AN ARCH-IN-MEAN MODEL WITHOUT INTERCEPT

Christian Hafner, Arie Preminger

LIDAM Discussion Paper ISBA
2026 / 17



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication

An ARCH-in-Mean Model without Intercept

Christian M. Hafner* Arie Preminger†

May 27, 2026

Abstract

This paper introduces an ARCH-in-mean model without intercept, which is characterized by nonstationarity irrespective of the parameters. A critical parameter region is derived for which the volatility process behaves like a random walk. We establish the asymptotic properties of the quasi maximum likelihood estimator (QMLE) for the volatility and risk premium parameters. It is shown that the existence of the second and fourth order moment of the innovations is sufficient to derive the consistency and the asymptotic normality of the QMLE, respectively. The variance of the estimator of the volatility parameter does not depend on the underlying process. We further suggest a test for model stability and a test for the risk premium, which is non-standard as the parameter is on the boundary under the null. The finite sample performance of the estimator and the stability test are investigated in a simulation study. Finally, the model is applied to a monthly stock index series.

Keywords: Volatility, non-stationarity, risk premium, asymptotic theory, maximum likelihood

MSC2020 classifications: 62M10, 91B84

*Corresponding author, UCLouvain, LIDAM, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, e-mail christian.hafner@uclouvain.be

†Academic College of Ramat Gan, Ramat Gan, Israel. Email : ariepr25@iac.ac.il

The authors would like to thank participants of the conference "Score-driven and nonlinear time series models, with applications in economics, finance and climate change", at Ca'Foscari University of Venice, May 2026, for helpful comments and suggestions.

1 Introduction

The relation between risk and return has a long history in finance, going back at least to the portfolio selection theory of Markowitz (1952) and the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965). In a dynamic framework, the intertemporal CAPM of Merton (1973) establishes a positive link between conditional expected returns and the conditional variance, as long as the representative agent has a positive risk aversion coefficient. Empirical studies, such as Ghysels et al (2005) confirm a positive risk-return relationship for stock market data at various frequencies.

On the econometrics side, the most popular model linking conditional expectations of returns with conditional variances is the ARCH-in-mean or ARCH-M model of Engle et al (1987), which extends the classical ARCH and GARCH models of Engle (1982) and Bollerslev (1986) to include a risk premium that directly depends on the conditional variance. The ARCH-M model has been used in numerous empirical studies, see e.g. the handbook of Bauwens et al. (2012). It is however only recently that the asymptotic theory of the maximum likelihood estimator has been thoroughly established by Conrad and Mammen (2016) for the classical GARCH-in-mean model, Hafner and Kyriakopoulou (2021) for EGARCH-in-mean, and Christensen et al. (2012) for a modified version where volatility depends directly on observed returns rather than on innovations.

Most papers on the theory of MLE in ARCH-type models work under the assumption of stationarity, see e.g. the monograph of Francq and Zakoïan (2019) for a detailed treatment, and so do Conrad and Mammen (2016) and Hafner and Kyriakopoulou (2021) for the models of the "in-mean" type. Once stationarity is put into question, the model itself may need to be revised. For example, in the classical GARCH model, there is no consistent estimator for the constant intercept term in the volatility equation if the process is nonstationary, as shown by Francq and Zakoïan (2012). The same issue arises in some versions of score-driven models (Creal et al. (2013), Harvey (2013)), such as the Beta-t-GARCH model of Harvey and Chakravarty (2008).

In this paper, we consider an ARCH-in-mean model without intercept, which is extension of the ARCH model without intercept proposed by Hafner and Preminger (2015) to include a risk premium. Li et al. (2018) extended the model of Hafner and Preminger

(2015) to the GARCH case without intercept, but without risk premium. The volatility equation of our model is in the spirit of Christensen et al. (2012), where volatility is a linear function of lagged squared returns rather than lagged squared error terms. The peculiarity of the models without intercept is that they are nonstationary, irrespective of the parameter values. Nelson (1990) showed that in the classical GARCH model without intercept, volatility converges either to zero almost surely if the Lyapounov exponent is negative, or diverges to infinity if it is positive. There is however an intermediate case with a Lyapounov exponent of zero for which volatility behaves like a random walk. This is the practically relevant case of what we call a "stable" process.

We establish the asymptotic properties of the quasi maximum likelihood estimator (QMLE) for the ARCH-M model without intercept. We show that the existence of the second and fourth order moment of the innovation term is sufficient to derive the consistency and asymptotic normality of the QMLE, respectively. The variance of the estimated volatility parameter does not depend on the underlying process. We also suggest a test for the model stability.

The paper is organized as follows. Section 2 presents the model and states our results. In Section 3 we perform a simulation study, in Section 4 the empirical study, and Section 5 concludes. The proofs are delegated to the appendix.

The following notation will be used throughout. The Euclidian norm is $\|\cdot\|$, the symbol $\xrightarrow{\mathcal{D}}$ and $\xrightarrow{a.s.}$ denote convergence in distribution and almost surely, respectively, and $o_{a.s.}(1)$ denotes a sequence of random variables that converges to zero almost surely.

2 The Model

For a time series of financial returns $\{y_t\}, t \in \mathbb{Z}$, we consider a first order ARCH model without intercept defined by

$$y_t = \lambda\sigma_t + \sigma_t z_t \tag{1}$$

with volatility process σ_t^2 given by

$$\sigma_t^2 = \alpha y_{t-1}^2 \tag{2}$$

where $\{z_t\}$ is a sequence of independent and identically distributed (iid) random variables such that $P(z_t = 0) = 0$. This model is in the spirit of Christensen et al. (2012) and Ling (2004), where volatility is a linear function of lagged squared returns, y_{t-1}^2, rather than lagged squared error terms, $(y_{t-1} - \lambda\sigma_{t-1})^2$ as in classical GARCH models. Obviously we have to assume $\alpha > 0$ as otherwise σ_t^2 would be zero or negative. We will also assume $\lambda \geq 0$ to ensure a non-negative risk premium compatible with finance theory. For the case $\lambda = 0$ the model reduces to that considered by Hafner and Preminger (2015). Furthermore, we assume that the process is initiated at $t = 0$ with initial values $\sigma_0^2 > 0$ a.s. which ensures the positivity of the conditional variance. From (1)–(2) we note that $\sigma_t^2 = \sigma_0^2 \prod_{j=0}^{t-1} \eta_j^2$ where $\eta_t := \frac{y_t}{|y_{t-1}|} = \sqrt{\alpha}(\lambda + z_t)$, which is i.i.d. with mean $\lambda\sqrt{\alpha}$ and variance α. Thus

$$\log \sigma_t^2 = \sum_{j=0}^{t-1} \log \eta_j^2 + \log \sigma_0^2. \tag{3}$$

Define the Lyapunov exponent

$$\gamma_0 = \mathbb{E} \log(\eta_t^2) \tag{4}$$

and let $B(s)$ be a standard Brownian motion, and $\mathbb{D}[0, 1]$ be the space of functions defined on $[0, 1]$, which are right continuous, have left limits and are endowed with Skorokhod topology. Given (3) we can apply directly Donsker's Theorem from Billingsley (1999) to obtain the following result.

Theorem 1. *Suppose that* $\sigma_\gamma^2 := Var(\log \eta_t^2) < \infty$ *and* $\sigma_0^2 > 0$ *a.s. Then, in* $\mathbb{D}[0, 1]$,

$$\begin{aligned} \exp(-s\gamma_0\sqrt{T})\ \sigma_{[sn]}^{2/\sqrt{T}} &\xrightarrow{\mathcal{D}} \exp(\sigma_\gamma B(s)) \\ \exp(-s\gamma_0\sqrt{T})\ |y_{[sn]}|^{2/\sqrt{T}} &\xrightarrow{\mathcal{D}} \exp(\sigma_\gamma B(s)). \end{aligned}$$

Theorem 1 implies that y_t is always nonstationary, which is not the case if the volatility process has a constant term and $\gamma_0 < 0$ (see e.g. Bougerol and Picard (1992) for the GARCH model). Thus, if $\gamma_0 = 0$, then $\log \sigma_t^2$ is a driftless random walk and $|y_t|$ in (1) oscillates randomly between zero and infinity over time. Hence, when $\gamma_0 = 0$, we can say that the process is "stable" (not stationary) in the sense that the observed process has a non-degenerate distribution. If $\gamma_0 \neq 0$, we can say that the process is "unstable", in

the sense that y_t converges to zero or diverges with probability one. See e.g. Hafner and Preminger (2015) and Li et al. (2018).

For example, if $\lambda = 0$ and z_t is standard Gaussian, then the process is stable if $\alpha_0 = \frac{1}{2}\exp(-\Psi(\frac{1}{2})) \approx 3.56$, where $\Psi(\cdot)$ is the Euler psi function. In our case $\lambda \geq 0$, so we consider the set of points

$$\mathcal{S} = \{(\alpha, \lambda) \in \mathbb{R}^2_{++} : \gamma_0(\alpha, \lambda) = 0,\ z_t \sim \mathrm{N}(0,1)\}, \quad \gamma_0 = \mathbb{E}\log(\eta_t).$$

and the critical curve $(\alpha, \lambda) \in \mathcal{S}$ is plotted in Figure 1.

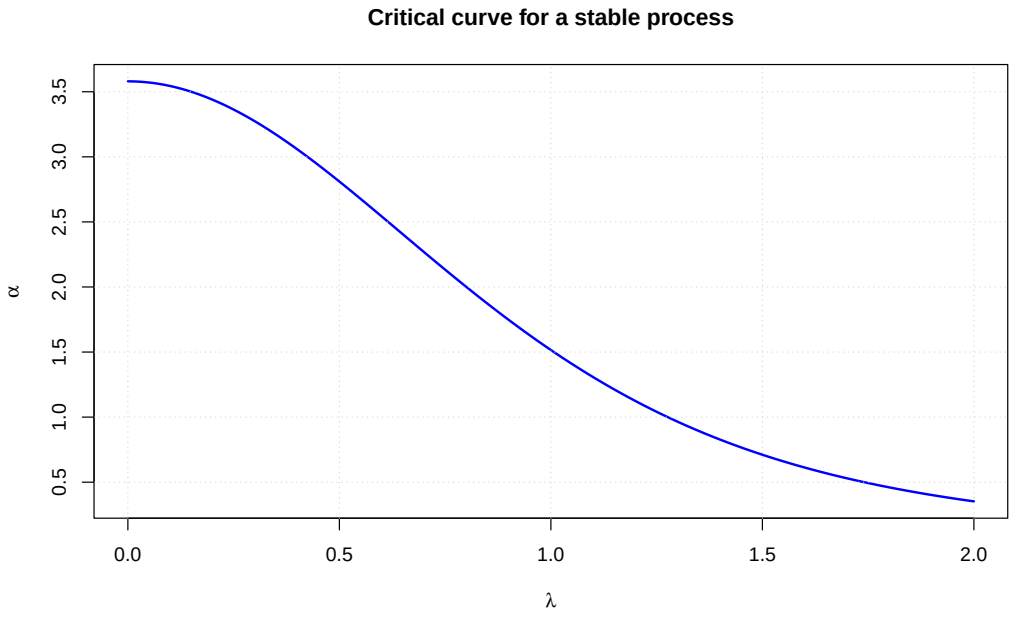


Figure 1: Critical curve of combinations (α, λ) for which the Lyapunov exponent is zero, i.e. the process is stable.

Let $\theta = (\lambda, \alpha)$. The QMLE $\hat{\theta}_T$ of the true parameter θ_0 is defined as the maximizer of a Gaussian quasi log-likelihood function, equivalently $\hat{\theta}_T = \arg\max_{\theta\in\Theta} L_T(\theta)$, where

$$L_T(\theta) = \max_{\theta\in\Theta} \frac{1}{T}\sum_{t=1}^{T} \ell_t(\theta), \quad \ell_t(\theta) = -\frac{1}{2}\left(\log\sigma_t^2 + \frac{(y_t - \lambda\sigma_t)^2}{\sigma_t^2}\right) \tag{5}$$

where $\Theta \subset \mathbb{R}^2$ is a compact set, $\alpha > 0$, $\lambda \geq 0$, and $\sigma_t^2(\theta) = \alpha y_{t-1}^2$, $t = 1, \ldots, T$, with initial condition $y_0^2 > 0$. Note that this is not the true likelihood unless z_t has a Gaussian distribution but allows us to obtain consistent estimators of the true parameters.

We can obtain an analytical solution to the maximization problem. Let $\mu = \lambda\sqrt{\alpha}$ and $\hat{\mu} = \frac{1}{T-1}\sum_{t=2}^{T}\eta_t$. Then the MLEs are given by

$$\begin{aligned}\hat{\alpha} &= \frac{1}{T-1}\sum_{t=2}^{T}\left(\frac{y_t}{|y_{t-1}|} - \hat{\mu}\right)^2, \\ \hat{\lambda} &= \max\left(\hat{\mu}/\sqrt{\hat{\alpha}}, 0\right).\end{aligned}$$

Extensions: Other distributions, such as Student's t-distribution can be employed as they capture the fat tails typical of financial data that uses such models. Furthermore, log-likelihood methods based on t-distributions were considered in the literature (see e.g. Pedersen and Rahbek (2016) and Li and Zhu (2020)) and as such can be considered as an interesting topic for future work.

To obtain the properties of the MLE the following assumptions are made.

Assumption 1. *The random variables $\{z_t\}$ have a non-degenerate distribution with support on at least three points, with $E(z_t) = 0$, $E[z_t^2] = 1$ and $\mathrm{P}(z_t = 0) = 0$.*

Assumption 2. $\mathrm{E}(z_t^4) < \infty$ *and θ_0 belongs to the interior of Θ.*

Theorem 2. *(a) Under Assumption 1, $\hat{\theta}_T \xrightarrow{a.s.} \theta_0$. (b) Further, under Assumptions 1 and 2, $\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{D}} N\big(0, J^{-1}VJ^{-1}\big)$, where*

$$J = -\begin{bmatrix} 1 & \frac{\lambda_0}{2\alpha_0} \\ \frac{\lambda_0}{2\alpha_0} & \frac{\lambda_0^2+2}{4\alpha_0^2} \end{bmatrix}, \quad V = \begin{pmatrix} 1 & \frac{\lambda_0+\kappa_3}{2\alpha_0} \\ \frac{\lambda_0+\kappa_3}{2\alpha_0} & \frac{\lambda_0^2-1+\kappa_4+2\lambda_0\kappa_3}{4\alpha_0^2} \end{pmatrix}$$

and $\kappa_j = \mathrm{E}(z_t^j)$.

Our results are independent of the value of γ_0. Thus, regardless of whether the underlying process is stable or not we will get a strongly consistent estimate with the same variance. Furthermore, if z_t is Gaussian, then $\kappa_3 = 0$, $\kappa_4 = 3$ and the Fisher information matrix equality holds, so $\sqrt{T}(\hat{\theta}_T - \theta_0) \to_D N(0, V^{-1})$. However, this assumption is not required to obtain consistent estimates of the model parameters.

Consistent estimates of $\gamma_0 = \mathbb{E}\log\eta_t^2$ and $\sigma_\gamma^2 = \mathrm{Var}(\log\eta_t^2)$ are given by $\hat{\gamma}_T = T^{-1}\sum_{t=1}^{T}\log\eta_t^2$ and $\hat{\sigma}_\gamma^2 = T^{-1}\sum_{t=1}^{T}(\log\eta_t^2 - \hat{\gamma}_T)^2$ respectively. The following result gives the asymptotic distribution of $\hat{\gamma}_T$.

Theorem 3. *If* $0 < \sigma_\gamma^2 < \infty$ *then,* $\sqrt{T}(\hat{\gamma}_T - \gamma_0) \xrightarrow{\mathcal{D}} N(0, \sigma_\gamma^2)$.

We can use this result to test if $\gamma_0 = 0$, which is equivalent to testing the null hypothesis that the process is stable. Note that we can define the critical curve of α for which $\gamma_0 = 0$ as $\tilde{\alpha}(\lambda) = \exp\left(-\mathbb{E}\log(\lambda + z_t)^2\right)$. This curve may be very different from that shown in Figure 1, which was obtained for normal innovations, if e.g. the innovation distribution has fat tails. For a given λ, a consistent estimator of $\tilde{\alpha}$ is given by $\tilde{\alpha}_T = \exp\left(-\frac{1}{T}\sum_{t=1}^{T}\log(\lambda + z_t)^2\right)$. However, estimation of $\tilde{\alpha}$ is not necessary for the stability test outlined above.

3 Test of the risk premium

Suppose we wish to test that the risk premium parameter is zero, $H_0 : \lambda = 0$. We note that when $\lambda_0 = 0$, the parameter lies on the boundary. In this case, the distribution of the test statistic may not be standard; see e.g., Self and Liang (1987), Andrews (1999, 2001), and for GARCH see Francq and Zakoïan (2007). Let $Z = (Z_\alpha, Z_\lambda)^T$ be a Gaussian random vector with mean zero and covariance matrix $I_\lambda = H_\lambda^{-1} V_\lambda H_\lambda^{-1}$, where V_λ and J_λ are given in Theorem 2 and evaluated at $\lambda = 0$. We can use the likelihood ratio test (LRT) statistic, denoted by $\Lambda_T = 2T(L_T(\hat{\theta}_T) - L_T(\hat{\theta}_{TR}))$, where $\hat{\theta}_{TR}$ is the estimator under the null. The following theorem provides the asymptotic properties of the LRT statistic.

Theorem 4. *Assume Assumptions 1-2 hold, and let* χ_k^2 *denote a chi-squared distributed random variable with* k *degrees of freedom. Then,*

(a) Testing $H_0 : \lambda = 0$*:*

$$\Lambda_T \xrightarrow{\mathcal{D}} \frac{1}{2}\delta_0 + \frac{1}{4}(\kappa_4 - 1)\chi_1^2$$

where δ_0 *is a point mass at zero.*

(b) Testing $H_0 : \lambda = 0, \alpha = \alpha_0$*:*

$$\Lambda_T \xrightarrow{\mathcal{D}} \frac{1}{2}\chi_1^2 + \frac{1}{2}(Z_\alpha^2 + \frac{1}{2\alpha_0^2}Z_\lambda^2)$$

where $Z = (Z_\alpha, Z_\lambda)^\intercal \sim N(0, I_\lambda)$, $I_\lambda = \begin{pmatrix} 1 & \alpha_0\kappa_3 \\ \alpha_0\kappa_3 & \alpha_0^2(\kappa_4 - 1) \end{pmatrix}$.

Under Gaussian innovations:

$$\Lambda_T \xrightarrow{\mathcal{D}} \frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$$

We note that under the null of no risk premia, the model reduces to an ARCH(1) without intercept and we can use part (a) to test it. Under Gaussian innovations, it is clear that $\Lambda_T \xrightarrow{\mathcal{D}} \frac{1}{2}\delta_0 + \frac{1}{2}\chi_2^2$. Furthermore, part (b) deals with joint hypothesis regarding also the value of α_0. For instance, given Figure 1, it would be interesting to test $H_0 : \lambda = 0, \alpha_0 = 3.56$, which implies that under the null, the true model is ARCH(1) with no intercept and Gaussian innovations. We will consider the test of this hypothesis in the simulation study of the next section.

Note that under misspecification, the QMLE remains consistent for the true values, allowing us to obtain reliable estimates for potential fat tailed distributions of the innovations (e.g., t-distribution) and refine our null hypothesis by choosing α_0 for which (4) is zero and the model is stable. This is left for future work.

4 Simulation Study

We conduct two simulation studies to show the empirical performance of the QMLE and the stability test in finite samples. First, Table 1 shows the estimation results of $K = 100,000$ independent Monte Carlo replications. We choose four different levels of the risk premium parameter, $\lambda \in \{0.1, 0.5, 1, 2\}$, and then determine α such that the process is stable, see Figure 1. We report the means and standard deviations of QMLE across the replications for two different sample sizes, $T = 100$ and $T = 1000$. For all considered scenarios, we see that already at $T = 100$ parameter estimates are close to the true values. Moreover, increasing the sample size from $T = 100$ to $T = 1000$, standard deviations decrease roughly with a factor of $10^{-1/2}$, as expected.

For each scenario, we analyze the performance of the likelihood ratio tests of Section 3 to test the hypotheses $H_0 : \lambda = 0$ and $H_0 : \alpha = 3.56, \lambda = 0$, respectively. The results are reported in the rows LRT1 and LRT2 of Table 1. We see that the empirical size of the tests is very close to the nominal level of 5% already for $T = 100$, while the empirical power increases quickly with increasing λ and sample size T. The power of the composite

test is slightly lower than the power of the simple test, because we simulate processes using Gaussian innovations, so that the difference with respect to the null is only due to λ, not α.

To show that our theory holds also for the case of unstable processes with either positive or negative Lyapunov exponents, we produce analogous simulation results for these cases in Tables 2 and 3, using the same values of λ as in Table 1, but with α either smaller (Table 2) or larger (Table 3) than the critical α, i.e. the value for which $\gamma_0 = 0$. The difference to the critical α is fixed at 0.5, except for the last scenario with $\lambda = 2$ for which it is fixed at 0.2. We see that the estimation results for α and λ are very similar to those in the stable case. Also, the test results of $H_0 : \lambda = 0$ using LRT1 are largely the same, but for testing $H_0 : \alpha = 3.56, \lambda = 0$, LRT2 has obviously more power now because both parts of the hypothesis are violated, not only the first part.

Furthermore, we provide simulation results for the test of $H_0 : \gamma_0 = 0$ in terms of power and size, for the two sample sizes, $T = 100$ and $T = 1000$, see Table 4. We again choose $\lambda \in \{0.1, 0.5, 1, 2\}$. For the size properties, we determine α as above to obtain a stable process, let's call these α_0. For the power properties, we consider values $\alpha_0 \pm 0.25$ and $\alpha_0 \pm 0.5$, except for the case $\lambda = 2$ where we choose $\alpha_0 \pm 0.1$ and $\alpha_0 \pm 0.2$. This gives us the parameter values in Table 2. The numbers in the table are the proportions of rejection of H_0 for $K = 100,000$ independent Monte Carlo replications at a nominal level of 5%. We see that for $T = 100$ is slightly oversized, but at $T = 1000$, the empirical size is very close to the nominal size. In all cases, we see that the power increases as we move away from the stability case, and that it increases substantially when increasing the sample size from $T = 100$ to $T = 1000$. The test therefore shows a good empirical performance.

5 Empirical application

We fit the ARCH-in-mean model without intercept to the monthly stock index FTSE 100, from January 1984 to October 2025, 502 observations. The upper panel of Figure 2 depicts the FTSE 100 index time series.

Summary statistics of the returns of this index are given in Table 3, showing a negative

	$T = 100$		$T = 1000$		$T = 100$		$T = 1000$	
	$\lambda = 0$, $\alpha = 3.5621$							
LRT1	0.0516		0.0499					
LRT2	0.0505		0.0496					
	$\lambda = 0.1$, $\alpha = 3.5442$				$\lambda = 1$, $\alpha = 1.5156$			
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
$\hat{\alpha}$	3.5444	0.5037	3.5416	0.1591	1.5171	0.2185	1.5167	0.0672
$\hat{\lambda}$	0.0985	0.1018	0.1002	0.0319	1.0062	0.1245	1.0005	0.0391
LRT1	0.2569		0.9344		1.0000		1.0000	
LRT2	0.1779		0.8610		1.0000		1.0000	
	$\lambda = 0.5$, $\alpha = 2.8102$				$\lambda = 2$, $\alpha = 0.3525$			
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
$\hat{\alpha}$	2.8098	0.3991	2.8119	0.1263	0.3517	0.0497	0.3524	0.0158
$\hat{\lambda}$	0.5037	0.1088	0.5001	0.0337	2.0175	0.1773	2.0016	0.0549
LRT1	0.9995		1.0000		1.0000		1.0000	
LRT2	0.9981		1.0000		1.0000		1.0000	

Table 1: QMLE estimation results for the stable ARCH-in-mean model with parameters λ and α. Reported are the means and standard deviations of parameter estimates for $K = 100,000$ independent Monte Carlo replications. LRT1 and LRT2 give the proportion of rejections of $H_0 : \lambda = 0$ and $H_0 : \alpha = 3.56, \lambda = 0$, respectively, using the likelihood ratio test and nominal level 5%.

	$T = 100$		$T = 1000$		$T = 100$		$T = 1000$	
	$\lambda = 0$, $\alpha = 3.0621$							
LRT1	0.0511		0.0502					
LRT2	0.1476		0.8828					
	$\lambda = 0.1$, $\alpha = 3.0442$				$\lambda = 1$, $\alpha = 1.0156$			
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
$\hat{\alpha}$	3.0452	0.4352	3.0437	0.1361	1.0161	0.1450	1.0154	0.0452
$\hat{\lambda}$	0.1089	0.0879	0.1001	0.0317	1.0072	0.1245	1.0007	0.0386
LRT1	0.2574		0.9346		1.0000		1.0000	
LRT2	0.2827		0.9949		1.0000		1.0000	
	$\lambda = 0.5$, $\alpha = 2.3102$				$\lambda = 2$, $\alpha = 0.1525$			
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
$\hat{\alpha}$	2.3104	0.3305	2.3104	0.1034	0.1524	0.0218	0.1524	0.0096
$\hat{\lambda}$	0.5035	0.1079	0.5003	0.0335	2.0165	0.1778	2.0032	0.0579
LRT1	0.9994		1.0000		1.0000		1.0000	
LRT2	0.9995		1.0000		1.0000		1.0000	

Table 2: QMLE estimation results for the ARCH-in-mean model. The parameter α is chosen such that the Lyapunov exponent γ_0 is negative, i.e. the ARCH-in-mean model is unstable. Other comments as in the caption of Table 1.

	$T = 100$		$T = 1000$		$T = 100$		$T = 1000$	
	$\lambda = 0$, $\alpha = 4.0621$							
LRT1	0.0522		0.0499					
LRT2	0.1368		0.7869					
	$\lambda = 0.1$, $\alpha = 4.0442$				$\lambda = 1$, $\alpha = 2.0156$			
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
$\hat{\alpha}$	4.0454	0.5791	4.0444	0.1812	2.0142	0.2883	2.0156	0.0901
$\hat{\lambda}$	0.1093	0.0883	0.1001	0.0317	1.0078	0.1250	1.0006	0.0386
LRT1	0.2589		0.9360		1.0000		1.0000	
LRT2	0.2780		0.9866		1.0000		1.0000	
	$\lambda = 0.5$, $\alpha = 3.3102$				$\lambda = 2$, $\alpha = 0.5525$			
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
$\hat{\alpha}$	3.3118	0.4721	3.3100	0.1482	0.5526	0.0786	0.5524	0.0246
$\hat{\lambda}$	0.5039	0.1079	0.5003	0.0335	2.0152	0.1770	2.0016	0.0547
LRT1	0.9996		1.0000		1.0000		1.0000	
LRT2	0.9973		1.0000		1.0000		1.0000	

Table 3: QMLE estimation results for the ARCH-in-mean model. The parameter α is chosen such that the Lyapunov exponent γ_0 is positive, i.e. the ARCH-in-mean model is unstable. Other comments as in the caption of Table 1.

$\lambda = 0.1$			$\lambda = 1$		
α	$T = 100$	$T = 1000$	α	$T = 100$	$T = 1000$
3.0442	0.0828	0.5517	1.0156	0.4661	1.0000
3.2942	0.0555	0.1530	1.2656	0.1115	0.7813
3.5442	**0.0592**	**0.0533**	**1.5156**	**0.0587**	**0.0512**
3.7942	0.0866	0.1918	1.7656	0.1511	0.6186
4.0442	0.1298	0.5004	2.0156	0.3129	0.9833
$\lambda = 0.5$			$\lambda = 2$		
2.3102	0.1176	0.7970	0.1525	1.0000	1.0000
2.5602	0.0595	0.2428	0.2525	0.6427	1.0000
2.8102	**0.0599**	**0.0506**	**0.3525**	**0.0641**	**0.0505**
3.0602	0.0940	0.2539	0.4525	0.4343	0.9986
3.3102	0.1582	0.6584	0.5525	0.8041	1.0000

Table 4: Monte Carlo results for the test of the stability hypothesis $H_0 : \gamma_0 = 0$. The numbers are the proportions of rejection of H_0 for $K = 100,000$ independent replications at a nominal level of 5%. The lines in bold correspond to the empirical size of the test, the other lines to the power properties.

Mean	0.004401	$\rho(1)$	0.004	$\rho_2(1)$	0.121
Std.dev.	0.043052	$\rho(2)$	-0.074	$\rho_2(2)$	0.054
Skewness	-1.082147	$\rho(3)$	-0.012	$\rho_2(3)$	0.029
Kurtosis	8.187252	$\rho(4)$	0.062	$\rho_2(4)$	0.008

Table 5: *Summary statistics for FTSE100 monthly returns.* $\rho(k)$ *is the autocorrelation of returns of order* k, $\rho_2(k)$ *is the autocorrelation of squared returns of order* k.

skewness due to some large negative returns, and leptokurtosis. As expected, autocorrelations of returns are negligible, while the first order autocorrelation of squared returns is significant. Hence, ARCH models of low order appear appropriate to model volatility clustering. We estimate an ARCH(1) model both with and without risk premium, and report the estimation results in Table 4.

The ARCH-in-mean model passes a portmanteau specification test for the squared standardized residuals. The statistic is insignificant up to a lag order of twenty, with a p-value of 0.92. The critical curve $\tilde{\alpha}(\lambda)$, evaluated at $\hat{\lambda}$, for which $\gamma_0 = 0$ is estimated as $\tilde{\alpha}_T = 4.7668$, while the estimate of γ_0 is $\hat{\gamma} = 0.0013$ with a standard error of 0.1318.

Parameter	ARCH(1)	ARCH(1)-M
α	4.8979 (0.64)	4.7734 (0.52)
λ		0.1240 (0.04)
Lik	875.92	879.74

Table 6: *Parameter estimates of the ARCH(1) model with and without risk premium, and corresponding values of the log likelihood function. Standard errors are in parentheses.*

Thus, the stability test proposed in Section 2 yields a test statistic of 0.01051, and the null of $\gamma_0 = 0$ is clearly not rejected. Hence, the conclusion is that there is no statistical evidence against the hypothesis that log volatility of this monthly stock index follows a random walk.

The likelihood ratio statistic to test $H_0 : \lambda = 0$ takes the value $\Lambda_T = 7.64$, which is significant at level 5% with critical value 2.70. We therefore reject the hypothesis of a model without risk premium. The composite hypothesis $H_0 : \lambda = 0, \alpha = 3.56$ is rejected even stronger, as the critical α is much higher than the value under H_0. The test statistic takes the value 17.24, larger than the critical value of 5.14.

The center and lower panels of Figure 2 show the estimated conditional mean and the log conditional standard deviation of the stock index. In periods of high volatility such as around the stock market crash of 1987, the financial crisis of 2008 and the Covid crisis of 2020, the estimated risk premium rises to 3-4%, and even to 8% for the 1987 crash. For log volatility, as there is no lower bound, it shows presence of negative outliers, which are even more frequent than positive ones.

6 Conclusion

We have proposed a new ARCH-in-mean model without intercept, which is characterized by non-stationarity but possibly stable trajectories. We derived the asymptotic theory of the QMLE and developped a test for stability as well as for the risk premium. In a simulation study, we showed the good finite sample performance of the estimator and the test. Finally, we applied the model to a monthly stock index series.

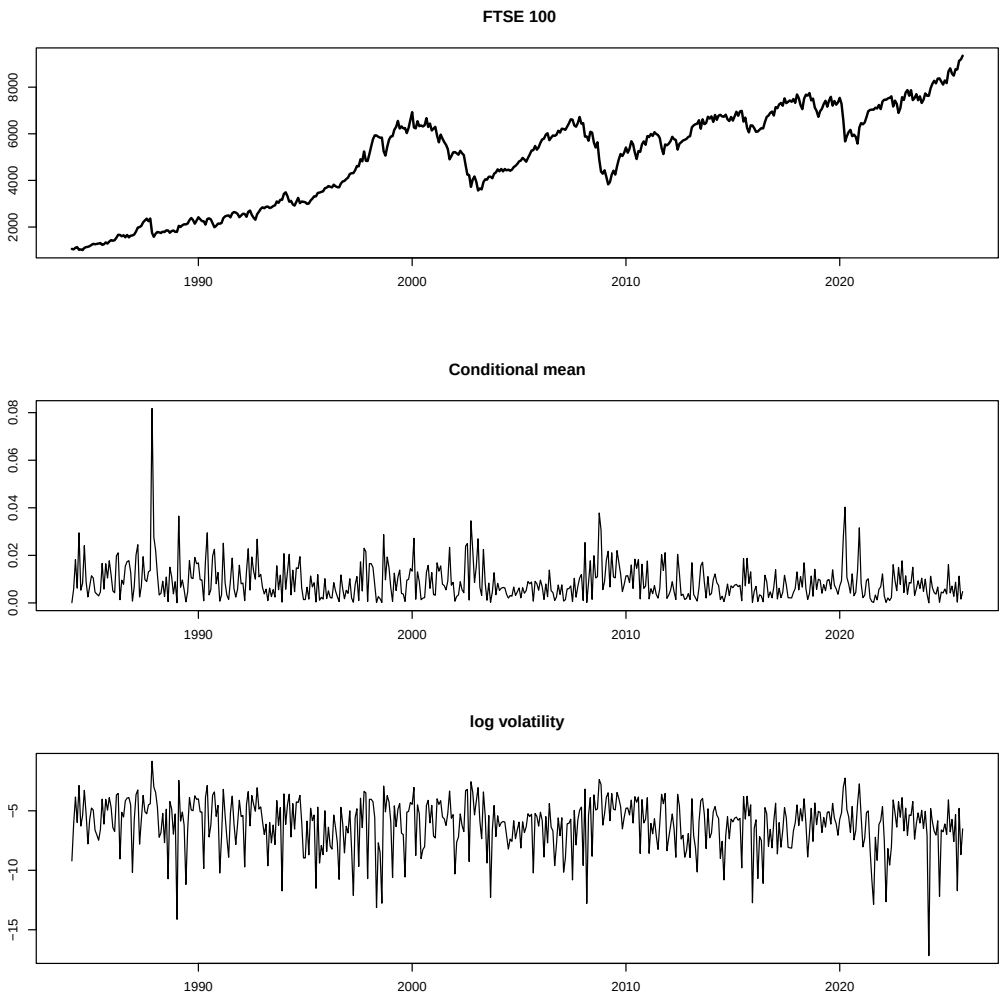


Figure 2: *Upper panel: Monthly series of the FTSE 100 from 1984 to 2025. Middle panel: Estimated conditional mean of returns. Lower panel: Estimated log volatility of returns.*

Further extensions are possible, e.g. by considering fat-tailed distributions for the innovation term. To increase the memory in volatility, one might add an autoregressive term in volatility as in Li et al (2018), or embed the model into a multivariate system with one lag to obtain long memory in the individual components, as in Bauwens et al (2023).

Appendix

Proof of Theorem 1

The result follows directly from Donsker's theorem.

Proof of Theorem 2

(a) From (2) and (5) we have

$$\ell_t(\theta) = -\frac{1}{2}\left(\log\sigma_t^2 + \frac{(y_t - \lambda\sigma_t)^2}{\sigma_t^2}\right) = -\frac{1}{2}\left(\log\alpha y_{t-1}^2 + \frac{(\eta_t - \lambda\sqrt{\alpha})^2}{\alpha}\right) \tag{6}$$

where

$$\eta_t = \frac{y_t}{|y_{t-1}|} = \lambda_0\sqrt{\alpha_0} + \sqrt{\alpha_0}z_t \tag{7}$$

is an i.i.d. random variable.

We first establish consistency. Note that $\hat{\theta}_T = \arg\min_{\theta\in\Theta} Q_T(\theta)$ where $Q_T(\theta) = -\frac{2}{T}\sum_{t=1}^{T}(\ell_t(\theta) - \ell_t(\theta_0))$. From (6)–(7) we get

$$\begin{aligned} Q_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\log\frac{\alpha}{\alpha_0} - z_t^2 + \frac{(\eta_t - \lambda_0\sqrt{\alpha_0} + \lambda_0\sqrt{\alpha_0} - \lambda\sqrt{\alpha})^2}{\alpha} \qquad (8)\\ &= \frac{1}{T}\sum_{t=1}^{T}\log\frac{\alpha}{\alpha_0} + \left(\frac{\alpha_0}{\alpha} - 1\right)z_t^2 + 2\sqrt{\alpha^{-2}\alpha_0}\left(\lambda_0\sqrt{\alpha_0} - \lambda\sqrt{\alpha}\right)z_t + \frac{(\lambda_0\sqrt{\alpha_0} - \lambda\sqrt{\alpha})^2}{\alpha} \end{aligned}$$

Rearranging terms yields

$$\begin{aligned} Q_T(\theta) &= \left(\frac{\alpha_0}{\alpha} - 1\right) - \log\frac{\alpha_0}{\alpha} + \frac{(\lambda\sqrt{\alpha} - \lambda_0\sqrt{\alpha_0})^2}{\alpha} \qquad (9)\\ &\quad + \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\alpha_0}{\alpha} - 1\right)(z_t^2 - 1) + 2\sqrt{\alpha^{-2}\alpha_0}\left(\lambda_0\sqrt{\alpha_0} - \lambda\sqrt{\alpha}\right)z_t \\ &= Q(\theta) + \frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{\alpha_0}{\alpha} - 1\right)(z_t^2 - 1) + 2\sqrt{\alpha^{-2}\alpha_0}\left(\lambda_0\sqrt{\alpha_0} - \lambda\sqrt{\alpha}\right)z_t\right] \end{aligned}$$

where $Q(\theta) = \left(\frac{\alpha_0}{\alpha} - 1 - \log\frac{\alpha_0}{\alpha}\right) + \left(\lambda - \lambda_0\sqrt{\frac{\alpha_0}{\alpha}}\right)^2$.

Set $x = \alpha_0/\alpha$ and note that $Q(\theta) \geq 0$ since the first term is non-negative as $\log x \leq x - 1$ for all $x > 0$ and equals zero iff $x = 1$. The first term in $Q(\theta)$ is zero if $x = 1$, i.e. $\alpha = \alpha_0$ which implies that the second term $(\lambda - \lambda_0)^2$ is zero iff $\lambda = \lambda_0$, so unless $\theta = \theta_0$, we have $Q(\theta) > 0$.

Next, we note that by definition $Q_T(\hat{\theta}_T) \leq Q_T(\theta_0)$ and $Q(\theta_0) \leq Q(\hat{\theta}_T)$. Using these inequalities we have

$$Q_T(\hat{\theta}_T) - Q(\hat{\theta}_T) \leq Q_T(\hat{\theta}_T) - Q(\theta_0) \leq Q_T(\theta_0) - Q(\theta_0) \tag{10}$$

Therefore,

$$\begin{aligned}
\left|Q_T(\hat{\theta}_T) - Q(\theta_0)\right| &\leq \max\left(Q_T(\hat{\theta}_T) - Q(\hat{\theta}_T), Q_T(\theta_0) - Q(\theta_0)\right) \\
&\leq \sup_{\theta\in\Theta} |Q_T(\theta) - Q(\theta)| \leq \sup_{\theta\in\Theta} \left|\frac{\alpha_0}{\alpha} - 1\right| \frac{1}{T}\sum_{i=1}^{T}(z_i^2 - 1) \\
&\quad + \sup_{\theta\in\Theta} \left|2\sqrt{\alpha^{-2}\alpha_0}(\lambda_0\sqrt{\alpha_0} - \lambda\sqrt{\alpha})\right| \frac{1}{T}\sum_{i=1}^{T} z_i \xrightarrow{a.s.} 0
\end{aligned}$$

as $T \to \infty$ by Assumption 1 and the strong law of large numbers. Now, since $Q(\theta)$ has a unique minimum, $\left|Q_T(\hat{\theta}_T) - Q(\theta_0)\right| \xrightarrow{a.s.} 0$ implies that $\hat{\theta}_T \xrightarrow{a.s.} \theta_0$.

(b) We proceed by showing the asymptotic normality for the MLE. Let $s_t(\theta) = \frac{\partial \ell_t(\theta)}{\partial \theta}$ and $H_t(\theta) = \frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'}$ and denote $V = \mathbb{E}\left(\frac{\partial \ell_t(\theta_0)}{\partial \theta}\frac{\partial \ell_t(\theta_0)}{\partial \theta'}\right)$, $J = \mathbb{E}\left(\frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'}\right)$. We split the proof into several intermediate steps.

(i) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_t(\theta_0) \xrightarrow{\mathcal{D}} N(0, V)$ where

$$V = \begin{pmatrix} 1 & \frac{\lambda_0+\kappa_3}{2\alpha_0} \\ \frac{\lambda_0+\kappa_3}{2\alpha_0} & \frac{\lambda_0^2-1+\kappa_4+2\lambda_0\kappa_3}{4\alpha_0^2} \end{pmatrix}$$

is a positive definite matrix.

(ii) If $\bar{\theta}_T \xrightarrow{a.s.} \theta_0$ then $T^{-1}\sum_{t=1}^{T} H_t(\bar{\theta}_T) \xrightarrow{a.s.} J$ where

$$J = -\begin{bmatrix} 1 & \frac{\lambda_0}{2\alpha_0} \\ \frac{\lambda_0}{2\alpha_0} & \frac{\lambda_0^2+2}{4\alpha_0^2} \end{bmatrix}$$

is a negative definite matrix.

To prove (i) we first calculate $s_t(\theta)$ from (6), thus

$$\frac{\partial \ell_t}{\partial \alpha} = -\frac{1}{2\alpha} + \frac{\lambda}{2\alpha^{3/2}}(\eta_t - \lambda\sqrt{\alpha}) + \frac{1}{2\alpha^2}(\eta_t - \lambda\sqrt{\alpha})^2 \tag{12}$$

$$\frac{\partial \ell_t}{\partial \lambda} = \frac{-2\sqrt{\alpha}(\eta_t - \lambda\sqrt{\alpha})}{2\alpha} = \frac{\eta_t - \lambda\sqrt{\alpha}}{\sqrt{\alpha}} \tag{13}$$

Recall from (7) that $z_t = (\eta_t - \lambda_0\sqrt{\alpha_0})/\sqrt{\alpha_0}$ so by (12)–(13), we have $s_t(\theta_0) = \left(z_t, \frac{(\lambda_0 z_t + (z_t^2 - 1))}{2\alpha_0}\right)$. Next, using the triangle inequality, we get

$$\mathbb{E}\left\|s_t(\theta_0)\right\|^2 \leq C_1 + C_2\mathbb{E}\left|z_t^2\right| + C_3\,\mathbb{E}\left|z_t^4\right| < \infty \tag{14}$$

where the last inequality follows by Assumption 2.

$$\mathrm{V} = \mathbb{E}(s_t(\theta_0)s_t(\theta_0)') = \begin{pmatrix} 1 & \frac{\lambda_0+\kappa_3}{2\alpha_0} \\ \frac{\lambda_0+\kappa_3}{2\alpha_0} & \frac{\lambda_0^2-1+\kappa_4+2\lambda_0\kappa_3}{4\alpha_0^2} \end{pmatrix} \tag{15}$$

where $\kappa_j = \mathbb{E}(z_t^j)$.

Now suppose $c^\intercal \mathrm{V} c = \mathbb{E}(c^\intercal s_t(\theta_0))^2 = 0$ for some $c = (c_1, c_2)^\intercal$ hence $c^\intercal s_t(\theta_0) = 0$ almost surely which implies $z_t(c_2 + \frac{c_1\lambda_0}{2\alpha_0}) + \frac{c_1 z_t^2}{2\alpha_0} - \frac{c_1}{2\alpha_0} = 0$. This contradicts Assumption 1 as z_t can take on at least three distinct values with a non-zero probability. Hence V is positive definite. Using the arguments above, (14), and that $s_t(\theta_0)$ are i.i.d. with mean zero, we can apply the Lindeberg-Lévy central limit theorem to get

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_t(\theta_0) \xrightarrow{\mathcal{D}} N(0, V) \tag{16}$$

Next, we show (ii). It follows from (12)–(13) that the elements of $H_t(\theta)$ are given by

$$\begin{aligned} \frac{\partial^2 \ell_t}{\partial \alpha^2}(\theta) &= \frac{2-\lambda^2}{4\alpha^2} - \frac{5\lambda(\eta_t - \lambda\sqrt{\alpha})}{4\alpha^{5/2}} - \frac{(\eta_t - \lambda\sqrt{\alpha})^2}{\alpha^3} \\ \frac{\partial^2 \ell_t}{\partial \alpha \partial \lambda}(\theta) &= \frac{-2\eta_t}{2\alpha^{3/2}} \quad \text{and} \quad \frac{\partial^2 \ell_t}{\partial \lambda^2} = -1 \end{aligned} \tag{17}$$

Recall from (7) that $\eta_t = \sqrt{\alpha_0}(\lambda_0 + z_t)$ so from Assumption 2 and the c_r inequality $(E|X+Y|^r \leq 2^{r-1}(E|X|^r + E|Y|^r), r \geq 1)$ we have

$$\mathbb{E}(|\eta_t|^p) \leq \alpha_0^{p/2}\mathbb{E}|\lambda_0 + z_t|^p = C_1 + C_2\mathbb{E}|z_t|^p < \infty \text{ for } p = 1, 2. \tag{18}$$

Hence, by (18) and applying the c_r inequality again we obtain

$$\begin{aligned} \mathbb{E}\sup_{\theta\in\Theta}\left|\frac{\partial^2 \ell_t}{\partial \alpha^2}(\theta)\right| &= C_1 + C_2\mathbb{E}(|\eta_t|) + C_3\mathbb{E}(|\eta_t^2|) < \infty \\ \mathbb{E}\sup_{\theta\in\Theta}\left|\frac{\partial^2 \ell_t}{\partial \alpha \partial \lambda}(\theta)\right| &= C_1\mathbb{E}(|\eta_t|) < \infty \end{aligned} \tag{19}$$

Since η_t is i.i.d., we can use (19) to apply the uniform law of large numbers (White (1996, pp. 351–352)) to get that

$$\sup_{\theta\in\Theta}\left\|T^{-1}\sum_{t=1}^{T}H_t(\theta)-\mathbb{E}H_t(\theta)\right\|\xrightarrow{a.s.}0 \tag{20}$$

Hence, if $\bar{\theta}_T\xrightarrow{a.s.}\theta_0$ we have by the triangle inequality

$$\begin{aligned}\left\|T^{-1}\sum_{t=1}^{T}H_t(\bar{\theta}_T)-\mathbb{E}(H_t(\theta_0))\right\| &\leq \left\|T^{-1}\sum_{t=1}^{T}H_t(\bar{\theta}_T)-\mathbb{E}H_t(\bar{\theta}_T)\right\| \\ &\quad+\left\|\mathbb{E}H_t(\bar{\theta}_T)-\mathbb{E}(H_t(\theta_0))\right\| \\ &\leq \sup_{\theta\in\Theta}\left\|T^{-1}\sum_{t=1}^{T}H_t(\theta)-\mathbb{E}H_t(\theta)\right\|+o_{a.s.}(1)=o_{a.s.}(1)\end{aligned}$$

The first term after the second inequality follows from (20), and the second term follows from the continuity of $\mathbb{E}H_t(\theta)$. We further note that for $\theta=\theta_0$

$$J=\mathbb{E}(H_t(\theta_0))=-\begin{bmatrix}1 & \frac{\lambda_0}{2\alpha_0}\\ \frac{\lambda_0}{2\alpha_0} & \frac{\lambda_0^2+2}{4\alpha_0^2}\end{bmatrix}$$

is invertible because it is negative definite as

$$\det(J)=-\frac{1}{2\alpha_0^2}<0. \tag{21}$$

Lastly, by the strong consistency shown in the first part of the theorem, we have for sufficiently large T, that $\hat{\theta}_T$ is contained a.s. in an arbitrary small neighborhood of θ_0. Hence, the mean-value theorem expansion of the score function around θ_0 gives

$$\sqrt{T}(\hat{\theta}_T-\theta_0)=\left[\frac{1}{T}\sum_{t=1}^{T}H_t(\bar{\theta}_T)\right]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(\theta_0\right) \tag{22}$$

where $\overline{\theta}_T$ is between $\hat{\theta}_T$ and θ_0.

In (i) we show that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(\theta_0\right)$ obeys the CLT and in (ii) we show that the terms inside the square brackets converge to J, a non-singular matrix. The desired result follows from the continuity of the inverse operator and Slutsky's lemma.

Proof of Theorem 3

The result follows directly from the Lindeberg-Lévy central limit theorem.

Proof of Theorem 4

The log-likelihood function is three times continuously differentiable and from (7) and (17), and using similar arguments to Theorem 2, we can further show that the third derivative of the log likelihood function is bounded over Θ. Thus, all the conditions of Theorem 3 of Self and Liang (1987) are satisfied and the LRT statistic converges in distribution to

$$\Lambda_\tau \xrightarrow{\mathcal{D}} \inf_{\psi \in C_0} (Z-\psi)^\intercal V_\lambda (Z-\psi) - \inf_{\psi \in C_1} (Z-\psi)^\intercal V_\lambda (Z-\psi) \tag{23}$$

where $\psi = (\psi_\lambda, \psi_\alpha)^\intercal$, $Z = (Z_\lambda, Z_\alpha)^\intercal$, and C_0, C_1 are the cones under the null and the alternative hypothesis, respectively. Also $Z \sim N(0, I_\lambda)$, $I_\lambda = J_\lambda^{-1} V_\lambda J_\lambda^{-1}$ where V_λ and J_λ are the covariance matrix of the score function and the expected Hessian evaluated at $\lambda = 0$, respectively. It is straightforward to show that $I_\lambda = \begin{pmatrix} 1 & \alpha_0\kappa_3 \\ \alpha_0\kappa_3 & \alpha_0^2(\kappa_4 - 1) \end{pmatrix}$.

Because $V_\lambda = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2\alpha_0^2} \end{pmatrix}$, the quadratic form in (23) separates:

$$(Z_\alpha - \psi_\alpha)^2 + \frac{1}{2\alpha_0^2}(Z_\lambda - \psi_\lambda)^2 \tag{24}$$

(a) Consider $H_0 : \lambda = 0$. The cones are $C_0 = \mathbb{R} \times \{0\}$ and $C_1 = \mathbb{R} \times [0, \infty)$. Minimizing (24) over C_0 ($\psi_\lambda = 0$) gives $\psi_\alpha = Z_\alpha$, yielding $\inf_{C_0} = \frac{1}{2\alpha_0^2} Z_\lambda^2$ and over C_1 ($\psi_\lambda \geq 0$) gives $\psi_\alpha = Z_\alpha$ and $\psi_\lambda = \max(0, Z_\lambda)$, hence $\inf_{C_1} = \frac{1}{2\alpha_0^2}(Z_\lambda - \max(0, Z_\lambda))^2$. Thus the LRT statistic converges in distribution to:

$$\Lambda_T \xrightarrow{\mathcal{D}} \begin{cases} \frac{1}{2\alpha_0^2} Z_\lambda^2 & Z_\lambda \geq 0 \\ 0 & Z_\lambda < 0 \end{cases}$$

Given $Z_\lambda \sim N(0, (\kappa_4 - 1)\alpha_0^2)$, we have $\frac{Z_\lambda^2}{2\alpha_0^2} \sim \frac{1}{2}(\kappa_4 - 1) \cdot \chi_1^2$ and because $P(Z_\lambda \geq 0) = \frac{1}{2}$, the desired result follows.

(b) For the joint hypothesis, $H_0 : \lambda = 0$, $\alpha = \alpha_0$, the cones are $C_0 = \{(0, 0)\}$ and $C_1 = \mathbb{R} \times [0, \infty)$ and similarly to part (a) the infima difference of (23) becomes

$$\Lambda_T \xrightarrow{\mathcal{D}} \begin{cases} Z_\alpha^2 + \frac{1}{2\alpha_0^2} Z_\lambda^2 & Z_\lambda \geq 0 \\ Z_\alpha^2 & Z_\lambda < 0 \end{cases}$$

Again, $P(Z_\lambda \geq 0) = \frac{1}{2}$ and noting that $Z_\alpha^2 \sim \chi_1^2$ (since $Var(Z_\alpha) = 1$) gives the desired result for the LRT statistics. Under Gaussian innovations, $Z_\alpha \sim N(0,1)$, $Z_\lambda \sim N(0, 2\alpha_0^2)$ and $Z_\alpha \perp Z_\lambda$, so $Z_\alpha^2 + \frac{1}{2\alpha_0^2} Z_\lambda^2 \sim \chi_2^2$, which completes the proof.

References

Bauwens, L, Chevillon, G. and Laurent, S. (2023), We modeled long memory with just one lag!, *Journal of Econometrics*, 236, 105467.

Bauwens, L., Hafner, C.M. and Laurent, S. (2012), *Handbook of Volatility Models and Their Applications*, Wiley.

Billingsley, P. (1999): *Convergence of Probability Measures*. Wiley, Second edition.

Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307-327.

Bougerol, P. and Picard, N. (1992), Stationarity of GARCH process and some nonnegative time series, *Journal of Econometrics*, 52, 115–127.

Creal, D., Koopman, S.J. and Lucas, A. (2013), Generalized autoregressive score models with applications, *Journal of Applied Econometrics*, 28, 777-795.

Christensen, B.J., Christian M. Dahl, C.M. and Iglesias, E.M. (2012), Semiparametric inference in a GARCH-in-mean model, *Journal of Econometrics*, 167, 458-172.

Conrad, C. and Mammen, E. (2016), Asymptotics for parametric GARCH-in-Mean models, *Journal of Econometrics*, 194, 319-329.

Engle, R. F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 50, 987–1007.

Engle, R. F., Lilien, D. M., & Robins, R. P. (1987), Estimating Time Varying Risk Premia in the Term Structure: The Arch-M Model, *Econometrica*, 55, 391–407.

Francq, C. and Zakoïan, J.-M. (2012), Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroskedasticity models, *Econometrica*, 80, 821-861.

Francq, C. and Zakoïan, J.-M. (2019) *GARCH Models: Structure, Statistical Inference and Financial Applications*, 2nd ed., Wiley.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2005), There is a risk-return trade-off after all, *Journal of Financial Economics*. 76, 509-548.

Hafner, C. M. and Kyriakopoulou, D. (2021), Exponential-Type GARCH Models With Linear-in-Variance Risk Premium, *Journal of Business & Economic Statistics*, 39, 589–603.

Hafner, C. M. and Preminger, A. (2015), An ARCH model without intercept, *Economics Letters*, 129, 13-17.

Harvey, A.C. (2013), Dynamic models for volatility and heavy tails: with applications to financial and economic time series, Cambridge University Press.

Harvey, A.C. and Chakravarty, T. (2008), Beta-t-(E)GARCH, Cambridge University working paper.

Li, D., Zhang, X., Zhu, K. and Ling, S. (2018), The ZD-GARCH model: A new way to study heteroscedasticity, *Journal of Econometrics*, 202, 1–17.

Li, D. and Zhu, K. (2020), Inference for asymmetric exponentially weighted moving average models, *Journal of Time Series Analysis*, 41, 154–162.

Ling, S. (2004), Estimation and Testing Stationarity for Double-Autoregressive Models, *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 66, 63–78.

Lintner, J. (1965), The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets, *The Review of Economics and Statistics*, 47, 13–37.

Markowitz, H. (1952), Portfolio Selection, *The Journal of Finance*, 7, 77–91.

Merton, R. C. (1973), An Intertemporal Capital Asset Pricing Model, *Econometrica*, 41, 867–887.

Nelson, D. B. (1990), Stationarity and Persistence in the GARCH$(1,1)$ Model, *Econometric Theory*, 6, 318–334.

Pedersen, R.S. and Rahbek, A. (2016), Nonstationary GARCH with t-distributed innovations, *Economics Letters*, 138, 19–21.

Sharpe, W.F. (1964), Capital asset prices: A theory of market equilibrium under conditions of risk, *The Journal of Finance*, 19, 425-442.

White, H. (1996), *Estimation, Inference and Specification Analysis*. First edition, Cambridge University Press.