

Herwartz, Helmut

Conference Paper

Are GARCH innovations independent - a long term assessment for the S&P 500

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Financial Econometrics, No. B22-V3

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Herwartz, Helmut (2015) : Are GARCH innovations independent - a long term assessment for the S&P 500, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Financial Econometrics, No. B22-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/113109>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Are GARCH innovations independent - a long term assessment for the S&P 500

Helmut Herwartz*

February 28, 2015

Abstract

GARCH specifications have been widely applied in financial literature and practice. For purposes of (Quasi) ML (QML) estimation innovations to GARCH processes are assumed identically and independently distributed (iid) with mean zero and unit variance. In this note GARCH innovations entering daily S&P 500 quotes are diagnosed to lack independence and to signal ex-ante the directions of stock price changes.

Keywords: GARCH, independence testing, stock return predictability.

JEL Classification: C22, G14.

*Department of Economics, Georg-August-University of Göttingen, Platz der Göttinger Sieben 5, D-37073 Göttingen, Germany, Tel: +49 551 39 7350, Fax +49 551 39 7279, email: hherwartz@uni-goettingen.de. The author is grateful to Fabian H. C. Raters for excellent support in data handling and the provision of efficient software modules.

1 Introduction

Numerous variants of (G)ARCH models benefit volatility estimation and risk management.¹ Incorporating the prominent leverage effect the threshold GARCH model (TGARCH(1,1,1)) introduced by Glosten et al. (1993) will be considered throughout in this study. To implement QML estimation it is common to assume strong (T)GARCH forms, implying that innovations are iid with zero mean and unit variance. Assuming less restrictive frameworks of semi-strong or weak TGARCH, higher order dependence in innovation processes might carry informational content for out-of-sample stock returns. In empirical practice such higher order dependence is typically not subjected to specification testing, and yet its eventual content for ex-ante return prediction lacks consideration.

The purpose of this note is to unravel if, for a longitudinal sample drawn from the S&P 500, TGARCH innovations conform to the iid assumption. Consecutive innovation estimates are subjected to independence testing, and show strong indications of higher order dependence. In addition, nonparametric innovation forecasts ex-ante indicate price developments. Finally, inferential results (p -values) are combined with sign predictions to arrive at refined prediction schemes. The next Section sketches the data and outlines the empirical model and the design of the ex-ante analysis. Empirical results are provided in Section 3. Section 4 concludes.

2 The model and the forecasting design

2.1 Data and modelling framework

The data consists of 16348 daily observations of S&P 500 closing prices and covers the period from January 3, 1950 to December 19, 2014.² The following representation of returns r_t is used for an empirical assessment of innovation independence in TGARCH(1,1,1) models:

$$r_t = \nu + \phi r_{t-1} + e_t, \quad (1)$$

$$= \nu + \phi r_{t-1} + \sigma_t \xi_t,$$

$$\sigma_t^2 = \gamma_0 + \gamma_1 e_{t-1}^2 + \gamma_1^- e_{t-1}^2 I(e_{t-1} < 0) + \beta_1 \sigma_{t-1}^2, \quad (2)$$

$$\xi_t = e_t / \sigma_t \stackrel{iid}{\sim} N(0, 1), \quad \boldsymbol{\xi}_t = (\xi_{t-1}, \xi_t)', \quad t = \tau - G, \dots, \tau. \quad (3)$$

Conditional first and second order return characteristics are given in (1) and (2), respectively, where $I(\cdot)$ is an indicator function. In (3) τ indicates a rolling forecast origin and G is the size

¹For a textbook treatment of the family of GARCH models along with lead references see Mills and Markellos (2008).

²The sample was lastly updated from the Yahoo Finance historical prices database on December 22, 2014 at 7 pm. The prices were transformed to log returns and evaluated in R (R Core Team, 2014) by means of the R Package "rugarch" (Ghalanos, 2014) for univariate GARCH models, and the function "indepTest" from the R package "copula" (Hofert et al., 2014).

of (rolling) time windows of sample information (including presample values). Ex-ante forecasts are determined for forecast origins $\tau = G + 1, G + 2, \dots, T - 1$ where $G = 1000$, $T = 16347$. To immunize against the eventual detection of innovation dependence as a reflection of first order autoregressive dynamics the model in (1) includes an autoregressive component.³

2.2 Testing for independence of innovations

In an iid framework the vector $\boldsymbol{\xi}_t$ defined in (3) comprises independent elements. Samples $\Omega_m = \{\boldsymbol{\xi}_t\}_{t=\tau-G_m}^\tau$ are subjected to independence testing by means of the Cramér-von-Mises statistic which was proposed in Genest et al. (2007). Independence tests will be applied to nonoverlapping and overlapping/rolling samples of distinct sizes G_m . Specifically are the sample sizes $G = G_1 = 1000$, $G_2 = 500$, $G_3 = 250$ and $G_4 = 100$. Hence, sample information shrinks from an approximate 4 year period ($m = 1$) to some 'local' time support of 5 months ($m = 4$) of stock trading. Throughout, the notation is informative with respect to local time and the magnitude of sample information, i.e. independence diagnostics are denoted $\mathcal{B}_{\tau,m}$.

2.3 Prediction under dependence

In the case that the elements of $\boldsymbol{\xi}_t$ lack independence one has for the conditional expectation $E[\xi_t|\xi_{t-1}] \neq 0$. Hence, most recent innovations or their estimates ξ_τ might carry predictive content for future speculative returns. In this study nonparametric estimates $\hat{\mu}_{\tau+1,m} = E[\xi_{\tau+1}|\xi_\tau, \Omega_m]$ are evaluated for their potential to ex-ante signal the sign of 'centered' out-of-sample returns $\tilde{e}_{\tau+1} = r_{\tau+1} - \hat{\nu}_\tau - \hat{\phi}_\tau r_\tau$.⁴ Cumulated statistics of directional accuracy (DA) measure predictive performance, i.e.

$$\overline{\text{DA}}_m = \frac{1}{T-G} \sum_{\tau=G}^{T-1} I(\hat{\mu}_{\tau+1,m} \tilde{e}_{\tau+1} > 0) - I(\hat{\mu}_{\tau+1,m} \tilde{e}_{\tau+1} < 0). \quad (4)$$

In (4) time specific indicators of directional accuracy are +1 (-1) if the nonparametric predictor matches (does not match) in sign with the centered ex-ante return. Hence, a particular prediction scheme is considered 'successful' if $\text{DA}_m > 0$. To improve the scale of DA_m counts are divided by (rolling) numbers of overall predictions.⁵

³Positive autocorrelation of stock returns might be attributed to time varying risk premia (Engle et al., 1987). In empirical practice, however, the fit of autoregressive models often outperforms the one of risk premium models (Hafner and Herwartz, 2000). In Section 3 the case of imposing $\phi = 0$ will be shortly addressed.

⁴Conditional expectations $\hat{\mu}_{\tau+1,m}$ are determined by means of the Nadaraya-Watson estimator with Gaussian Kernel and an ad-hoc bandwidth $1.06G_m^{-0.2}$. OLS estimates $\hat{\nu}_\tau$ and $\hat{\phi}_\tau$ process sample information $\{r_t\}_{t=\tau-G}^\tau$.

⁵Given the return specification in (1) it appears natural to diagnose DA for AR(1) centered out-of-sample returns. Alternatively one may diagnose DA by means of combining $\hat{\mu}_{\tau+1,m}$ with either $\tilde{e}_{\tau+1} = r_{\tau+1} - \hat{\nu}_\tau$ (centering with mean return) or $\tilde{e}_{\tau+1} = r_{\tau+1}$ (no centering). Respective DA statistics strongly confirm the discussion in Section 3. Explicit results are available from upon request.

m	No pretest				With pretest ($p(\mathcal{B}_{\tau,m}) < 0.1$)			
	1	2	3	4	1	2	3	4
DA, cDA ($\cdot 100$)	3.343	2.417	2.639	2.587	3.906	5.453	5.279	7.937
$\tilde{T}/15347$	1	1	1	1	.586	.384	.262	.144
'hits'	.517	.512	.513	.513	.520	.527	.526	.540
sd ($\cdot 10E03$)	4.04	4.04	4.04	4.04	5.27	6.51	7.89	10.6

Table 1: Cumulated Directional Accuracy

The table documents the excess of cases of sign match over sign mismatch of ex-ante innovation predictions and centered future returns relative to evaluated predictions ($\overline{\text{DA}}, \overline{\text{cDA}}$). The row labeled ' $\tilde{T}/15347$ ' provides counts of evaluated predictions as a fraction of all iterations. Unconditional forecasts are compared with the outcome of a pretest scheme. Relative success measures are significantly positive with 1% significance. The standard errors of sample means from respective Bernoulli distributions with success probability 0.5 are shown in the bottom row of the Table. Frequencies of sign matching (e.g. $0.5 + \overline{\text{DA}}/2$) are given in the row labelled 'hits'.

2.4 A pretest approach

Under innovation independence the performance of ex-ante predictors $\hat{\mu}_{\tau+1,m}$ is random. A particular underpinning of dependence in TGARCH innovations obtains if, specifically, low p -values attached to $\mathcal{B}_{\tau,m}$, $p(\mathcal{B}_{\tau,m}) < 0.1$ say, are informative for the local ex-ante performance of $\hat{\mu}_{\tau+1,m}$. Respective conditional DA (cDA) statistics read as

$$\overline{\text{cDA}}_m = \frac{1}{\tilde{T}_m} \sum_{\tau=G}^{T-1} (I(\hat{\mu}_{\tau+1,m} \tilde{e}_{\tau+1} > 0) - I(\hat{\mu}_{\tau+1,m} \tilde{e}_{\tau+1} < 0)) I(p(\mathcal{B}_{\tau,m}) < 0.1), \quad (5)$$

where $\tilde{T}_m = \sum_{\tau=G}^{T-1} I(p(\mathcal{B}_{\tau,m}) < 0.1)$.

3 Empirical results

3.1 Inference in disjoint subsamples

Offering a first impression of eventual dependence in TGARCH innovations, S&P 500 returns are divided into 16 subsamples each comprising 1001 observations.⁶ Returns are subjected either to centering by their empirical mean ($\tilde{e}_t = r_t - \hat{\nu}_\tau$) or transformed into residuals from an AR(1) regression $\tilde{e}_t = r_t - \hat{\nu}_\tau - \hat{\phi}_\tau r_{t-1}$. Then, in total, 16 samples of innovation tuples $\boldsymbol{\xi}_t$ with $G = G_1 = 1000$ observations are obtained and subjected to independence testing. Moreover, each sample $\Omega_1 = \{\boldsymbol{\xi}_t\}_{t=\tau-G_1}^\tau$ is divided into 2, 4, and 10 nonoverlapping subsamples, comprising 500, 250 and 100 observations, respectively. Hence, 16, 32, 64 and 160 independence tests for

⁶For the analysis of nonoverlapping subsamples the most recent 331 observations are left out from the analysis.

(p_l, p_u)		[0,0.1)	[0.1,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,0.9)	[0.9,1)	
H_0		.10	.15	.25	.25	.15	.10	$U(0, 1)$
G_m	#	$\tilde{e}_t = r_t - \hat{\nu}$						
1000	16	.750	.188	.000	.062	.000	.000	.000
500	32	.656	.156	.188	.000	.000	.000	.000
250	64	.531	.156	.172	.094	.031	.016	.000
100	160	.331	.206	.169	.150	.069	.075	.000
		$\tilde{e}_t = r_t - \hat{\nu} - \hat{\phi}r_{t-1}$						
1000	16	.625	.125	.188	.062	.000	.000	.000
500	32	.438	.219	.125	.156	.031	.031	.000
250	64	.312	.094	.281	.188	.078	.047	.000
100	160	.125	.144	.238	.231	.162	.100	.911

Table 2: Independence diagnosis in disjoint samples

The table documents empirical frequencies of p -values $p(\mathcal{B}_m)$ between a lower bound p_l and an upper bound p_u . The row labeled ' H_0 ' provides the nominal frequency counterparts that are expected under the null hypothesis of innovation independence, and '#' is the number of samples subjected to independence testing. The column labelled ' $U(0, 1)$ ' shows p -values for a $\chi^2(5)$ test of a uniform distribution of the empirical distribution of $p(\mathcal{B}_m)$. Almost throughout, the null hypothesis of uniform p -values is rejected.

nonoverlapping samples are done. Under independence p -values $p(\mathcal{B}_\tau)$ should exhibit a uniform distribution. Table 2 documents the frequency estimates for particular ranges of the obtained p -values. Diagnostic results for samples Ω_1 are at odds with the independence assumption. Neglecting (accounting for) the first order dynamics in returns, 12 (10) out of 16 samples of ξ_t are diagnosed dependent with 10% significance, respectively. For smaller samples comprising either 500 (Ω_2) or 250 observations (Ω_3) the empirical pattern of p -values is also noticeably different from a uniform distribution. After adjusting for the first order autoregressive dynamics, 43.8% and 31.2% of all samples Ω_2 and Ω_3 are diagnosed dependent with 10% significance, respectively. Conditionally on small samples of 100 observations (Ω_4) the independence test appears to suffer from low power. In this case, the empirical distribution of the 160 p -values is in line with a uniform distribution if the TGARCH model is applied to residuals from linear autoregressions. Interestingly, the null hypothesis of independence is rejected with 10% significance for one third of the 160 samples when the raw returns are not subjected to AR(1) prewhithening.

3.2 Inference and prediction in rolling samples

Rolling independence diagnosis Figure 1 displays the empirical pattern of consecutive p -values, $p(\mathcal{B}_{\tau,m})$, $\tau = G, G + 1, \dots, T - 1$, $m = 1, 2, 3, 4$.⁷ Conditioning on Ω_1 the time profile of $p(\mathcal{B}_{\tau,1})$ reveals that small p -values cluster in the beginning and at the end of the rolling modelling design. While empirical p -values are somewhat larger for subsamples ending during

⁷Subsamples of Ω_1 are determined at the end of rolling samples.

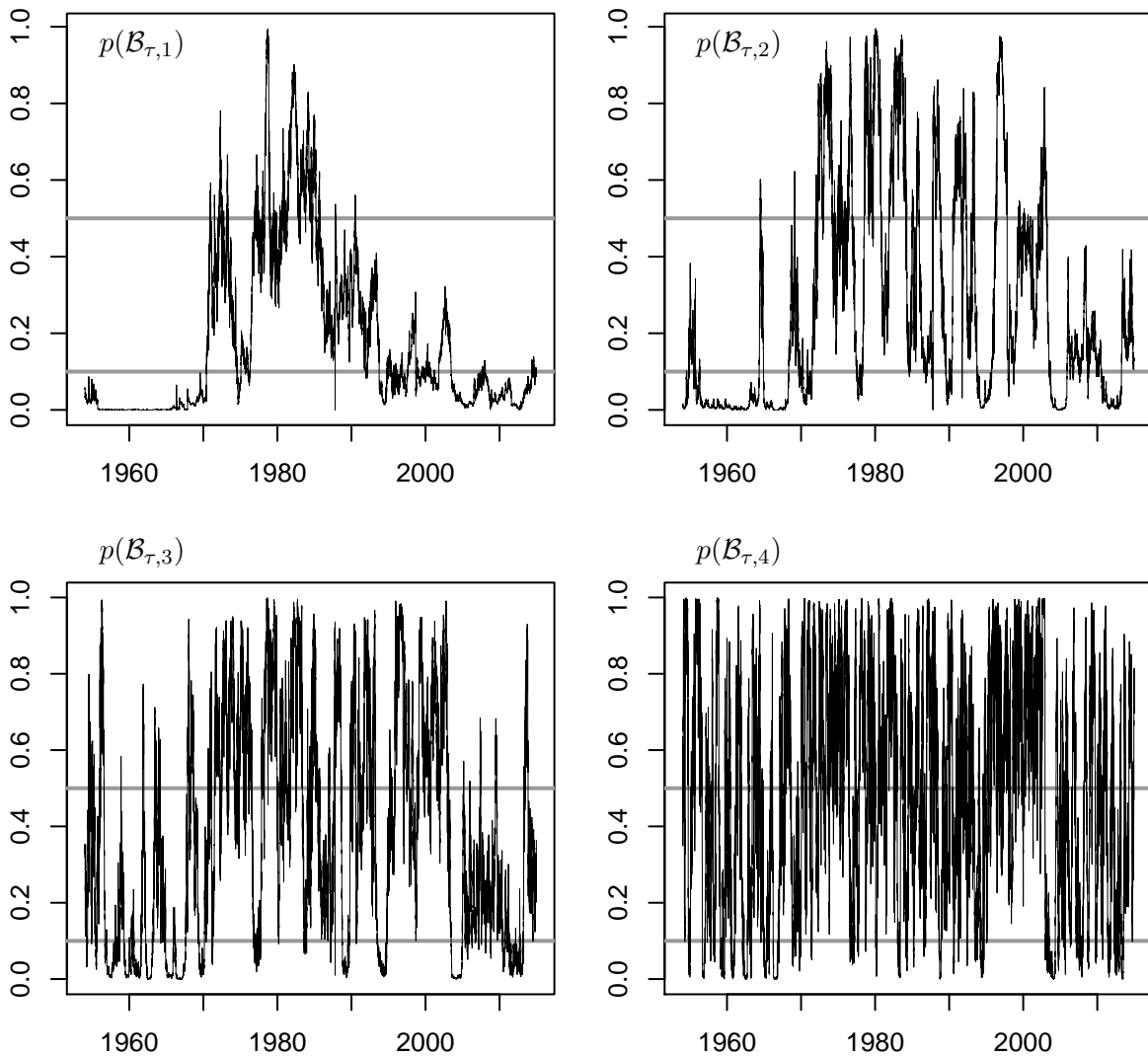


Figure 1: Rolling p -values $p(\mathcal{B}_{\tau,m})$ from independence testing in samples $\Omega_m, m = 1, 2, 3, 4$. To facilitate the interpretation of the graphs horizontal lines indicate p -values of 0.5 and 0.1.

the late 1970s and 1980s, the overall pattern of rolling p -values is at odds with independence. Using smaller samples the occurrence of larger p -values turns out to be more frequent, although rolling p -values obtained from samples Ω_2 are far from supporting the independence assumption. Conditional on further reduced sample information (Ω_3 and Ω_4), the respective test statistics ($\mathcal{B}_{\tau,m}, m = 3, 4$) reveal some power loss of the testing approach. However, rolling p -values indicate that the strength of innovation dependence is likely time varying.

Innovation forecasts Table 1 shows full sample statistics describing the forecasting performance. Nonparametric innovation forecasts $\hat{\mu}_{\tau+1,m}$ carry predictive content for the sign of out

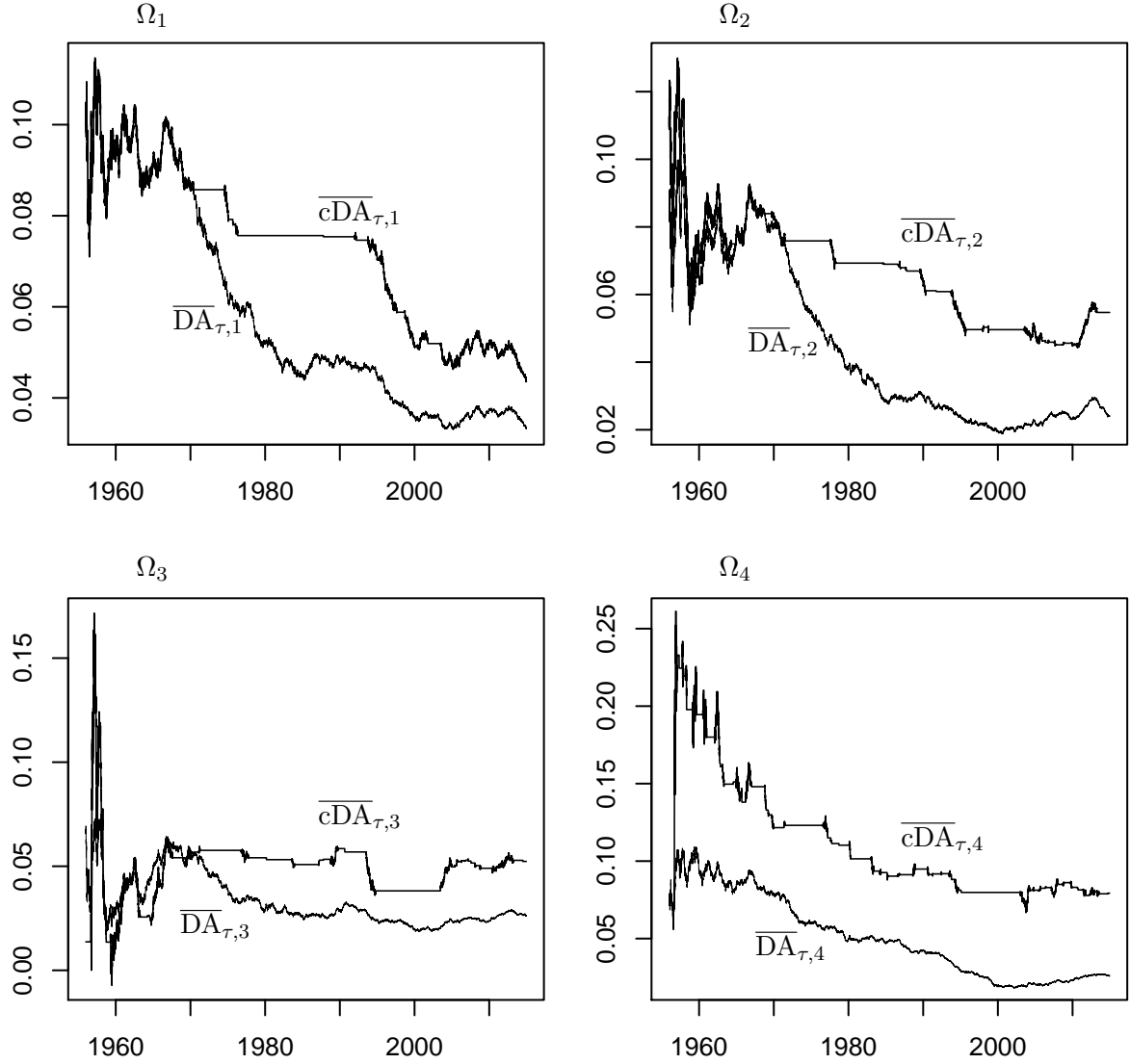


Figure 2: Rolling frequencies for ex-ante directional accuracy, unconditional (\overline{DA}) vs. pretest ($p(\mathcal{B}_{\tau,m}) < 0.1$, \overline{cDA}).

of sample (centered) returns $\tilde{e}_{\tau+1} = r_{\tau+1} - \hat{\nu}_{\tau} - \hat{\phi}_{\tau} r_{\tau}$. Overall, out of the 15347 single predictions the excess of sign matches over sign mismatch amounts to some percentage between 2.41% (Ω_2) and 3.34% (Ω_1). Noting the huge number of determined predictors the documented success frequencies are significant.⁸

Pretesting Evaluating directional accuracy only at time instances when $p(\mathcal{B}_{\tau,m}) < 0.1$ improves the documented success ratios markedly. With rolling samples Ω_2 the relative excess of accurate

⁸Interestingly, predicting naively a positive sign for $\tilde{e}_{\tau+1}$ also obtains a relative excess of matches over mismatches of 2.41%. Hence, with regard to this particular prediction outcome the naive strategy appears, unconditionally, as valuable as evaluating $\hat{\mu}_{\tau+1,2}$.

predictors raises from 2.41% to 5.45%. Conditioning, e.g., on $p(\mathcal{B}_{\tau,2}) < 0.1$ reduces the number of experiments to 38.4% of all rolling estimates. Still, however, $\tilde{T}_2 = 0.384 \cdot 15347 = 5905$ experiments enter the performance assessment for $\hat{\mu}_{\tau,2}$ such that the documented relative performance improvement is significant. Figure 2 shows rolling frequencies for DA statistics.⁹ As it turns out, the average success frequencies are time stable and conditioning on pretests dominates the performance of unconditional prediction schemes robustly.

4 Conclusions

In their strongest form GARCH models formalize that returns of speculative assets can be decomposed into a measurable conditional standard deviation and an independent identically distributed innovation process with mean zero and unit variance. For a longitudinal sample of S&P 500 returns this work shows that estimated innovations are mostly at odds with the independence assumption. The level of actual dependencies is likely to be time varying. Independence diagnostics can be fruitfully combined with conditional innovation forecasts to improve the performance of stock return prediction schemes. Although the analysis does not address profitable trading, it provides, to some extent, an empirical support for (time varying) market inefficiency.

References

- Engle, R. F.; Lilien, D. M. and Robins, R. P. (1987): Estimating Time Varying Risk Premia in the Term Structure: The Arch-M Model, in: *Econometrica*, 55, p. 391.
- Genest, C.; Quessy, J. and Rémillard, B. (2007): Asymptotic Local Efficiency of Cramér-von-Mises Tests for Multivariate Independence, in: *The Annals of Statistics*, 35, pp. 166–191.
- Ghalanos, A. (2014): rugarch: Univariate GARCH models. R package version 1.3-4.
- Glosten, L. R.; Jagannathan, R. and Runkle, D. E. (1993): Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, in: *Journal of Finance*, 48, pp. 1779–1801.
- Hafner, C. and Herwartz, H. (2000): Testing for Linear Autoregressive Dynamics under Heteroskedasticity, in: *The Econometrics Journal*, 3, pp. 177–197.
- Hofert, M.; Kojadinovic, I.; Maechler, M. and Yan, J. (2014): copula: Multivariate Dependence with Copulas, R package version 0.999-12.
- Mills, T. and Markellos, R. (2008): The Econometric Modelling of Financial Time Series, 3rd, New York: Cambridge University Press.
- R Core Team (2014): R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

⁹To improve the scale of the graphs, a 'burn-in' period of the first 500 ex-ante statistics has been removed from the display.