1 - \lambda_2^2} \right].
\]

Interestingly, if \(\lambda_1 = \phi\) and \(\lambda_2 = c\), then the above expressions reduce to
\[
\lambda^{(k)}_g = \frac{(\phi c/g)^k}{1 - (\phi c/g)^2}, \quad \lambda^{(k)}_g = \frac{(\phi c/g)^{1+k}}{[1 - (\phi c/g)^2](1 - \phi c)}, \quad \lambda^{(k)}_g = \frac{1}{g} \left[ \frac{g^{1+k} - (\phi c/g)^{1+k}}{1 - (\phi c/g)^2} \right].
\]

**Proposition 6** Assume that \(\Sigma\) exists. That is \(0 < \Sigma < \infty\).\textsuperscript{4} Then the (co)variances of \(y_t\), \(\text{Cov}_k(y_t)\), \(k \in \mathbb{N}\), are given by
\[
\text{V}(y_t) = \sigma_c^2 \lambda^{(0)}_c + \sigma^2 v(\partial \alpha)^2 \lambda^{(0)} + 2 \sigma v \partial \alpha \lambda^{(0)}_c, \quad (22)
\]
\[
\text{Cov}_k(y_t) = \lambda^{(k)}_c + \sigma^2 v(\partial \alpha)^2 \lambda^{(k)} + \sigma v \partial \alpha \lambda^{(k)}_c + \lambda^{(k-1)}_c, \quad k \geq 1. \quad (23)
\]

where all the \(\lambda\)’s have been defined above.

\textsuperscript{4}Matrix inequality sign \(\Sigma < \infty\) represents element-by-element inequality.
A.3 Correlation Structure of $h^\delta_t$ and Cross-correlations

Proposition 7 Suppose that $\Sigma$ exists. Then the autocovariances of $h^\delta_t$, $\text{Cov}_k(h^\delta_t)$, $k \in \mathbb{N}$, are given by

$$\text{V}(h^\delta_t) = \sigma^2 \epsilon \gamma^2 \lambda^{(0)} + \sigma^2 \epsilon \alpha^2 \lambda^{(0)} + 2 \sigma \epsilon \gamma \alpha \tilde{\lambda}^{(0)},$$

(24)

$$\text{Cov}_k(h^\delta_t) = \sigma^2 \epsilon \gamma^2 \lambda^{(k)} + \sigma^2 \epsilon \alpha^2 \lambda^{(k)} + \sigma \epsilon \alpha \alpha (\tilde{\lambda}^{(k)} + \tilde{\lambda}^{(k-1)}), \quad k \geq 1.$$  

(25)

Moreover, the autocovariances of $f(\epsilon_t)$ are given by

$$\text{V}[f(\epsilon_t)] = \text{V}(h^\delta_t) + \sigma^2 \epsilon,$$

$$\text{Cov}_k[f(\epsilon_t)] = \text{Cov}_k(h^\delta_t) + \sigma^2 \epsilon \psi_k^{(hv)}, \quad k \geq 1.$$  

Moreover, the autocovariances of $f(\epsilon_t)$ are given by

$$\text{V}[f(\epsilon_t)] = \text{V}(h^\delta_t) + \sigma^2 \epsilon,$$

$$\text{Cov}_k[f(\epsilon_t)] = \text{Cov}_k(h^\delta_t) + \sigma^2 \epsilon \psi_k^{(hv)}, \quad k \geq 1.$$  

Lemma 3 When $\lambda_1 = \phi$ and $\lambda_2 = c$, we have

$$\text{V}(h^\delta_t) = \frac{1}{1 - c^2} \left\{ \sigma^2 \epsilon^2 + \frac{\gamma}{1 - \phi c} \left[ \frac{\sigma^2 \epsilon \gamma (1 + \phi c)}{1 - \phi^2} + 2 \sigma \epsilon \alpha c \right] \right\},$$

$$\text{Cov}_k(h^\delta_t) = \sigma^2 \epsilon \gamma^2 \lambda^{(k)} + \frac{\sigma^2 \epsilon \alpha^2 c^k}{1 - c^2} + \frac{\sigma \epsilon \alpha \alpha}{(1 - \phi c)(\phi - c)} \left[ \phi^{1 + k} + \frac{c^k (\phi - 2c + \phi c^2)}{1 - c^2} \right].$$

Further, if $\gamma = 0$, then the above expressions reduce to $\text{Cov}_k(h^\delta_t) = \frac{\sigma^2 \epsilon^2 c^k}{1 - c^2}$, $k \geq 0$, which are the autocovariances of the APARCH(1,1) model.

Next, define

$$\lambda_c^{(0)} = \lambda \left\{ \frac{\lambda_1 [-c + (1 + \phi) \lambda_1 - \phi \lambda_1^2]}{1 - \lambda_1^2} - \frac{\lambda_1 [-c + (1 + \phi) \lambda_1 - \phi \lambda_1^2]}{1 - \lambda_2^2} \right\},$$

$$\lambda_c^{(1)} = \lambda \left\{ \frac{\lambda_2 [-c + (1 + \phi) \lambda_1^{-1} - \phi]}{1 - \lambda_1^2} - \frac{\lambda_2 [-c + (1 + \phi) \lambda_1^{-1} - \phi]}{1 - \lambda_2^2} \right\},$$

$$\lambda_c^{(k)} = \lambda \left\{ \frac{\lambda_1^{1+k} [-c + (1 + \phi) \lambda_1^{-1} - \phi \lambda_1^2]}{1 - \lambda_1^2} - \frac{\lambda_2^{1+k} [-c + (1 + \phi) \lambda_2^{-1} - \phi \lambda_2^2]}{1 - \lambda_2^2} \right\}, \quad k \geq 2.$$  

Interestingly, if $\lambda_1 = \phi$ and $\lambda_2 = c$, then the above expressions reduce to
\[ \lambda^{(0)}_{c\phi} = \frac{\phi}{1 - \phi c}, \quad \lambda^{(1)}_{c\phi} = \frac{1}{1 - \phi c}, \quad \lambda^{(k)}_{c\phi} = \frac{c^k}{1 - \phi c}. \]

**Proposition 8** If \(0 < \Sigma < \infty\) then the covariances between \(y_{t-k}\) and \(h_k, k \in \mathbb{N}\), are given by

\[
\text{Cov}(y_t, h_t^2) = \sigma^2 \gamma \lambda^{(0)}_c + \sigma^2 \alpha^2 \lambda^{(0)}_\phi + \sigma_{ev} \alpha (\lambda^{(0)}_{c\phi} + \gamma \vartheta \lambda^{(0)}), \tag{26}
\]

\[
\text{Cov}(y_{t-k}, h_t^2) = \sigma^2 \gamma \lambda^{(k-1)}_c + \sigma^2 \alpha^2 \lambda^{(k-1)}_\phi + \sigma_{ev} \alpha (\lambda^{(k)}_{c\phi} + \gamma \vartheta \lambda^{(k)}), \quad k \geq 1. \tag{27}
\]

**Lemma 4** Interestingly if \(\lambda_1 = \phi, \lambda_2 = c\) then

\[
\text{Cov}(y_t, h_t^2) = \frac{\sigma^2 \phi \gamma}{(1 - \phi c)(1 - \phi^2)} + \frac{\sigma^2 \alpha^2}{(1 - \phi c)(1 - c^2)} + \sigma_{ev} \alpha \left[ \frac{\phi}{1 - \phi c} + \gamma \vartheta \lambda^{(0)} \right],
\]

\[
\text{Cov}(y_{t-1}, h_t^2) = \frac{\sigma^2 \gamma}{(1 - \phi c)(1 - \phi^2)} + \frac{\sigma^2 \alpha^2}{(1 - \phi c)(1 - c^2)} + \sigma_{ev} \alpha \left[ \frac{1}{1 - \phi c} + \gamma \vartheta \lambda^{(1)} \right].
\]

Further, if \(\gamma = \sigma_{ev} = 0\), then \(\text{Cov}(y_{t-k}, h_t^2) = \frac{\sigma^2 \alpha^2 c^k}{(1 - \phi c)(1 - c^2)}\).

Next define

\[ \zeta_1 = \mathbb{E}(e_t^2) \gamma^2 \lambda^{(0)}, \quad \zeta_2 = \tilde{\kappa} \alpha^2 \lambda^{(0)}_\phi, \quad \zeta_3 = 2 \mathbb{K} \gamma \alpha \lambda^{(0)}_\phi, \quad \zeta = \zeta_1 + \zeta_2 + \zeta_3. \]

**Proposition 9** Let Assumption (A1) be satisfied and assume \(\delta > 1\). If \(\mu_{1+1/\delta} < \infty\) and \(\zeta_2 < 1\) then \(\mu_2\) exists and it is given by

\[
\mu_2 = \frac{\mu_1^2 + \mu_2 \zeta_1 + \mu_{1+1/2} \zeta_3}{1 - \zeta_2}. \tag{28}
\]

Moreover, if \(\delta = 1\) and \(\zeta < 1\), then \(\mu_2\) exists and it is given by

\[
\mu_2 = \frac{\mu_1^2}{1 - \zeta}. \tag{29}
\]
Lemma 5 When $\gamma = \zeta = 0$ and $e_t$ is normally distributed then equation (28) reduces to

$$\mu_2 = \frac{\mu_1^2(1-\beta^2)}{(1-3\alpha^2-\beta^2-2\alpha\beta)}.$$  

(30)

### A.4 Measures of Persistence for $h_t^{r/2}$

Next, we will derive the persistence measures for $h_t^{r/2}$ given standardized mean and volatility shocks. As before, the persistence measures are obtained by decomposing the variance of $h_t^{r/2}$ in the two parts which are due to $\tilde{e}_t$ and $\tilde{v}_t$:

$$\text{Var}(h_t^{r/2}) = P_\infty(h_t^{r/2} | \tilde{e}_t) + P_\infty(h_t^{r/2} | \tilde{v}_t)$$

where

$$P_\infty(h_t^{r/2} | \tilde{e}_t) = \sigma_e^2 P_\infty(h_t^{r/2} | e_t) + \rho_{ev}^2 \sigma_v^2 P_\infty(h_t^{r/2} | v_t) + 2\sigma_{ev} P_\infty(h_t^{r/2} | \sqrt{\tilde{e}_t \tilde{v}_t})$$  

(31)

and

$$P_\infty(h_t^{r/2} | \tilde{v}_t) = \sigma_v^2(1 - \rho_{ev}^2) P_\infty(h_t^{r/2} | v_t)$$  

(32)

with individual components $P_\infty(h_t^{r/2} | e_t) = \sum_{k=0}^{\infty} (\psi_{he}^{(k)})^2$, $P_\infty(h_t^{r/2} | v_t) = \sum_{k=0}^{\infty} (\psi_{hv}^{(k)})^2$ and $P_\infty(h_t^{r/2} | \sqrt{\tilde{e}_t \tilde{v}_t}) = \sum_{k=0}^{\infty} \psi_{he}^{(k)} \psi_{hv}^{(k)}$.

**Proposition 10** Assume that $\Sigma$ exists. Then $P_\infty(h_t^{r/2} | \tilde{e}_t)$ and $P_\infty(h_t^{r/2} | \tilde{v}_t)$ are given by equations (31) and (32) respectively, where

$$P_\infty(h_t^{r/2} | \tilde{e}_t) = \frac{\gamma^2 (1 + \lambda_1 \lambda_2)}{(1 - \lambda_1^2)(1 - \lambda_2^2)(1 - \lambda_1 \lambda_2)},$$  

(33)

$$P_\infty(h_t^{r/2} | \tilde{v}_t) = \alpha^2 \left\{ 1 + \frac{1}{(1 - \lambda_1 - \lambda_2)^2} \left[ \frac{\lambda_1^2 (\lambda_1 - \phi)^2}{(1 - \lambda_1^2)} + \frac{\lambda_2^2 (\lambda_2 - \phi)^2}{(1 - \lambda_2^2)} - \frac{2\lambda_1 \lambda_2 (\lambda_1 - \phi)(\lambda_2 - \phi)}{(1 - \lambda_1 \lambda_2)} \right] \right\},$$  

(34)

$$P_\infty(h_t^{r/2} | \sqrt{\tilde{e}_t \tilde{v}_t}) = \gamma \alpha \left[ 1 + \frac{(\lambda_1 - \phi) + \lambda_2 (1 - \phi \lambda_1)}{(1 - \lambda_1^2)(1 - \lambda_2^2)(1 - \lambda_1 \lambda_2)} \right].$$
Lemma 6 Interestingly if $\lambda_1 = \phi$ and $\lambda_2 = c$ we obtain

$$P_\infty(h_t|\varepsilon_t) = \frac{\gamma^2(1 + \phi c)}{(1 - \phi^2)(1 - c^2)(1 - \phi c)}.$$  
$$P_\infty(h_t|v_t) = \frac{\alpha^2}{1 - c^2};$$  
$$P_\infty(h_t|\sqrt{\varepsilon_t}v_t) = \gamma \alpha [1 + \frac{c}{(1 - \phi c)(1 - c^2)}].$$

Obviously, in this case the expression for $P_\infty(h_t|v_t)$ is the one for the GARCH(1,1) process (see Fiorentini and Sentana, 1998). Further, notice that if $\gamma = 0$ $P_\infty(h_t|\varepsilon_t) = P_\infty(h_t|\sqrt{\varepsilon_t}v_t) = 0$.

A.5 Proofs

Proof of Proposition 1. Multiplying equation (1) by $(1-cL)$ and substituting (3) into equation (1) gives (8). Similarly, multiplying equation (3) by $(1-\phi L)$ and substituting (1) into equation (3) gives (9).

Proof of Proposition 2. Taking expectations from both sides of equation (9) yields (10).

Proof of Proposition 3. On account of (8)-(9) and equation (A.1) in Karanasos (2007), we obtain equation (15) by straightforward manipulation.

Proof of Proposition 4. The desired result is obtained straightforwardly from Proposition 3.

Proof of Proposition 5. The proof follows from the ARMA representation of $y_t$ in equation (8) and equation (A.10) in Karanasos (2001).

Proof of Proposition 6. The proof follows from the ARMA representation of $y_t$ in equation (8) and Proposition 2 in Karanasos (2007).

Proof of Proposition 7. The proof follows from the ARMA representation of $h_t^{\frac{\phi}{2}}$ in equation (9) and Proposition 2 in Karanasos (2007).
Proof of Proposition 8. The proof follows from the ARMA representations of $y_t$ and $h_t^\delta$ in Lemma 1 and Proposition 3 in Karanasos (2007).

Proof of Proposition 9. From Proposition 7 and the fact that $\mathbb{V}(h_t^\delta) = \mu_2 - \mu_1^2$, $\sigma^2 = \mu_2/\delta \mathbb{E}(e_t^2)$, $\sigma^2_v = \mu_2 \bar{k}$ and $\sigma_{\epsilon \epsilon} = \mu_1 + \frac{1}{\delta} \bar{k}$ we obtain

$$\mu_2 = \mu_1^2 + \mu_2/\delta \zeta_1 + \mu_2 \zeta_2 + \mu_1 + \frac{1}{\delta} \zeta_3.$$  

Notice that if $\delta < 1$, both $2/\delta$ and $1 + \frac{1}{\delta}$ are greater than 2. When $\delta \geq 1$ the desired result is obtained straightforwardly from the above equation.

Proof of Proposition 10. The desired result is obtained straightforwardly from Proposition 3.
Figure 1: $\sigma_{w\theta}$ as a function of $\zeta$, the other parameters are given by: $\varphi = 0.05$, $\phi = 0.4$, $\vartheta = 0.5$, $\omega = 0.1$, $\alpha = 0.1$, $\beta = 0.7$, $\gamma = 0.0$, $\delta \in \{1, 1.5, 2\}$.

Figure 2: $\theta$ as a function of $\alpha$, the other parameters are given by: $\varphi = 0.05$, $\phi = 0.4$, $\vartheta = 2$, $\omega = 0.1$, $\gamma = 0$, $\beta = 0.6$, $\zeta = 0$ (solid), $\zeta = -0.3$ (dotted) and $\zeta = -0.5$ (dashed). Left panel: $\delta = 2$, right panel: $\delta = 1$. Lower panel: $c$ is kept constant.
Figure 3: $\theta$ as a function of $\varsigma$, the other parameters are given by: $\varphi = 0.05$, $\phi = 0.4$, $\vartheta = 2$, $\omega = 0.1$, $\alpha = 0.1$, $\gamma = 0$, $\beta = 0.6$, $\delta = 2$ (solid), $\delta = 1$ (dotted).

Figure 4: $\theta$ as a function of $\omega$, the other parameters are given by: $\varphi = 0.05$, $\phi = 1.0$, $\vartheta = 2$, $\alpha = 0.1$, $\gamma = 0$, $\beta = 0.7$, $\delta = 2$, $\varsigma = 0$ (solid), $\varsigma = -0.3$ (dotted) and $\varsigma = -0.5$ (dashed).
Figure 5: $\vartheta$ between -3.0 and 3.0. Other parameters: $\varphi = 0.05$, $\phi = 0.0$, $\omega = 0.1$, $\alpha = 0.1$, $\beta = 0.8$, $\gamma = 0.0$, $\delta = 2$, $\varsigma = 0$. Note, that neither $c$ nor $\vartheta$ depend on the value of $\phi$. In particular, it does not matter whether $\phi = 1$ or $\phi < 1$.

Figure 6: $\vartheta$ between -3.0 and 3.0. Other parameters: $\varphi = 0.05$, $\phi = 0.0$, $\omega = 0.1$, $\alpha = 0.1$, $\beta = 0.8$, $\gamma = 0.0$, $\delta = 2$, $\varsigma \in \{-0.5, 0.5\}$. Note, that the value of $c$ depends only on $|\varsigma|$ (but not on the sign).

Figure 7: $\vartheta$ between -3.0 and 3.0. Other parameters: $\varphi = 0.05$, $\phi = 0.0$, $\omega = 0.1$, $\alpha = 0.1$, $\beta = 0.8$, $\gamma = 0.0$, $\delta = 1$, $\varsigma = 0.5$. 

38
Figure 8: ACF of $y_t$ for $\phi = 0.1$ and $\delta$ 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 $\varphi = 0$, $\omega = 0.1$, $\beta = 0.8$, $\gamma = 0$.

Figure 9: ACF of $y_t$ for $\phi = 0.5$ and $\delta$ 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 $\varphi = 0$, $\omega = 0.1$, $\beta = 0.8$, $\gamma = 0$.

Figure 10: ACF of $y_t$ for $\phi = 0.9$ and $\delta$ 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 $\varphi = 0$, $\omega = 0.1$, $\beta = 0.8$, $\gamma = 0$. 

39
Figure 11: ACF of $y_t$ for $\phi = 0.1$ and $\vartheta$ 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 $\varphi = 0$, $\omega = 0.1$, $\alpha = 0.1$, $\gamma = 0$.

Figure 12: ACF of $y_t$ for $\phi = 0.5$ and $\vartheta$ 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 $\varphi = 0$, $\omega = 0.1$, $\alpha = 0.1$, $\gamma = 0$.

Figure 13: ACF of $y_t$ for $\phi = 0.9$ and $\vartheta$ 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 $\varphi = 0$, $\omega = 0.1$, $\alpha = 0.1$, $\gamma = 0$. 

40
Figure 14. Rejection rates of Dickey-Fuller Test (including a constant) as a function of $\vartheta$ between $-2.0$ and $2.0$. Other parameters: $\varphi = 0$, $\phi = 1.0$, $\omega = 0.1$, $\gamma = 0.0$, $\delta = 2$, $\varsigma = 0$. 
Figure 15. Rejection rates of Dickey-Fuller Test (including a constant) as a function of $\theta$ between $-2.0$ and $2.0$. Other parameters: $\varphi = 0$, $\phi = 1.0$, $\omega = 0.1$, $\gamma = 0.0$, $\delta = 2$. 
Figure 16. Rejection rates of Dickey-Fuller Test (including a constant) as a function of \( \vartheta \) between \(-2.0\) and \(2.0\). Other parameters: \( \varphi = 0 \), \( \phi = 0.9 \), \( \omega = 0.1 \), \( \gamma = 0.0 \), \( \delta = 2 \), \( \zeta = 0 \).