Missing mean does no harm to volatility!

Stanislav Anatolyev*[1], Irina Tarasyuk[2]

[1,2]New Economic School

June 2015

Abstract

Many empirical studies of financial volatility within the GARCH framework tend to exclude terms in the mean equation that are often proved to be statistically significant. We analyze analytically how omission of various mean terms affects the value of the ARCH parameter in the simple ARCH(1) model. We focus on the following terms missing from the mean equation when they are in fact present: a constant, a seasonal dummy, an autoregressive term, and a time-varying risk premium. We track how the relative distortion in the value of the ARCH parameter depends on amount of misspecification, and calibrate it to actual daily and monthly returns. It turns out that the effect on the variance equation of missing elements in the mean equation tends to be quite benign.

Keywords: misspecification, ARCH, financial returns

*Corresponding author. Address: Stanislav Anatolyev, New Economic School, 100A Novaya Street, Skolkovo, Moscow, 143026 Russia. E-mail: sanatoly@nes.ru. The authors are grateful to a referee whose comments and suggestions improved the presentation.

1 Introduction

It is well known that any misspecification in the mean equation passes over to the variance equation and distorts its estimation and all practically important output such as volatility and value-at-risk predictions. There is widespread evidence that the mean equation for financial returns may contain various deterministic trends (constant terms, seasonal dummies) and/or stochastic components (lagged returns, time-varying risk premia). For example, Berument & Kiymaz (2001) show that the day of the week effect is present in the return equation; Li et al (2005) find statistically significant ARCH-in-mean effect (Engle et al, 1987) with a carefully specified volatility equation.

Applied researchers usually specify the mean equation rather arbitrarily and often neglect what may in fact be present. Typically, they include a constant, sometimes they invoke seasonal dummies if the data exhibit some periodicity. Rarely do they include an autoregressive component which, if included, serves as a guard against possible serial correlation. And hardly ever do they include risk premium terms, except when studying evolution of risk premia is of explicit interest.

There is a bulk of literature that studies impact of misspecification on the ARCH component, see Lumsdaine & Ng (1999), van Dijk & Frances (1999) and Blake & Kapetanios (2007), among others. However, this literature concentrates on the impact on properties of tests for ARCH effects, while the question how misspecification distorts the ARCH equation seems to remain unanswered. In this paper, we shed some light on how important it is to keep the mean equation correctly specified.

We derive the formulas for relative distortions of the ARCH parameter in the simple ARCH(1) model resulting from omission of components from the mean equation. We focus on effects of omission of the following four terms, one by one, when they are in fact present: a constant, a seasonal dummy, an autoregressive term, and a time-varying risk premium. We analyze these formulas using 3-D graphs in order to track how the relative distortion in the value of the ARCH parameter depends on the value of the ARCH parameter and on amount of misspecification.

We intentionally focus on a 'toy' volatility model from the GARCH class, as it allows to obtain *analytical* results for the quantity of interest. Another advantage of the ARCH(1) structure is that it has a minimal number of parameters of interest that are inputs to any forecasts, the ARCH coefficient being most important. If we get clearcut conclusions (as we do) for the simplest ARCH model, they may be thought of being applicable to more complex situations under GARCH modeling.

When calibrated to typical daily and monthly returns, the relative distortions tend to be pretty small, often negligible. The relative distortion that is likely to occur in practice is of order of meager few percent, much smaller than typically the standard error for the ARCH parameter is. One can then conclude that in typical scenarios the effect on the variance equation of missing elements in the mean equation must be quite benign and may be ignored.

2 Model

Consider the ARCH(1) model (Engle, 1982) superimposed on an arbitrary mean equation:

$$r_t = \mu_t + \epsilon_t, \tag{1}$$

$$\epsilon_t = \sigma_t z_t, \quad z_t|\mathcal{F}_{t-1} \sim i.i.d.\mathcal{N}(0,1)\,,$$

and

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2, \tag{2}$$

where $\mathcal{F}_{t-1}$ is natural filtration, $\alpha > 0$, $\omega > 0$. We impose the condition $\alpha < 1/\sqrt{3}$ for fourth moments of ϵ_t to exist. Denote the unconditional variance of ϵ_t by σ^2.

Suppose now that a researcher omits the term μ_t from the conditional mean and estimates the same ARCH(1) model without the mean part. The missing μ_t is absorbed into the process for σ_t^2 distorting the true values of ω and α. Our purpose is to analyze the distortions to α under various scenarios for μ_t.

For analytical tractability, we consider estimation of the weak ARCH representation, see Drost & Nijman (1993) and Francq & Zakoïan (2000):

$$\epsilon_t^2 = \omega + \alpha\epsilon_{t-1}^2 + v_t, \tag{3}$$

where $\mathbb{P}\left(v_t|\mathcal{F}_{t-1}\right) = 0$. Then the distorted (pseudotrue) values of ω and α, say ω' and α', are determined from the representation

$$r_t^2 \;=\; \omega' + \alpha' r_{t-1}^2 + v_t',$$

where $\mathbb{P}\left(v_t'|\mathcal{F}_{t-1}\right) = 0$. That is, the pair (ω', α') is determined from a population least squares projection of squared return on its lag. In particular,

$$\alpha' \;=\; \frac{Er_t^2 r_{t-1}^2 - \left(Er_t^2\right)^2}{Er_t^4 - \left(Er_t^2\right)^2}. \tag{4}$$

The reason for utilizing the weak representation of the strong ARCH process (in the terminology of Drost & Nijman (1993)) is availability of the closed-form expression (4). A more popular estimation method in this context is quasi-maximum likelihood (QML) whose pseudotrue parameter values differ from (ω', α'). Both estimation methods are based on the moment condition $E[(r_t^2 - \omega - \alpha r_{t-1}^2)\zeta_{t-1}] = 0$, where ζ_{t-1} is an instrument different in the degree of optimality: it is $\zeta_{t-1} = (1, r_{t-1}^2)$ in case of the weak representation and $\zeta_{t-1} = (\omega + \alpha r_{t-1}^2)^{-2}(1, r_{t-1}^2)$ in the case of QML. Thus, these methods are expected to yield pseudotrue values that are close.[1]

[1]We evaluated, via simulations, the difference between the pseudotrue values for processes from section 3. When the parameters belong to typical sets described in section 4, we obtained differences of no more than 10%. When we blew up the degree of misspecification s hundredfold, these differences went up, but their proportion to the gap between the true and pseudotrue values never exceeded 40%.

3 Misspecification scenarios

We consider separately four types of mean specification omitted from the mean equation:

$$\begin{array}{rrcl}
\text{`lost constant'}: & \mu_t & = & \mu, \\
\text{`lost seasonality'}: & \mu_t & = & \frac{1}{2}(-1)^t\mu, \\
\text{`lost autoregression'}: & \mu_t & = & \rho r_{t-1}, \\
\text{`lost risk premium'}: & \mu_t & = & \delta\sigma_t^2.
\end{array}$$

These four cases imitate the usual practice when applied researchers omit such terms from the mean equation despite numerous evidence that they may be statistically significant in the real data.

Let us introduce a scale parameter $s \geq 0$ that summarizes influence of misspecification and thus can be thought of as the degree thereof. It is different under different scenarios: $s = (\mu/\sigma)^2$ under (qualitatively equivalent) 'lost constant' and 'lost seasonality,' $s = \rho^2$ under 'lost autoregression,' and $s = (\delta\sigma)^2$ under 'lost risk premium.' For each of the four types of misspecification we derive the function $\alpha' = \alpha'(\alpha, s)$, and analyze the relative difference[2]

$$\partial\alpha(\alpha, s) := \frac{\alpha'(\alpha, s) - \alpha}{\alpha}.$$

This measure does not depend on data generation parameters other than α and s. The computations and final formulas are given in an online Appendix at is.gd/missmean.

4 Results

We present the function $\partial\alpha$ using 3-D surface graphs produced by Wolfram *Mathematica*®. In all graphs below we restrict α to lie within the interval $[\underline{\alpha}, \bar{\alpha})$, where $\underline{\alpha}$ and $\bar{\alpha}$ are lower and upper bounds for α. We set $\underline{\alpha} = 0.1$ which is lower than is observed in returns data of high enough frequency. The upper bound $\bar{\alpha}$ is such that all moments needed for existence of α' exist; it varies with a type of misspecification. We restrict the 'amount of misspecification' parameter s as well: $s \in [0, \bar{s}]$.

The choice of $\bar{s}$ is driven by the following considerations. We have calibrated a mean-GARCH model with our four specifications of the mean to daily and monthly returns from a few stock indexes and exchange rates. More precisely, we have run rolling window (of 2000 observations for daily data and 500 observations for monthly data with a step of 100 and 10 observations, respectively) regressions on a few last decades of log returns, and read off maximal values of s.[3] For the 'lost constant' and 'lost seasonality' scenarios they lie within the range [0.002,0.030] and reach maximal values for monthly stock returns; for the 'lost autoregression' scenario they lie within the range [0.001,0.100] and reach maximal values for monthly exchange rate returns; for the 'lost risk premium' scenario they lie

[2] As long as α is far enough from zero, as is the case with financial returns of high enough frequency, there are no problems with computing such ratio. The logarithmic measure $\log(\alpha'/\alpha)$ would yield similar numerical results but the formulas would be less neat.

[3] In the 'lost seasonality' scenario, we used day-of-week dummies and tracked (half of) maximal discrepancy among dummy estimates, which is a worst-case analog of the mean estimate in the 'lost constant' scenario.

within the range [0.005,0.040] and reach maximal values for monthly stock returns. These intervals include values implied by typical evidence from empirical finance literature (e.g., see examples in Tsay (2002), chapter 3).

Under the 'lost constant' and 'lost seasonality' scenarios, $\bar{\alpha} = 1/\sqrt{3} \approx 0.577$ and $\bar{s} = 0.03$. In both cases the function $\partial\alpha\,(\alpha, s)$ is the same and is presented in Figure 1.

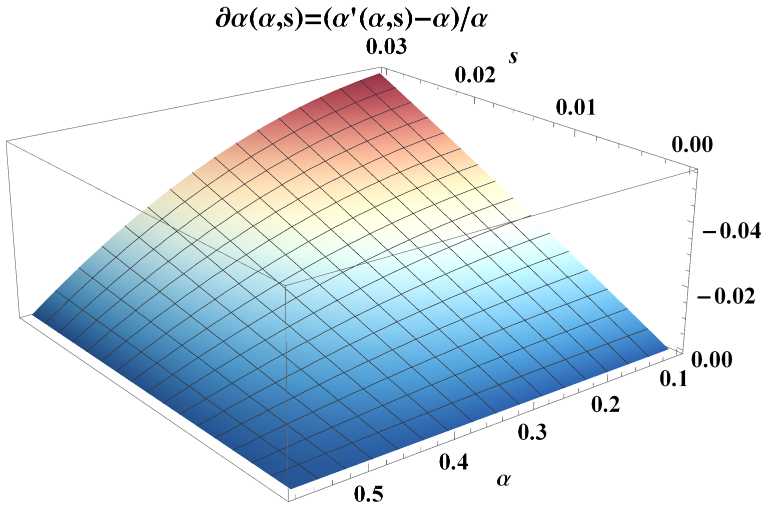

Figure 1: Distortions under 'lost constant' and 'lost seasonality' scenarios

The relative distortion is negative and decreases with the strength of heteroskedasticity. For typical values of s and likely values of α it amounts to about 1–2%. Even in extreme cases with lower values of α and high s the relative distortion will not exceed 5–6%.

Under the 'lost autoregression' scenario, $\bar{\alpha} = 1/\sqrt{3} \approx 0.577$ and $\bar{s} = 0.10$. The function $\partial\alpha\,(\alpha, s)$ is presented in Figure 2.

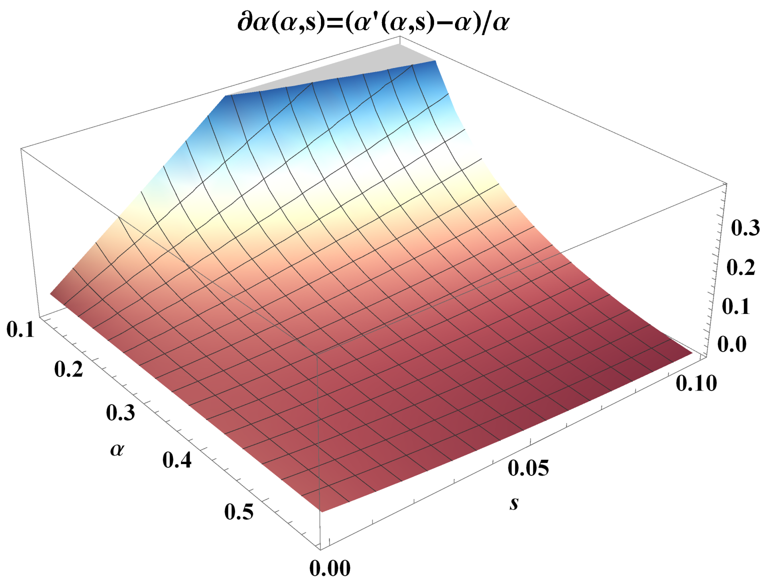

Figure 2: Distortions under 'lost autoregression' scenario

The relative distortion may be positive or negative (which is unrelated to the sign of serial correlation). It sharply rises when heteroskedasticity gets weaker and can attain dozens of percent. However, for typical values of s and likely values of α it fluctuates around 1%, and may even decrease when s is much larger (like that for monthly exchange rate returns) if heteroskedasticity is strong enough.

Under the 'lost risk premium' scenario, $\bar{\alpha} = 1/\sqrt[4]{105} \approx 0.312$ and $\bar{s} = 0.04$. The function $\partial\alpha\,(\alpha, s)$ is presented in Figure 3.

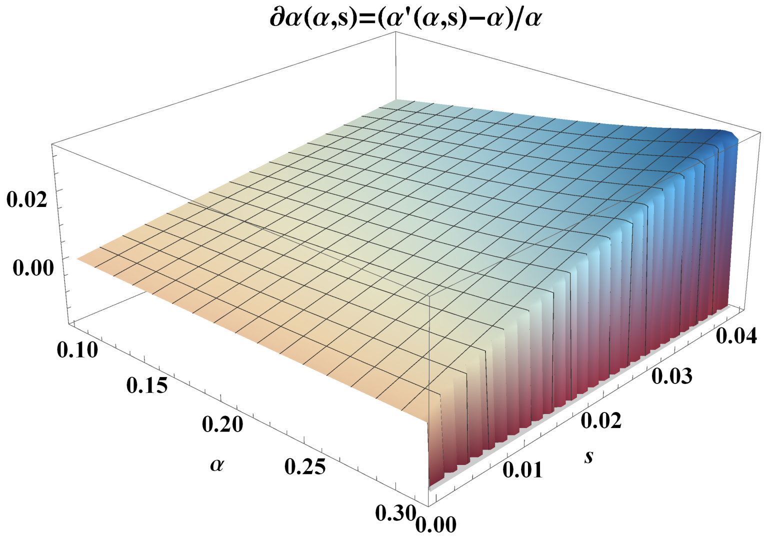

Figure 3: Distortions under 'lost risk premium' scenario

The relative distortion may be positive or negative (which is unrelated to the sign of the risk premium term). It is pretty flat over the region except near the edge of $\bar{\alpha}$ where the surface sharply changes its shape. However, for typical values of s the relative distortion never exceeds 2–3% by absolute value.

Note that the ARCH parameter is typically estimated with such precision that its standard error has values up to 20% of value of the parameter (e.g., Frances & van Dijk (2000), chapter 4; Tsay (2002), chapter 3). This is much larger than the relative distortions obtain above, even for worst case scenarios.

5 Concluding remarks

Based on our findings, we can conclude that the effect on the variance equation of missing elements in the mean equation must be quite benign, and applied researchers need not worry about the issue.

We have intentionally exploited a very simple ARCH model in order to obtain analytical clear-cut conclusions. Setups with more realistic GARCH processes and possible omissions of combinations of various mean components may be explored in a Monte-Carlo framework.

References

Berument, H. & Kiymaz, H. (2001). The day of the week effect on stock market volatility. *Journal of Economics and Finance*, 25(2), 181–193.

Blake, A.P. & Kapetanios, G. (2007). Testing for ARCH in the presence of nonlinearity of unknown form in the conditional mean. *Journal of Econometrics*, 137(2), 472-488.

Drost, F.C. & Nijman, T.E. (1993). Temporal Aggregation of GARCH Processes. *Econometrica*, 61(4), 909–927.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.

Engle, R., Lilien, D. & Robins, R. (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica*, 55(2), 391–407.

Francq, C. & Zakoïan, J.-M. (2000). Estimating weak GARCH representations. *Econometric Theory*, 16(5), 692–728.

Frances, P.H. & van Dijk, D. (2000). *Nonlinear Time Series Models in Empirical Finance.* Cambridge University Press.

Li, Q., Yang, J., Hsiao, C. & Chang, Y.-J. (2005). The relationship between stock returns and volatility in international stock markets. *Journal of Empirical Finance*, 12(5), 650–665.

Lumsdaine, L. & Ng, S. (1999). Testing for ARCH in the presence of a possibly misspecified conditional mean. *Journal of Econometrics*, 93(2), 257-279.

Tsay, R.S. (2002). *Analysis of Financial Time Series.* Wiley.

van Dijk, D. & Frances, P.H. (1999). Testing For ARCH in the presence of additive outliers. *Journal of Applied Econometrics*, 14(5), 539–562.