Studies in Nonlinear Dynamics & Econometrics

Volume 11, *Issue* 1 2007 *Article* 1

Spurious Inference in the GARCH (1,1) Model When It Is Weakly Identified

Jun Ma* Charles R. Nelson†
Richard Startz‡

*University of Washington, jma@cba.ua.edu
†University of Washington, cnelson@u.washington.edu
‡University of Washington, startz@u.washington.edu

Spurious Inference in the GARCH (1,1) Model When It Is Weakly Identified*

Jun Ma, Charles R. Nelson, and Richard Startz

Abstract

This paper shows that the Zero-Information-Limit-Condition (ZILC) formulated by Nelson and Startz (2006) holds in the GARCH (1,1) model. As a result, the GARCH estimate tends to have too small a standard error relative to the true one when the ARCH parameter is small, even when sample size becomes very large. In combination with an upward bias in the GARCH estimate, the small standard error will often lead to the spurious inference that volatility is highly persistent when it is not. We develop an empirical strategy to deal with this issue and show how it applies to real datasets.

*The authors wish to thank Rob Engle for suggesting that the ZILC phenomenon may be present in the GARCH model. The authors gratefully acknowledge financial support from the Buechel Memorial Scholarship, the Ford and Louisa Van Voorhis Professorship and the Cecil and Jane Castor Professorship, respectively, at the University of Washington. We thank Torben Andersen, Richard Baillie, Russell Davidson, Michael Dueker, James Morley, Christopher Neely, Jeremy Piger, James Ramsey, Fallow Sowell, Mark Watson, Mark Wohar and Eric Zivot for helpful comments; but the authors are entirely responsible for the content of this paper.

1. Introduction

Capturing time-varying volatility is a key element in modeling time series data, especially for financial time series data. The ARCH (Autoregressive Conditional Heteroskedasticity) family, first proposed by Engle (1982), has been widely adopted to extract a latent volatility process and predict its future movement, especially since the generalization to the GARCH model by Bollerslev (1986). In allowing the conditional volatility to be linearly dependent upon both past squared shocks and the past conditional volatilities, GARCH type models can generate rich dynamics with few parameters. Indeed, the GARCH(1,1) is usually sufficient to provide a good fit (see Bollerslev, Chou and Kroner (1992)).

Nelson and Startz (2006) have shown that when identification of one parameter is conditional on another inference for the former will be misleading if the Zero-Information-Limit Condition (hereafter ZILC) holds. In models where ZILC holds, standard errors tend to be understated when the identifying parameter is small enough, no matter how large a given sample size. Examples include the 'weak instrument' problem, ARMA models with near cancellation, and certain nonlinear regression models. In this paper we show that ZILC holds in the GARCH(1,1) model and that estimated standard errors are too small when the ARCH effect is of the size commonly reported in the empirical literature. As a result, the actual size of the t-test for the GARCH coefficient is far too great, rejection of the true null hypotheses occurring too often. Thus, researchers unaware of this spurious effect may be tempted to infer that the persistence due to the GARCH effect is strong when in fact it is absent.

As a response to the danger of spurious inference we propose an empirical strategy based on a pure ARCH(q) approximation to GARCH(1,1) and show how it applies to real datasets.

This paper is organized in the following way: Section 2 demonstrates that ZILC holds in the GARCH(1,1) model. Section 3 presents evidence by Monte Carlo experiments to document the underestimation of standard errors when identification of the GARCH effect is weak. Section 4 proposes the empirical strategy and evaluates its validity. Section 5 presents the results for some real datasets. Section 6 concludes this paper.

2. The Zero-Information-Limit Condition in the GARCH(1,1) Model

The archetypal GARCH(1,1) model may be written[1]:

[1] The mean of equation (2.1) is set to be zero without loss of generality since the information matrix is block-diagonal, as shown by Bollerslev (1986).

$$\varepsilon_t = \sqrt{h_t} \cdot \xi_t, \xi_t \sim i.i.d.N(0,1) \qquad (2.1)$$

$$h_t = \omega + \alpha \cdot \varepsilon_{t-1}{}^2 + \beta \cdot h_{t-1} \qquad (2.2)$$

Note that h_t is the conditional variance and is driven by past realizations of ε_t with added persistence determined by β. In the case $\beta = 0$ the model reduces to the pure ARCH(1) model, and in the case $\alpha = 0$ the data are homoskedastic and the GARCH effect β is not identified. Following the literature, we impose the parameter restrictions $\omega > 0$ and $|\alpha + \beta| < 1$ so that the underlying process is strictly stationary with a finite second moment. Note that the asymptotic theory of GARCH(1,1) does not critically depend upon the latter inequality restriction (e.g., see Lumsdaine (1996), Jensen and Rahbek (2004)), but we impose this restriction to have a finite unconditional variance for ε_t and evaluate its estimation performance.

Following the standard treatment in Hamilton (1994), we present the following ARMA(1,1) representation for the GARCH(1,1) process:

$$\varepsilon_t{}^2 = \omega + (\alpha + \beta) \cdot \varepsilon_{t-1}{}^2 + w_t - \beta \cdot w_{t-1} \qquad (2.3)$$

The innovation $w_t = \varepsilon_t{}^2 - h_t = h_t[(\frac{\varepsilon_t}{\sqrt{h_t}})^2 - 1]$ is a Martingale Difference Sequence (MDS) with a time-varying conditional variance. Thus the GARCH(1,1) process turns out to be a particular ARMA(1,1) process with $(\alpha + \beta)$ being the AR coefficient and β being the MA coefficient, though the shocks are non-normal and heteroskedastic.

Nelson and Startz (2006) show that ZILC holds in an ARMA(1,1) model as the absolute difference between the AR coefficient and MA coefficient approaches zero. If ZILC applies to the GARCH(1,1) model as well, the reported standard error of the MLE estimator $\hat{\beta}$ will tend to be smaller than the true asymptotic standard deviation when the identifying parameter α is small. To check whether ZILC holds one needs the asymptotic variance but no closed form expression exists in the literature. Ma (2006) derives an analytical approximation for the case that α is small and an exact expression that may be evaluated by stochastic simulation for comparison. Based upon Ma's result it is straightforward to show that the inverse of the asymptotic variance of $\hat{\beta}$, the 'information' measure of Nelson and Startz (2006), goes to zero as α approaches zero, i.e., ZILC holds:

$$\lim_{\alpha \to 0} I_{\hat{\beta}}(\omega, \alpha, \beta) = 0 \qquad (2.4)$$

Appendix A gives a formal proof of (2.4) based upon Ma's (2006) analytical result. Furthermore, $\hat{\omega}$ has the same issue as $\hat{\beta}$, since it is also subject to ZILC. Indeed, Ma's approximation establishes that these two estimates are highly negatively correlated when α is small; Appendix B illustrates these algebraic results using the special case of $\beta = 0$. Asymptotic theory does hold in the GARCH as sample size grows, but for any given sample size one can find a value of α small enough that the ZILC effect on standard errors will be apparent. Finally, the identifying parameter α itself is still well identified, in the sense that ZILC does not hold.

3. Evidence of Spurious Inference from Monte Carlo Experiments[2]

We implement a series of Monte Carlo (MC) experiments to investigate whether spurious inference occurs when the GARCH(1,1) model is weakly identified. There have been a few papers which examine the performance of GARCH estimates in a finite sample through MC experiments but the focus has been on the well identified case; see Hong (1988), Bollerslev and Wooldridge (1992), Lumsdaine (1995), and Fiorentini, Calzolari and Panattoni (1996). In the empirical literature it is standard practice to rely on estimated standard errors for the GARCH parameter to make the inference that β is non-zero and in the typical case large with a small confidence interval. Thus, we are interested here in the potential for spurious inference when there is in fact no GARCH effect, or it is only moderate.

3.1 Inference when there is no GARCH effect

In this sequence of MC experiments, data is simulated from the GARCH(1,1) process defined by equation (2.1) and (2.2) with three sets of parameter values:

$$\begin{pmatrix} \omega \\ \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 \\ 0.01 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0.05 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0.10 \\ 0 \end{pmatrix}$$

The choices of α are motivated by the estimates typically reported in the empirical literature; some classical examples are Bollerslev (1987), Baillie and

[2] The estimation procedure is implemented by our own MATLAB codes, independent from the GARCH Toolbox in MATLAB. We tried both the restricted code which restricts the estimates to be positive through an exponential transformation and the unrestricted code which does not have this restriction. Similar results are obtained in both cases. Here we only report the results from the unrestricted code. Codes in both cases are available from the authors upon request. Our major findings can also be replicated in both Eviews 5.1 and the SPLUS Finmetrics Library 1.0.

Bollerslev (1989) and Engle, Ng, and Rothschild (1990). Since β is 0 in these experiments there is no GARCH effect and the process is actually an ARCH(1). The scale parameter ω is normalized to be unity. For each set of parameter values, we have three sample sizes T = 500, 1000 and 5000, respectively. For all 9 experiments, 1000 simulated paths of sample data of length T are generated. Table 1 gives the empirical sizes of *t*-test, Likelihood Ratio (LR) test, Lagrange Multiplier (LM) test at the nominal 5% level for all parameters and the frequency Schwarz Information Criterion (SIC) chooses GARCH(1,1) over ARCH(1).

Table 1: Size of various tests at 5% level and SIC in GARCH(1,1)

	T = 500	T = 1000	T = 5000
	True Parameter Values: $\omega=1, \alpha=0.01, \beta=0$		
t-test for ω	47.5%	45.2%	44.4%
t-test for α	21.8%	20.1%	20.8%
t-test for β	48.7%	45.6%	44.5%
LR test for β	13.0%	10.9%	8.3%
LM test for β	4.7%	5.2%	4.6%
SIC correct	5.7%	2.9%	1.0%
	True Parameter Values: $\omega=1, \alpha=0.05, \beta=0$		
t-test for ω	38.3%	35.7%	16.8%
t-test for α	19.8%	18.2%	6.7%
t-test for β	41.3%	36.0%	17.5%
LR test for β	11.1%	9.9%	7.3%
LM test for β	4.7%	6.0%	4.3%
SIC correct	5.1%	2.9%	0.8%
	True Parameter Values: $\omega=1, \alpha=0.10, \beta=0$		
t-test for ω	27.3%	19.4%	7.9%
t-test for α	12.6%	10.2%	5.2%
t-test for β	30.6%	21.0%	8.5%
LR test for β	8.9%	8.4%	5.6%
LM test for β	4.5%	6.0%	4.5%
SIC correct	3.7%	2.3	0.1%

In Table 1, when the ARCH coefficient is 0.01 the actual size of *t*-test for β is nearly 50% even for a large sample size. However, for sufficiently large α, and for sufficiently large sample size, the size distortion is greatly reduced. Note that the size distortion for $\hat{\omega}$ is as large as that for $\hat{\beta}$. Size distortion for $\hat{\alpha}$ is not as large as for $\hat{\beta}$, although not completely absent.

Fortunately for practitioners, the LR and LM tests perform much better than the *t*-test. The former indicates that the weakly identified model does not fit much better than the restricted model, hence little improvement in the likelihood value. The better performance of the LM test can be traced to the fact that it is calculated under the restriction on the weakly identified parameter; see Zivot, Startz and Nelson (1998) for discussion of this in the weak instrument case. Ma and Nelson (2006) are exploring approaches to obtaining valid tests based on what would be the Anderson-Rubin test in a linear approximation to a weakly identified model where ZILC holds. Note that the LM test in this context is a chi-square test for serial correlation in the squared residuals from the constrained model, in this case ARCH(1). SIC performs well in model selection which is consistent with findings on lag selection reported by Lutkepohl (1991).

To understand why the *t*-statistic does such a poor job, we separately examine the denominator and the numerator. In Table 2 we compare the median of the estimated standard error of $\hat{\beta}$ in the MC sample with the actual standard deviation of $\hat{\beta}$ in the MC sample as well as with two computed approximations to the asymptotic standard deviations, one using Ma's analytical approximation and the other evaluated by stochastic simulation; see Ma (2006) for details. This comparison is for the fixed sample size $T = 1000$.

Table 2: Estimated standard error vs. true asymptotic standard deviation
True Parameters values: $\omega = 1, \alpha = 0.01, 0.05, 0.10, \beta = 0$, $T = 1000$

Identifying Parameter α	Model Parameters Estimates	Median of Estimated S.E.	Standard Deviation of Estimates in MC	Asymptotic SD using Ma approx.	Asymptotic SD evaluated numerically
0.01	$\hat{\omega}$	0.3226	0.6161	3.1621	3.3549
	$\hat{\alpha}$	0.0266	0.0381	0.0313	0.0332
	$\hat{\beta}$	0.3164	0.6175	3.1303	3.3192
0.05	$\hat{\omega}$	0.3022	0.5532	0.6317	0.6712
	$\hat{\alpha}$	0.0349	0.0401	0.0300	0.0374
	$\hat{\beta}$	0.2874	0.5402	0.5993	0.6364
0.10	$\hat{\omega}$	0.2686	0.4083	0.3164	0.3513
	$\hat{\alpha}$	0.0408	0.0436	0.0282	0.0411
	$\hat{\beta}$	0.2394	0.3719	0.2817	0.3142

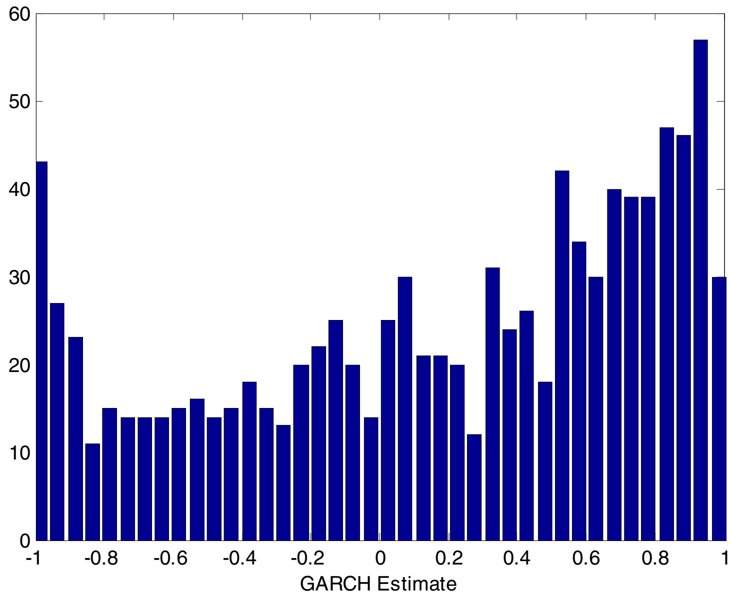
The standard error of $\hat{\beta}$ is indeed severely underestimated. For example, when $\alpha = 0.01$, the median estimated standard error of $\hat{\beta}$ is only about one tenth of the true (asymptotic) standard error. As pointed out by Nelson and Startz (2006) for the ARMA case, this is more surprising since variation in $\hat{\beta}$ is bounded by the stationarity requirement although the asymptotic formula does not take this into account. However, the estimated standard error for $\hat{\beta}$ is so much underestimated that it is well below the actual standard deviation, being about half of it. Even when $\alpha = 0.10$ the median estimated standard error of $\hat{\beta}$ is still well below both the asymptotic standard error and the actual standard deviation. While $\hat{\omega}$ has exactly the same issue, this is not true for $\hat{\alpha}$.

We present the histogram of $\hat{\beta}$ in Plot 1 from the experiment when $\omega = 1, \alpha = 0.01, \beta = 0$ and sample size T = 1000, corresponding to the first three rows in Table 2. An interesting "pile-up" phenomenon appears which reflects an upward bias in $\hat{\beta}$: the median of $\hat{\beta}$ is 0.3207. At the same time $\hat{\omega}$ is downward biased with the median being 0.6889.

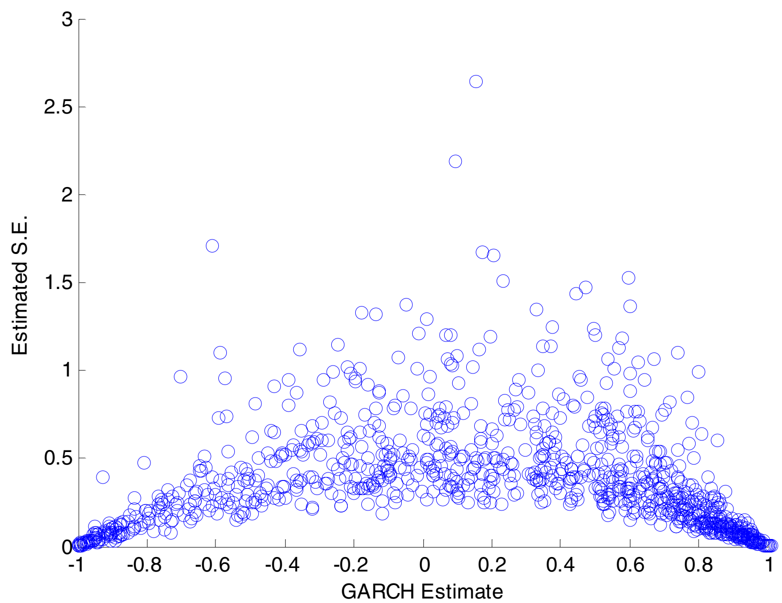
We also plot the estimated standard error of $\hat{\beta}$ against $\hat{\beta}$ in Plot 2. It is evident that there is a strong negative correlation between the absolute value of $\hat{\beta}$ and its standard error. Nelson and Startz (2006) show that a general property of models in which ZILC holds is *dependence* between absolute size of the numerator and denominator of the *t*-statistic, the sign of the correlation determining whether the *t*-test is under- or over-sized. In this case, large values of $\hat{\beta}$ are accompanied often by very small estimated standard errors, and vice versa, so there is an excess of large *t*-statistics and the test size is too great.

Another finding in the MC experiment is that very often the individual Profile Log-Likelihood Function (PLLF) displays multiple maxima (See Plot 3 for two typical examples). The PLLF is obtained by maximizing the log-likelihood function (LLF) subject to pre-specified values of β. This suggests that practitioners should be aware of the possibility of getting stuck in a local maximum in lieu of a global one when relying on a traditional gradient-based optimum searching algorithm. In our experiment we start optimizations with various initial values to avoid this pitfall. Furthermore, the PLLF varies little as the parameter is varied. Interestingly, Figlewski (1997) finds that it is difficult to get the algorithm to converge when estimating GARCH(1,1) for monthly stock return because the LLF is quite flat.

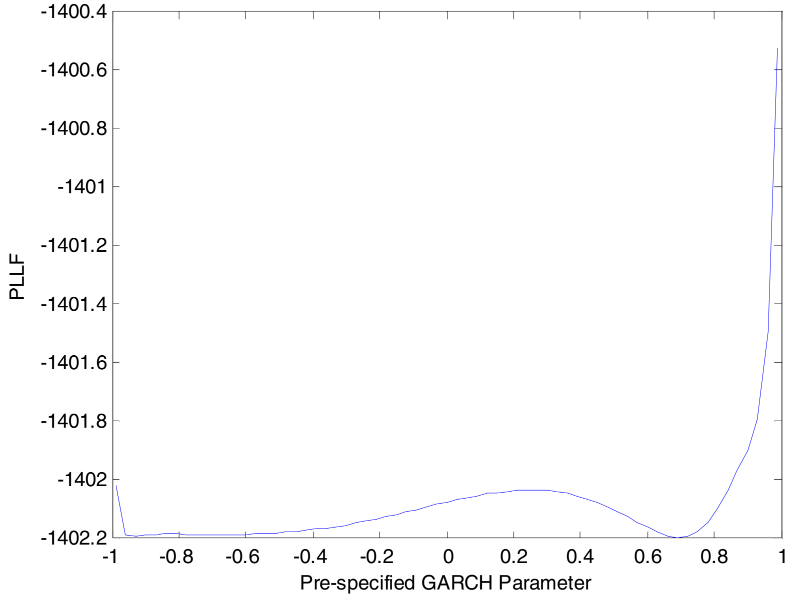
Plot 1: Histogram of $\hat{\beta}$, $\omega = 1, \alpha = 0.01, \beta = 0, T = 1000$

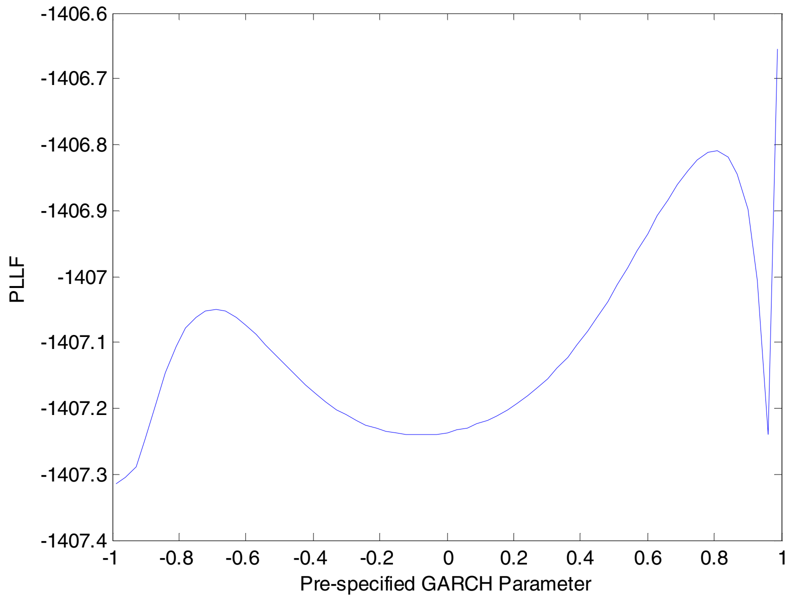


Plot 2: Scatter plot of estimated S.E. of $\hat{\beta}$ vs $\hat{\beta}$, $\omega = 1, \alpha = 0.01, \beta = 0, T = 1000$



Plot 3: Two examples of PLLF, $\omega = 1, \alpha = 0.01, \beta = 0, T = 1000$





3.2 When there is a moderate GARCH effect

It is important to note that ZILC holds whenever the ARCH coefficient α is small, regardless of the magnitude of true β. In this sequence of MC experiments we simulate data from the GARCH(1,1) process defined by equation (2.1) and (2.2) with moderate GARCH effect:

$$\begin{pmatrix} \omega \\ \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 \\ 0.01 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 1 \\ 0.05 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 1 \\ 0.10 \\ 0.5 \end{pmatrix}$$

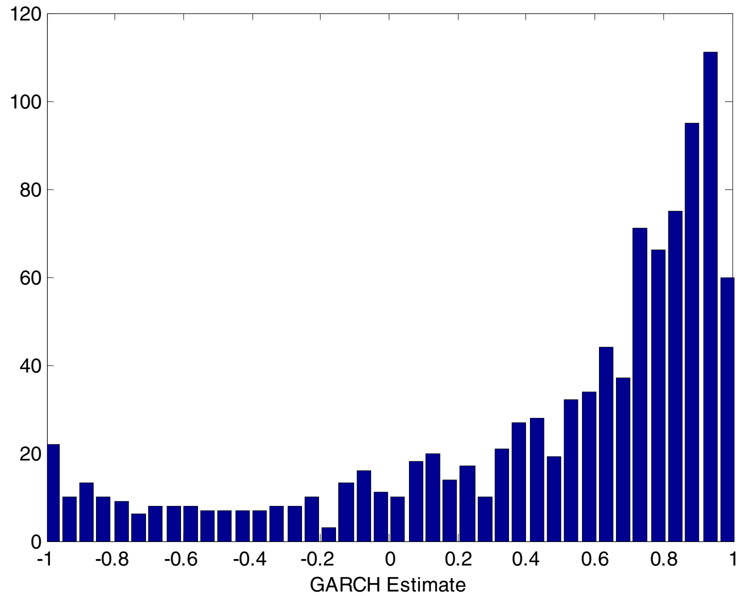
The sample size is fixed at 1000 and the number of simulation is 1000. Table 3 presents our major findings. No major difference has been found compared with Table 1 and 2. The standard error of $\hat{\beta}$ is underestimated, leading to a very large size distortion of *t*-test. Besides, $\hat{\beta}$ is upward biased (See Plot 4) and the median is 0.6834. The LR and LM test again perform much better than *t*-test and SIC is quite accurate in model selection.

Table 3: Inference for β in GARCH(1,1) with moderate GARCH effect

True parameters values: $\omega = 1, \alpha = 0.01, 0.05, 0.10, \beta = 0.5$, T = 1000

	True value of α		
	0.01	0.05	0.10
	Standard Deviation of $\hat{\beta}$		
Asy. (analyt. approx.)	2.0400	0.3957	0.1887
Asy. (num. eval.)	2.0665	0.4237	0.2149
Std Dev in MC sample	0.5499	0.4485	0.2768
MC median S.E.	0.2566	0.2332	0.1818
	Size of tests of null hypothesis $\beta = 0.5$ at nominal 5% level		
t-test	42.7%	29.2%	16.1%
LR test	8.4%	7.0%	6.7%
LM test	6.3%	5.4%	6.3%
SIC correct	2.7%	2.0%	1.3%

Plot 4: Histogram of $\hat{\beta}$, $\omega = 1, \alpha = 0.01, \beta = 0.5, T = 1000$



3.3 Persistence in the GARCH(1,1) model

In the GARCH(1,1) model $(\alpha + \beta)$ determines how long a random shock to volatility persists. To see this we rewrite equation (2.2) to obtain its AR representation:

$$h_t = \omega + (\alpha + \beta)h_{t-1} + \alpha w_{t-1}$$
$$where, w_{t-1} = \varepsilon_{t-1}{}^2 - h_{t-1} \qquad (3.3.1)$$

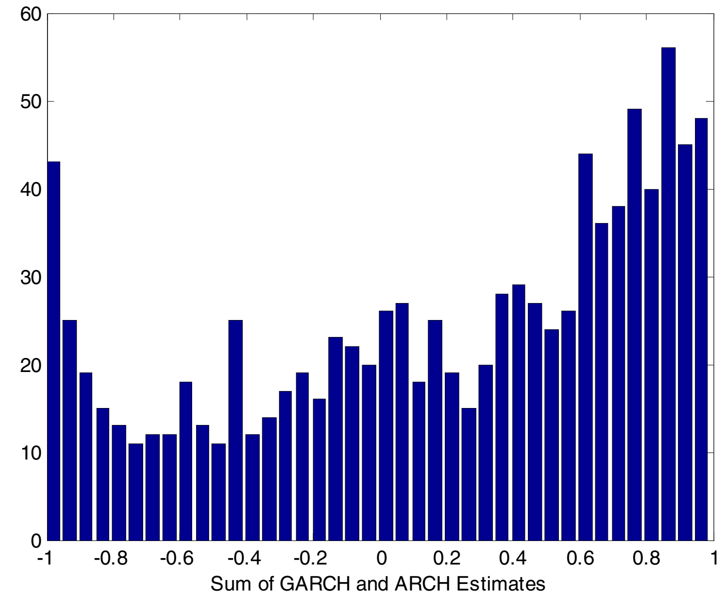
In empirical applications it is often this persistence in volatility that is of great interest and the magnitude of it usually makes a significant difference in terms of economic implications. For example, Bansal and Yaron (2000) present a potential resolution of the equity premium puzzle based on a large value of $(\alpha + \beta)$. So it is important to note that $(\hat{\alpha} + \hat{\beta})$ is upward biased and has an underestimated standard error when α is small. Plot 5 and 6 give the histograms of $(\hat{\alpha} + \hat{\beta})$ with parameter values $\omega = 1, \alpha = 0.01, \beta = 0$ and $\omega = 1, \alpha = 0.01, \beta = 0.5$, respectively, and T = 1000. Table 4 reports the size distortion of the t-test for $(\alpha + \beta)$ under both cases of no GARCH and moderate GARCH effect when the model is weakly identified, for the fixed sample size T = 1000. The size distortion of $(\hat{\alpha} + \hat{\beta})$ is comparable to that of $\hat{\beta}$. Table 4 also gives the power of *t*-test for the Integrated GARCH (IGARCH) process[3]. Notice that the power is very small when α is small. Furthermore, given the same α, the power is even smaller when the true β increases.

Table 4: Inference for $(\alpha+\beta)$ in GARCH(1,1), sample size $T = 1000$

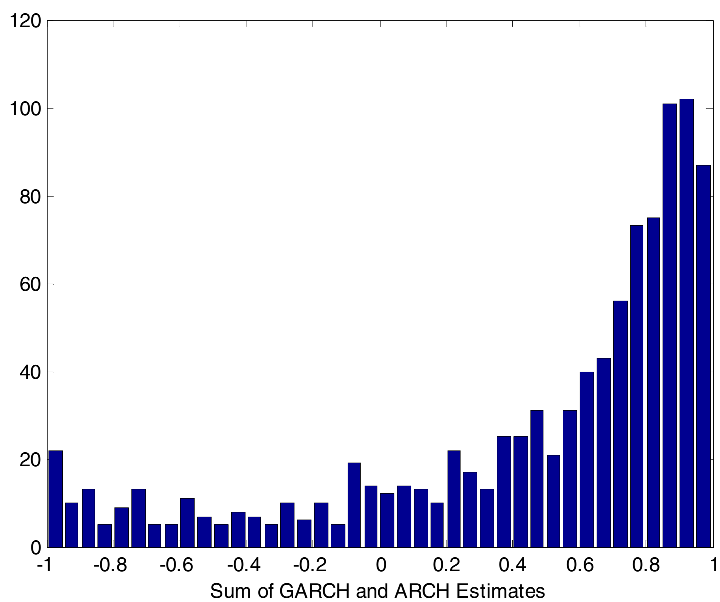
True parameters values: $\omega=1, \alpha=0.01, 0.05, 0.10, \beta=0$			
	True value of α		
	0.01	0.05	0.10
Standard Deviation of $\hat{\alpha}+\hat{\beta}$			
Asy. (analyt. approx.)	3.1402	0.6303	0.3130
Asy. (num. eval.)	3.1302	0.6301	0.3114
Std Dev in MC sample	0.6101	0.5292	0.3667
MC median S.E.	0.3188	0.2851	0.2404
Size of *t*-test of null hypothesis $\alpha+\beta$ equals its true value at nominal 5% level			
t-test	45.6%	35.2%	19.5%
Power of *t*-test for the hypothesis $\alpha+\beta=1$ at nominal 5% level			
t-test	40.4%	56.1%	80.1%
True parameters values: $\omega=1, \alpha=0.01, 0.05, 0.10, \beta=0.5$			
	True value of α		
	0.01	0.05	0.10
Standard Deviation of $\hat{\alpha}+\hat{\beta}$			
Asy. (analyt. approx.)	2.0519	0.4067	0.1957
Asy. (num. eval.)	2.0262	0.3817	0.1741
Std Dev in MC sample	0.5439	0.4387	0.2638
MC median S.E.	0.2522	0.2273	0.1677
Size of *t*-test of null hypothesis $\alpha+\beta$ equals its true value at nominal 5% level			
t-test	42.3%	29.2%	16.5%
Power of *t*-test for the hypothesis $\alpha+\beta=1$ at nominal 5% level			
t-test	25.1%	38.9%	69.8%

[3] The test is asymptotically valid since the GARCH estimates have regular properties even for an IGARCH process. See Lumsdaine (1995, 1996) and D. Nelson (1990).

Plot 5: Histogram of $(\hat{\alpha}+\hat{\beta})$, $\omega=1, \alpha=0.01, \beta=0, T=1000$



Plot 6: Histogram of $(\hat{\alpha}+\hat{\beta})$, $\omega=1, \alpha=0.01, \beta=0.5, T=1000$



3.4 Forecasting performance of the GARCH(1,1) model when it is weakly identified

Due to its practical interest, here we evaluate both the in-sample and out-of-sample forecasting performance of the GARCH(1,1) model when it is weakly identified. The in-sample forecasting is simply the estimated volatility which can be easily computed once the parameters estimates are obtained. Out-of-sample forecasting for horizon *k* is also straightforward as shown below:

$$E_t[h_{t+h}] = \omega \sum_{i=0}^{h-1} (\alpha+\beta)^i + (\alpha+\beta)^h h_t \qquad (3.4.1)$$

And $\lim_{h\to\infty} E_t[h_{t+h}] = \frac{\omega}{1-\alpha-\beta}$, given $|\alpha+\beta|<1$.

We work on the MC experiment of $\omega=1, \alpha=0.01, \beta=0, T=1000$. Since the ARCH(1) model is correctly specified given $\beta=0$, we use the ARCH(1) model as a benchmark. At the same time, we estimate the constant unconditional variance as another benchmark: $CONST.h = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2$.

Plot 7 presents a typical comparison of the in-sample forecasted volatility by three methods along with the true volatility (the illustrated sample period is chosen to be short to make the difference clear). The estimated volatility by the ARCH(1) and the constant unconditional variance measure resemble the underlying volatility process quite well, indicating a nearly homoskedastic process. However, the estimated volatility by the GARCH(1,1) displays a very persistent pattern. The out-of-sample comparison given by Plot 8 demonstrates the same idea.

We compute the Root Mean Squared Error (RMSE) across MC samples to summarize the predicting accuracy for various methods. Table 5 gives the in-sample RMSE and out-of-sample RMSE for GARCH(1,1), ARCH(1) and the constant unconditional variance measure. The in-sample comparison seems to be counter-intuitive to the common sense that the in-sample fitting should be always better with a more general model. The analogy to a linear estimation explains the puzzle. The forecasting measure $\sum (h_t - \hat{h}_t)^2$ here we use corresponds in a linear estimation to the Explained Sum of Squares (ESS) not the Sum of Squared Residuals (SSR), whose counterpart is $\sum (\varepsilon_t^2 - \hat{h}_t)^2$ instead. As the more general model decreases $\sum (\varepsilon_t^2 - \hat{h}_t)^2$, $\sum (h_t - \hat{h}_t)^2$, however, has to increase, given the fixed Total Sum of Squares (TSS) $\sum (\varepsilon_t^2 - h_t)^2$:

$$\sum (\varepsilon_t^2 - h_t)^2 = \sum (\hat{h}_t - h_t)^2 + \sum (\varepsilon_t^2 - \hat{h}_t)^2 + 2\sum (\hat{h}_t - h_t)(\varepsilon_t^2 - \hat{h}_t) \qquad (3.4.2)$$

Where, the term $\sum(\hat{h}_t - h_t)(\varepsilon_t^2 - \hat{h}_t)$ would be zero by construction in a linear context. In this nonlinear context, the value of this term is also close to zero in our MC experiment.

As for the out-of-sample forecasting performance, the GARCH(1,1) is also worse than both the ARCH(1) and constant variance measure for short horizons but all of them have almost the same performance for long enough horizons, indicating a well estimated unconditional variance $\hat{\omega} / 1 - \hat{\alpha} - \hat{\beta}$ even when the model is weakly identified. This is in contrast to Starica (2003) who investigate the forecasting performance of GARCH(1,1) model in the S&P 500 index return data and finds that it does a poor job in predicting the long run volatility during the period of his study. However, we want to point out that our MC experiments are implemented assuming a constant unconditional variance which may fail to hold for real data.

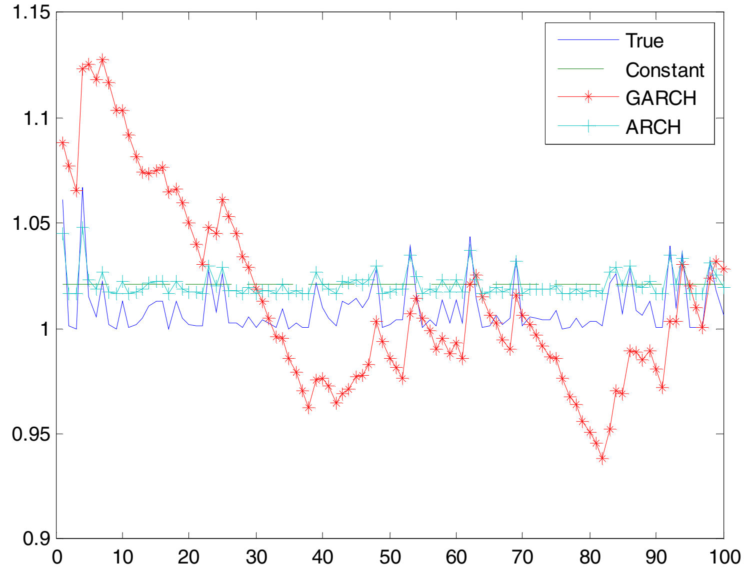
Table 5: Forecasting performance of GARCH(1,1)

$$\omega = 1, \alpha = 0.01, \beta = 0, T = 1000$$

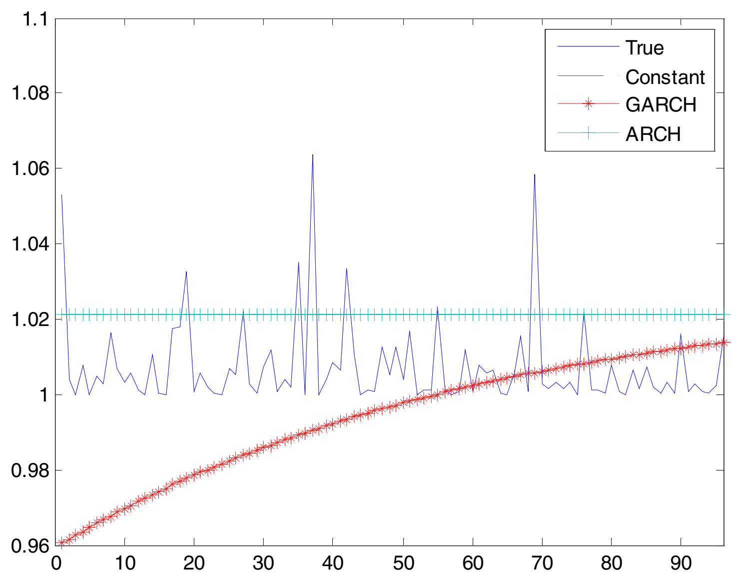
$$RMSE = \sqrt{\frac{1}{S}\sum_{i=1}^{S}(Forecast.h_i - True.h_i)^2}$$

	GARCH(1,1)	ARCH(1)	Constant
	In-sample RMSE		
Whole period	0.0895	0.0678	0.0485
	Out-of-sample RMSE for different horizons		
Horizon = 1	0.0714	0.0486	0.0485
Horizon = 3	0.0634	0.0495	0.0495
Horizon = 6	0.0582	0.0494	0.0494
Horizon = 9	0.0553	0.0484	0.0485
Horizon = 12	0.0541	0.0487	0.0487
Horizon = 24	0.0526	0.0493	0.0493
Horizon = 48	0.0498	0.0482	0.0482
Horizon = 96	0.0500	0.0490	0.0491

Plot 7: A typical comparison of the in-sample volatility forecast
$\omega = 1, \alpha = 0.01, \beta = 0, T = 1000$



Plot 8: A typical comparison of the out-of-sample volatility forecast
$\omega = 1, \alpha = 0.01, \beta = 0, T = 1000$



4. An Empirical Strategy for Detecting ZILC in the GARCH(1,1) Estimation

As suggested by our findings, a preliminary step to see whether one specific GARCH(1,1) estimation is subject to ZILC is to take a look at $\hat{\alpha}$ and the sample size since as α or sample size increases the ZILC issue becomes less severe. To facilitate this procedure, we provide a reference table (Table 6) for practitioners. We note that either sample size or the ARCH effect must be larger than generally encountered in the empirical literature for the ZILC problem to become moot.

Table 6: The reference table for practitioners - empirical size of *t*-test for testing $\hat{\beta}$ in the GARCH(1,1) model, $\omega = 1, \beta = 0$

Sample Size	True value of α						
	0.01	0.05	0.10	0.15	0.20	0.25	0.30
T = 250	52.8%	45.7%	37.9%	29.7%	24.4%	21.4%	21.5%
T = 500	48.7%	41.3%	30.6%	22.0%	17.4%	15.9%	--
T = 1000	45.6%	36.0%	21.0%	15.1%	11.9%	--	--
T = 5000	44.5%	17.5%	8.5%	6.8%	--	--	--

In our approach to GARCH(1,1) estimation, when $\hat{\alpha}$ and the sample size are in the left upper area of Table 6 we propose to estimate the ARCH(*q*) process and compare with the GARCH(1,1) estimation to see if there is any large discrepancy in the implied autocorrelation function (ACF) for h_t as a practical strategy for detecting a spurious result in estimating the weakly identified GARCH(1,1). The ARCH(*q*) process bears no ZILC concern since identification is not conditional on other parameters, as shown by its AR(*q*) representation:

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + w_t \qquad (4.1)$$

The GARCH(1,1) can be represented by an ARCH(∞) process theoretically:

$$\begin{aligned} \varepsilon_t^2 &= \omega + \alpha \cdot \varepsilon_{t-1}{}^2 + \beta \cdot h_{t-1} + w_t \\ &= \frac{\omega}{1-\beta} + \alpha \cdot \varepsilon_{t-1}{}^2 + \alpha\beta \cdot \varepsilon_{t-2}{}^2 + \ldots + \alpha\beta^{k-1} \cdot \varepsilon_{t-k}{}^2 + \ldots + w_t \end{aligned} \qquad (4.2)$$

In practice, an ARCH(*q*) process with sufficiently large lag q is able to approximate the GARCH(1,1) process very well. We verify this through a MC experiment.

We generate 1000 data paths of sample size T = 1000 by equation (2.1) and (2.2) with true parameters values $\omega = 1, \alpha = 0.3, \beta_1 = 0.6$. Given reference Table 6, this GARCH(1,1) process is well identified. This is also confirmed by the estimation result: the actual size of *t*-test for $\hat{\beta}$ is 5.1% for nominal size 5%.

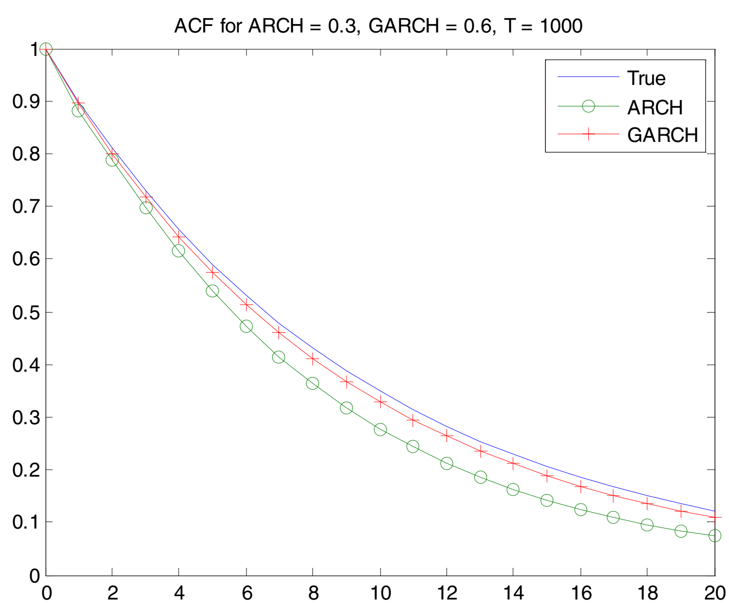
Besides, $\hat{\beta}$ is around its true value and there is no upward bias. For each data path, we estimate both the GARCH(1,1) and ARCH(q). To choose a proper lag q for the ARCH(q), we rely on both the SIC and LM test. We estimate the ARCH(q) up to lag 10 and find an optimal lag where SIC is minimal and LM test is not significant at 5% level.

After estimations, we compute and compare the theoretical ACF of the conditional variance implied by GARCH(1,1) and ARCH(q) estimates. Equation (3.3.1) shows that $(\alpha+\beta)$ fully determines the persistence of the conditional variance process in the GARCH(1,1). However, for the ARCH(q), the implied conditional variance has the ARMA(q, q-1) representation:

$$(1-\alpha_1 L-\ldots-\alpha_p L^p)(h_t - \frac{\alpha_0}{1-\alpha_1-\ldots-\alpha_p}) = (1-(-\frac{\alpha_2}{\alpha_1})L-\ldots-(-\frac{\alpha_p}{\alpha_1})L^{p-1})\alpha_1 w_{t-1} \quad (4.3)$$

We compare the median of theoretical ACF for the conditional variance across MC sample between the GARCH(1,1) and ARCH(q) estimations. We confirm that the ARCH(q) can approximate the GARCH(1,1) fairly well (see Plot 9). In the next section we examine some real datasets and experiment with this approach.

Plot 9



5. Issues in Real Data Analysis and the Example of S&P 500 Index Returns

Numerous applications of GARCH(1,1) appear in the literature and some generalizations are as follows. Very frequently a large value of $\hat{\beta}$ is reported, accompanied by a small standard error and large *t*-statistic. It is not uncommon to see a small $\hat{\alpha}$ along with a not very large sample size, the combination well in the area of Table 6 that suggests the danger of spurious inference. We cannot provide an explanation of the very frequently reported large values of $\hat{\beta}$ solely based on the results in this paper (the upward bias found in our Monte Carlo experiments are not sufficiently extreme). Dueker (1997) suggests that this may be due to the leptokurtic characteristic of the real data. Other studies such as Hamilton and Susmel (1994) and Cai (1994) attribute this to abrupt regime shifts of the unconditional variance.

In Engle, NG, Rothschild (1990), they estimate a GARCH(1,1)-mean model for monthly value-weighted stock index return data from August 1964 to November 1985. The number of observations is 256 and $\hat{\alpha}$ is slightly above 0.05. In contrast, $\hat{\beta}$ is quite large along with a very pronounced *t*-statistic. Their GARCH point estimates along with the *t*-ratios (in parentheses) are:

$$\hat{\omega} = 1.9348(1.68) \qquad \hat{\alpha} = 0.0518(1.79) \qquad \hat{\beta} = 0.8461(12.6)$$

In the first of two examples from Bollerslev (1987), the GARCH(1,1) estimation of daily U.S. dollar versus the British Pound exchange rate return data from March 1, 1980 to January 28, 1985 has the following point estimates and estimated standard errors (in parentheses):

$$\hat{\omega} = 0.96 \cdot 10^{-6}(0.46 \cdot 10^{-6}) \qquad \hat{\alpha} = 0.057(0.017) \qquad \hat{\beta} = 0.921(0.023)$$

And the GARCH(1,1) estimation of monthly S&P 500 index return data from 1947 January to 1984 September is (standard errors are in parentheses):

$$\hat{\omega} = 0.17 \cdot 10^{-3}(0.13 \cdot 10^{-3}) \qquad \hat{\alpha} = 0.074(0.045) \qquad \hat{\beta} = 0.768(0.148)$$

The former estimation gives $\hat{\alpha}$ as small as 0.057 with a sample size 1245. At the same time, $\hat{\beta}$ is very large and its estimated standard error is very small. The second estimation gives a slightly larger $\hat{\alpha}$ and a large $\hat{\beta}$ still, but with a much smaller sample size 453, and in this case neither ω nor α seems to be significantly different from zero at 5% level by a traditional *t*-test[4].

We take the monthly S&P 500 index return data as an example of our investigation. This dataset is obtained from the Eviews 5.1 DRI Database. We restrict our investigation to the sample period from 1947 January to 1984 September to make our estimation result comparable to Bollerslev (1987). Since the monthly price data is obtained by averaging the daily prices, there is a significant first order moving average correlation in the first moment equation, which is well known as the "Working Effect" (see Working (1960)). Therefore, we first estimate the MA(1) process for the return level data in EVIEWS 5.1 and store the residuals:

$$\hat{c}_0 = 0.005278(0.001901) \qquad \hat{\theta}_1 = 0.23694(0.052461)$$

c_0 is the constant in mean equation and θ_1 is the MA(1) coefficient. The White heteroskedasticity-consistent standard errors of estimates are in parentheses.

As suggested by Bollerslev (1988) as a routine check for the heteroskedasticity, the Ljung-Box test of squared residuals at log 10 is computed to be 26.9637, which is significant at 5% level. We provide two GARCH(1,1) estimation results. One is from EVIEWS 5.1 by directly estimating the MA(1) – GARCH(1,1) model. The other one is obtained by fitting the residuals from the first moment equation into the GARCH(1,1) model defined by equation (2.1) and (2.2) using our MATLAB code. Estimation results are reported below. To account for possible misspecification of conditional distribution for real data, we report the robust standard errors proposed by Bollerslev and Wooldridge (1992):

EVIEWS

$$\hat{\omega} = 0.16 \cdot 10^{-3} (0.14 \cdot 10^{-3}) \qquad \hat{\alpha} = 0.078(0.049) \qquad \hat{\beta} = 0.771(0.169)$$

MATLAB

$$\hat{\omega} = 0.16 \cdot 10^{-3} (0.14 \cdot 10^{-3}) \qquad \hat{\alpha} = 0.077(0.048) \qquad \hat{\beta} = 0.773(0.169)$$

These GARCH(1,1) estimations are quite similar to Bollerslev (1987) and all of them imply a persistent volatility process in that $\hat{\alpha} + \hat{\beta} \approx 0.85$. However, as we point out, the estimation result under this circumstance is probably subject to ZILC. To make a comparison, we fit the residuals into the ARCH(q) model. To determine the optimal lag, we estimate the ARCH(q) up to lag 10 and then identify the optimal lag where SIC achieves a local minimum and LM test is not

[4] The test of $\alpha = 0$ is non-standard, e.g., see the comment in Bollerslev, Engle and D. Nelson (1994). Davies (1977, 1987), Hansen (1996), Beg, Silvapulle, Silvapulle (2001) and Andrews (2001) have provided detailed discussions.

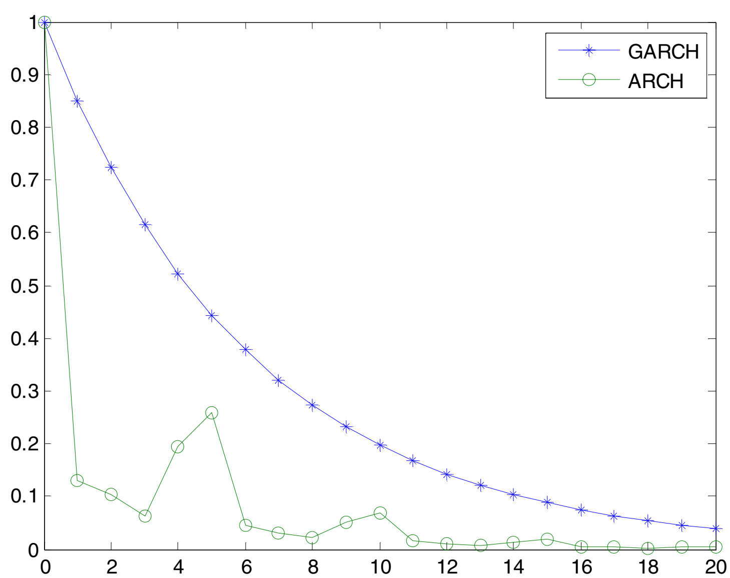
significant at 5% level[5]. This procedure results in the ARCH(5) and the estimation result is reported below. Again we report the robust standard errors proposed by Bollerslev and Wooldridge (1992):

$$\hat{\alpha}_0 = 0.73\cdot 10^{-3} \quad \hat{\alpha}_1 = 0.041 \quad \hat{\alpha}_2 = 0 \quad \hat{\alpha}_3 = 0.019 \quad \hat{\alpha}_4 = 0.008 \quad \hat{\alpha}_5 = 0.251$$
$$(0.21\cdot 10^{-3}) \quad (0.044) \quad (0.173) \quad (0.085) \quad (0.035) \quad (0.113)$$

The only significant lag is the 5th lag with a large magnitude. We find the same feature in the CRSP equal-weighted excess return data used in Kim, Nelson and Startz (1998). Oddly, Baillie and Bollerslev (1989) document a similar feature in the weekly exchange rate return data.

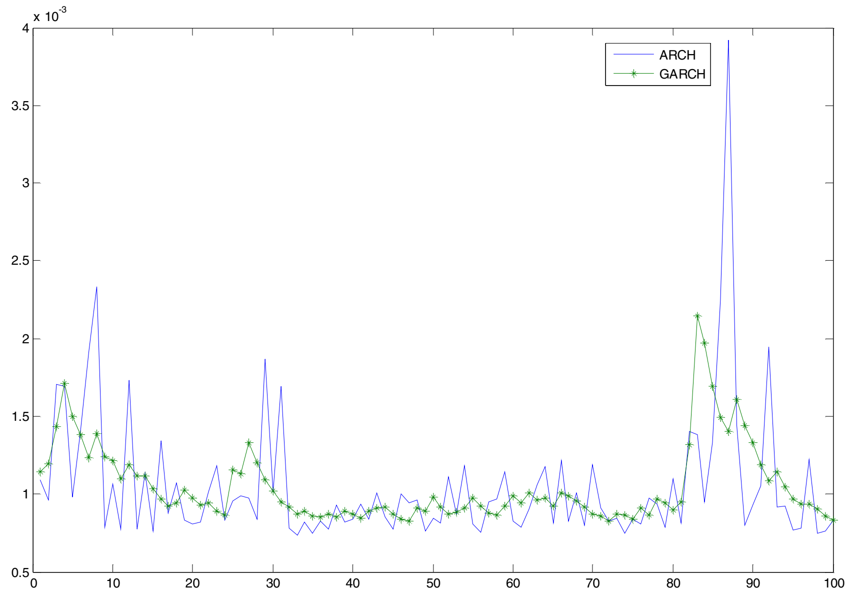
The theoretical ACF for the volatility process implied by both GARCH(1,1) and ARCH(5) estimates are given in Plot 10. The first order autocorrelation is 0.129 implied by the ARCH(5) estimation in a sharp contrast to 0.850 implied by the GARCH(1,1)estimation. The estimated conditional variance $\{\hat{h}_t\}_{t=1}^{T}$ from both estimations also differ greatly (See Plot 11). The PLLF for the GARCH(1,1) estimation is given by Plot 12. The PLLF turns out to be bimodal.

Plot 10: Theoretical ACF implied by GARCH(1,1) and ARCH(5) estimates,SP500

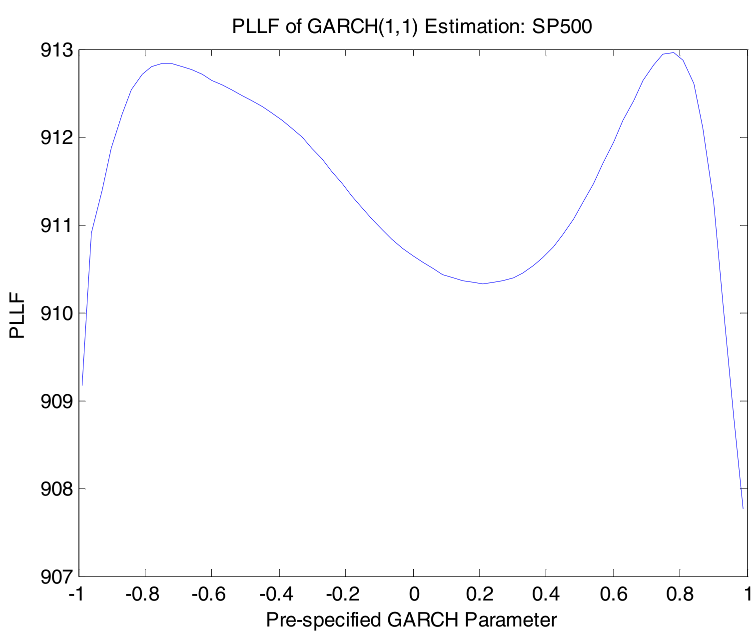


[5] We also use Akaike Information Criterion (AIC), which results in the same lag. When we look at the 6th up to 10th ARCH estimates in the ARCH(10) estimation, none of them is significant and the sum of them is negligibly small.

Plot 11: Estimated conditional variance of SP500 from GARCH(1,1) and ARCH(5) (A typical sub-sample is presented here to facilitate the visualization.)



Plot 12



We have also studied other datasets and the results are available upon request. Overall, $\hat{\alpha}$ is small but the sample size is not large enough to escape from the ZILC concern. Applying the proposed empirical strategy reveals a discrepancy between the theoretical ACF for the conditional variance implied by GARCH(1,1) and ARCH(q). For example, Baillie and Bollerslev (1989) note that there is almost no GARCH effect in the monthly exchange rate return data. However, the GARCH(1,1) estimation of the monthly exchange rate return data of U.S. dollar versus Japanese Yen in the sample period from 1971 January to 2006 January results in a large $\hat{\beta}$ with a very small standard error. Instead the ARCH(q) approach finds little persistence and the PLLF of the GARCH(1,1) is quite flat across the whole admissible region of β.

6. Conclusion

We show that the Zero-Information-Limit Condition (ZILC) formulated by Nelson and Startz (2006) holds in the GARCH(1,1) model so that the model is weakly identified when the ARCH coefficient is small. We present a sequence of Monte Carlo experiments and find that the GARCH estimate tends to have an underestimated standard error together with an upward bias when the ARCH coefficient is small even when sample size becomes very large, which results in a large size distortion of the *t*-test. We propose an empirical strategy for detecting ZILC and apply it to the real data. Our finding suggests that the concern raised by ZILC is quite relevant in empirical work.

References

Andrews, D. W. K. (2001): "Testing When a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683-734.

Baillie, R. T., and T. Bollerslev (1989): "The message in daily exchange rates: a conditional variance tale," *Journal of Business & Economic Statistics*, 7, 297-305.

Bansal, R., and A. Yaron (2000): "Risks for the Long Run: a Potential Resolution of Asset Pricing Puzzles," NBER Working Paper 8059.

Beg, R., M. Silvapulle, and P. Silvapulle (2001): "Tests against inequality constraints when some nuisance parameters are present only under the alternative: test of ARCH in ARCH-M models," *Journal of Business & Economic Statistics*, 19, 245-253.

Berkes, I., Horvath, L., and Kokoszka, P. (2003): "GARCH processes: structure and estimation," Bernoulli, 9, 201-227.

Box, G. E.P., and G. M. Jenkins (1976): *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

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

Bollerslev, T. (1987): "A conditional heteroskedastic time series model for speculative prices and rates of return," *The Review of Economics and Statistics*, 69, 542-547.

Bollersleve, T. (1988): "On the correlation structure for the generalized autoregressive conditional heteroskedastic process," *Journal of Time Series Analysis*, 9, 121-131

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994): "ARCH models," in: Engle, R. F., and D. L. McFadden, ed., *Handbook of Econometrics*, North Holland, Amsterdam, 2959-3038.

Bollerslev, T., and J. Wooldridge (1992): "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric Reviews*, 11, 143-172.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992): "ARCH modeling in finance-a review of the theory and empirical evidence," *Journal of Econometrics*, 52, 5-59.

Cai, J. (1994): "A markov model of switching-regime ARCH," *Journal of Business & Economic Statistics*, 12, 309-316.

Davies, R. B. (1977): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 64, 247-254.

Davies, R. B. (1987): "Hypothesis testing when a nuisance parameter is present only under the alternative", *Biometrika*, 74, 33-43.

Dueker, M. J. (1997): "Markov switching in GARCH processes and mean-reverting stock-market Volatility," *Journal of Business & Economic Statistics*, 15, 26-34.

Engle, R. (1982): “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation,” *Econometrica*, 50, 987-1008.

Engle, R., V. K. Ng, and M. Rothschild (1990): “Asset pricing with a factor-ARCH covariance structure-empirical estimates for treasury bills,” *Journal of Econometrics*, 45, 213-237.

Figlewski, S. (1997): “Forecasting volatility,” *Financial Markets, Institutions and Instruments*, 6, 1-88.

Fiorentini, G., G. Calzolari, and L. Panattoni (1996): ”Analytical derivatives and the computation of GARCH estimates,” *Journal of Applied Econometrics*, 11, 399-417.

Hamilton, J. (1994): *Time Series Analysis*. Princeton: Princeton University Press.

Hamilton, J., and R. Susmel (1994): “Autoregressive conditional heteroskedasticity and changes in regimes,” *Journal of Econometrics*, 64, 307-333.

Hansen, B. (1996): “Inference when a nuisance parameter is not identified under the null hypothesis,” *Econometrica*, 64, 413-430.

Hong, C. H. (1988): “The integrated generalized autoregressive conditional heteroskedastic model: the process, estimation and Monte Carlo experiments,” Unpublished Manuscript, Department of Economics, University of California, San Diego, CA.

Jensen, S. T., and Rahbek, A. (2004): “Asymptotic inference for nonstationary GARCH,” *Econometric Theory*, 20, 1203-1226.

Kim, C. J., C. R. Nelson, and R. Startz (1998): “Testing for mean reversion in heteroskedastic data based on Gibbs-sampling-augmented randomization,” *Journal of Empirical Finance*, 5, 131-154.

Lee, S. W., and B. Hansen (1994): “Asymptotic theory for The GARCH(1,1) quasi-maximum likelihood estimator,” *Econometric Theory*, 10, 29-52.

Lumsdaine, R. L. (1996): “Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models," *Econometrica*, 64, 575-596.

Lumsdaine, R. L. (1995): "Finite-sample properties of the maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models: a Monte Carlo investigation," *Journal of Business & Economic Statistics*, 13, No.1, 1-10.

Lutkepohl, H. (1991): *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

Ma, J. (2006): "A closed-form asymptotic variance-covariance matrix for the maximum likelihood estimator of the GARCH(1,1) model," working paper, Department of Economics, University of Washington.

Ma, J., C. R. Nelson (2006): "A general approach to constructing a valid test in weakly identified models where the Zero-Information-Limit Condition holds," working paper, Department of Economics, University of Washington.

Nelson, C. R., and R. Startz (2006): "The Zero-Information-Limit Condition and spurious inference in weakly identified models," forthcoming in the *Journal of Econometrics*.

Nelson, D. (1990): "Stationary and persistence in the GARCH(1,1) model," *Econometric Theory*, 6, 318-334.

Starica, C. (2003): "Is GARCH(1,1) as good a model as the Nobel Prize accolades would imply?" working paper, Department of Mathematical Statistics, Chalmers University of Technology.

Weiss, A. (1986): "Asymptotic theory for ARCH models: estimation and testing," *Econometric Theory*, 2, 107-131.

Working, H. (1960): "Note on the correlation of first differences of a random chain," *Econometrica*, 28, 916-918.

Zivot, E., R. Startz, and C. R. Nelson (1998): "Valid confidence intervals and inference in the presence of weak instruments," *International Economic Review*, 39, 1119-1144.

Zivot, E. and J. Wang (2002): *Modeling Financial Time Series with S-PLUS*. Springer-Verlag.

Appendix A

Ma (2006) gives an analytical information matrix for the GARCH(1,1) estimator $(\hat{\omega}, \hat{\alpha}, \hat{\beta})$:

$$I = \frac{(1-\alpha-\beta)^2}{2\omega^2}\begin{pmatrix} A & B & B \\ B & C & D \\ B & D & E \end{pmatrix} \qquad \text{(A.1)}$$

Where, $A = \dfrac{1}{(1-\beta)^2}$,

$B = \dfrac{1}{(1-\beta)^2} \cdot \dfrac{\omega}{1-\alpha-\beta}$,

$C = \dfrac{\omega^2}{(1-2\alpha\beta-\beta^2)(1-\alpha-\beta)}[\dfrac{3(1+\alpha+\beta)}{1-3\alpha^2-2\alpha\beta-\beta^2} + \dfrac{2\beta}{(1-\beta)^2}]$,

$$D = \frac{\omega^2(1+\alpha+\beta)}{(1-\alpha-\beta)(1-3\alpha^2-2\alpha\beta-\beta^2)(1-\beta^2)}(\frac{1}{1-\alpha\beta-\beta^2} + \frac{3\alpha\beta}{1-2\alpha\beta-\beta^2})$$
$$+ \frac{\omega^2\beta}{(1-\alpha-\beta)^2(1-\beta^2)}(\frac{2}{1-\beta} - \frac{\alpha+\beta}{1-\alpha\beta-\beta^2} - \frac{\alpha}{1-2\alpha\beta-\beta^2}),$$

$E = \dfrac{\omega^2}{(1-\beta^2)(1-\alpha\beta-\beta^2)(1-\alpha-\beta)}[\dfrac{(1+\alpha\beta+\beta^2)(1+\alpha+\beta)}{1-3\alpha^2-2\alpha\beta-\beta^2} + \dfrac{2\beta}{1-\beta}]$.

The 'information' measure for $\hat{\beta}$, defined to be the inverse of its variance by Nelson and Startz (2006), is derived as:

$$I_{\hat{\beta}}(\omega,\alpha,\beta) = \frac{T}{I^{-1}(3,3)} = T \cdot \frac{(1-\alpha-\beta)^2}{2\omega^2} \cdot \frac{B^2(2D-C-E) + A(CE-D^2)}{AC-B^2}$$

It is straightforward to show that

$$T \cdot \frac{(1-\alpha-\beta)^2}{2\omega^2} \to T \cdot \frac{(1-\beta)^2}{2\omega^2} \neq 0 \text{ as } \alpha \to 0 \qquad \text{(A.2)}$$

$$AC - B^2 \to \frac{2\omega^2}{(1-\beta)^6(1+\beta)} \neq 0 \text{ as } \alpha \to 0 \qquad \text{(A.3)}$$

However,

$$B^2(2D-C-E) + A(CE-D^2) \to 0 \text{ as } \alpha \to 0 \qquad \text{(A.4)}$$

This completes the proof of (2.4).

Appendix B

Here we impose $\beta = 0$ to demonstrate a few implications of ZILC in the GARCH(1,1):

$$Asy.Var\begin{pmatrix}\hat{\omega}\\ \hat{\alpha}\\ \hat{\beta}\end{pmatrix} = \begin{bmatrix} \frac{\omega^2(1+\alpha)}{\alpha^2(1-\alpha)} & 0 & -\frac{\omega(1-3\alpha^2)}{\alpha^2(1-\alpha)} \\ -- & \frac{(1-3\alpha^2)}{(1+\alpha)(1-\alpha)} & -\frac{(1-3\alpha^2)}{(1+\alpha)(1-\alpha)} \\ -- & -- & \frac{(1-3\alpha^2)}{\alpha^2(1+\alpha)(1-\alpha)} \end{bmatrix} \quad \text{(B.1)}$$

The 'information' measure of $\hat{\beta}$, again, approaches zero as α goes to 0:

$$I_{\hat{\beta}}(\omega,\alpha,\beta) = \frac{\alpha^2(1+\alpha)(1-\alpha)T}{(1-3\alpha^2)} \to 0, \text{as } \alpha \to 0 \quad \text{(B.2)}$$

$\hat{\omega}$ has the same issue as shown in the following:

$$I_{\hat{\omega}}(\omega,\alpha,\beta) = \frac{\alpha^2(1-\alpha)T}{\omega^2(1+\alpha)} \to 0, \text{as } \alpha \to 0 \quad \text{(B.3)}$$

Furthermore, $\hat{\omega}$ and $\hat{\beta}$ are highly negatively correlated when α is small:

$$Asy.Corr(\hat{\omega},\hat{\beta}) = -\sqrt{1-3\alpha^2} \to -1, \text{as } \alpha \to 0 \quad \text{(B.4)}$$

However, $\hat{\alpha}$ is well identified in that its information measure does not converge to zero:

$$I_{\hat{\alpha}}(\omega,\alpha,\beta) = \frac{(1+\alpha)(1-\alpha)T}{(1-3\alpha^2)} \to T \neq 0, \text{as } \alpha \to 0 \quad \text{(B.5)}$$