Essays on Inference in Weakly Identified Models in Macroeconomics and Finance

Jun Ma

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2007

Program Authorized to Offer Degree: Economics

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Jun Ma

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

Charles R. Nelson

Reading Committee:

Charles R. Nelson

Richard Startz

Eric Zivot

Date: 05/23/2007

In presenting this dissertation in partial fulfillment of the requirements for the doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of the dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, 1-800-521-0600, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature ____________

Date 06/05/2007

University of Washington

Abstract

Essays on Inference in Weakly Identified Models in Macroeconomics and Finance

Jun Ma

Chair of the Supervisory Committee:
Professor Charles R. Nelson
Department of Economics

This dissertation is concerned with the implications of weak identification in macroeconomics and finance: the risks of making spurious inferences, strategies for valid inference, and their economic implications. In the first essay I show that the standard estimation and *t*-test in the GARCH(1,1) model are spurious when the GARCH effect is weakly identified, implying strong and significant persistence of volatility when in fact there is little. This spurious inference is partly attributed to the severely under-estimated standard error for the estimated GARCH effect. A strategy for valid inference is suggested and seems to give robust results for this case. In my second essay I derive an analytical asymptotic variance matrix for the GARCH(1,1) Maximum Likelihood Estimator and show that the Zero-Information-Limit Condition (ZILC) of Nelson and Startz (2007) holds, accounting for spuriously large *t*-statistics. In the third essay I propose a general approach to valid inference in weakly identified models based on a common linear approximation and show that this general test strategy succeeds in obtaining a correct size in the presence of weak identification. In the fourth essay I apply this valid test to evaluate a recent resolution of the equity premium puzzle based on a high level of persistence in consumption growth. My results find little empirical evidence in support of this resolution.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

Table Number Page

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

By Jun Ma, Charles R. Nelson and Richard Startz

1.1 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 (2007) 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.

---

* A Paper based on this chapter has been published on the *Studies in Nonlinear Dynamics & Econometrics*: Vol. 11: NO. 1, Article 1.

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 1.2 demonstrates that ZILC holds in the GARCH(1,1) model. Section 1.3 presents evidence by Monte Carlo experiments to document the underestimation of standard errors when identification of the GARCH effect is weak. Section 1.4 proposes the empirical strategy and evaluates its validity. Section 1.5 presents the results for some real datasets. Section 1.6 concludes this paper.

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

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

$$\varepsilon_t = \sqrt{h_t} \cdot \xi_t, \xi_t \sim i.i.d.N(0,1) \tag{1.1}$$

$$h_t = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot h_{t-1} \tag{1.2}$$

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:

[1] The mean of equation (1.1) is set to be zero without loss of generality since the information matrix is block-diagonal, as shown by Bollerslev (1986).

$$\varepsilon_t^{\,2} = \omega + (\alpha+\beta)\cdot\varepsilon_{t-1}^{\,2} + w_t - \beta\cdot w_{t-1} \tag{1.3}$$

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 (2007) 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 (2007) (Chapter 2 of this dissertation) 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 (2007), goes to zero as α approaches zero, i.e., ZILC holds:

$$\lim_{\alpha\to 0} I_{\hat{\beta}}(\omega,\alpha,\beta) = 0 \tag{1.4}$$

Appendix A.1 gives a formal proof of (1.4) based upon Ma's (2007) 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 A.2 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.

1.3 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.

1.3.1 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 (1.1) and (1.2) 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 Bollerslev (1989)

[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.

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 1.1 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).

In Table 1.1, 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 (2007) 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 1.2 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 (2007) for details. This comparison is for the fixed sample size $T = 1000$. 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 (2007) 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 Figure 1.1 from the experiment when $\omega = 1, \alpha = 0.01, \beta = 0$ and sample size $T = 1000$, corresponding to the first three rows in Table 1.2. 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 Figure 1.2. It is evident that there is a strong negative correlation between the absolute value of $\hat{\beta}$ and its standard error. Nelson and Startz (2007) (Chapter 3 of this dissertation) 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 Figure 1.3 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.

1.3.2 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 (1.1) and (1.2) 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 1.3 presents our major findings. No major difference has been found compared with Table 1.1 and 1.2. 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 Figure 1.4) 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.

1.3.3 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 (1.2) to obtain its AR representation:

$$h_t = \omega + (\alpha + \beta)h_{t-1} + \alpha w_{t-1} \qquad (1.5)$$

Where $w_{t-1} = \varepsilon_{t-1}^2 - h_{t-1}$. 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. Figure 1.5 and 1.6 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 1.4 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 1.4 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.

1.3.4 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:

[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).

$$E_t[h_{t+h}] = \omega \sum_{i=0}^{h-1} (\alpha+\beta)^i + (\alpha+\beta)^h h_t \tag{1.6}$$

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: $CONST.h = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2$.

Figure 1.7 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 Figure 1.8 demonstrates the same idea.

We compute the Root Mean Squared Error (RMSE) across MC samples to summarize the predicting accuracy for various methods. Table 1.5 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) $\sum(\varepsilon_t^2 - h_t)^2$:

$$\sum(\varepsilon_t^2 - h_t)^2 = \sum(\hat{h}_t - h_t)^2 + \sum(\varepsilon_t^2 - \hat{h}_t)^2 + 2\sum(\hat{h}_t - h_t)(\varepsilon_t^2 - \hat{h}_t) \qquad (1.7)$$

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.

1.4 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 1.6) 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.

In our approach to GARCH(1,1) estimation, when $\hat{\alpha}$ and the sample size are in the left upper area of Table 1.6 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:

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + w_t \tag{1.8}$$

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

$$\begin{aligned} \varepsilon_t^2 &= \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot h_{t-1} + w_t \\ &= \frac{\omega}{1-\beta} + \alpha \cdot \varepsilon_{t-1}^2 + \alpha\beta \cdot \varepsilon_{t-2}^2 + \ldots + \alpha\beta^{k-1} \cdot \varepsilon_{t-k}^2 + \ldots + w_t \end{aligned} \tag{1.9}$$

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 (1.1) and (1.2) with true parameters values $\omega = 1, \alpha = 0.3, \beta_1 = 0.6$. Given reference Table 1.6, 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 (1.5) 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:

$$(1-\alpha_1 L-\ldots-\alpha_p L^p)(h_t - \frac{\alpha_0}{1-\alpha_1-\ldots-\alpha_p}) = (1-(-\frac{\alpha_2}{\alpha_1})L-\ldots-(-\frac{\alpha_p}{\alpha_1})L^{p-1})\alpha_1 w_{t-1} \tag{1.10}$$

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 Figure 1.9). In the next section we examine some real datasets and experiment with this approach.

1.5 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 1.6 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:

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

[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.

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 (1.1) and (1.2) 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

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

MATLAB

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

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 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):

[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.

$\hat{\alpha}_0 = 0.73 \cdot 10^{-3}$ $\hat{\alpha}_1 = 0.041$ $\hat{\alpha}_2 = 0$ $\hat{\alpha}_3 = 0.019$ $\hat{\alpha}_4 = 0.008$ $\hat{\alpha}_5 = 0.251$

$(0.21 \cdot 10^{-3})$ (0.044) (0.173) (0.085) (0.035) (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 Figure 1.10. 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 Figure 1.11). The PLLF for the GARCH(1,1) estimation is given by Figure 1.12. The PLLF turns out to be bimodal.

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.

1.6 Conclusion

We show that the Zero-Information-Limit Condition (ZILC) formulated by Nelson and Startz (2007) 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.

Table 1.1: 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%

Table 1.2: Estimated Standard Error versus 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

Table 1.3: 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%

Table 1.4: 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%

Table 1.5: 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

Table 1.6: The Reference Table for Practitioners - Empirical Size of t-test for $\hat{\beta}$ in the GARCH(1,1) Model, $\omega = 1, \beta = 0$

Sample	True value of α						
Size	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%	--	--	--

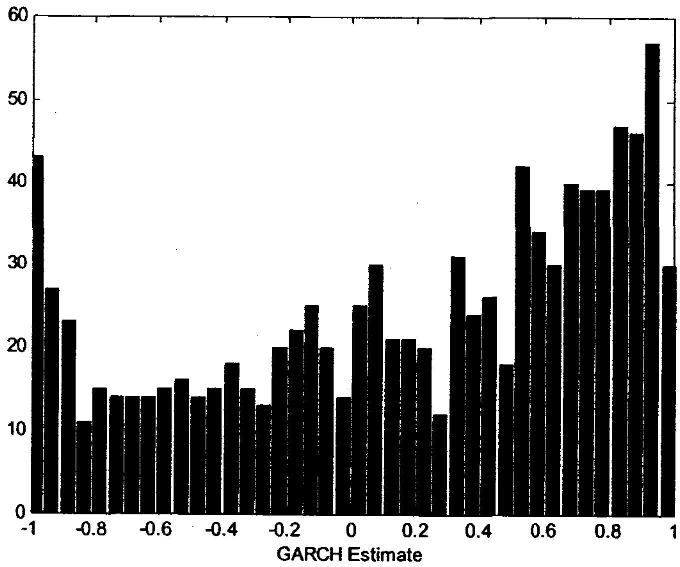


Figure 1.1: Histogram of $\hat{\beta}$ from the MC Experiment of GARCH(1,1) for $\omega = 1, \alpha = 0.01, \beta = 0, T = 1000$

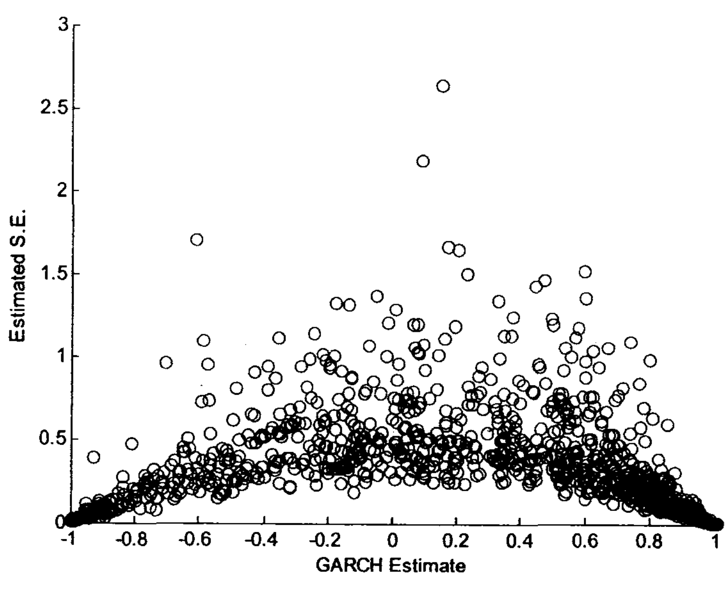
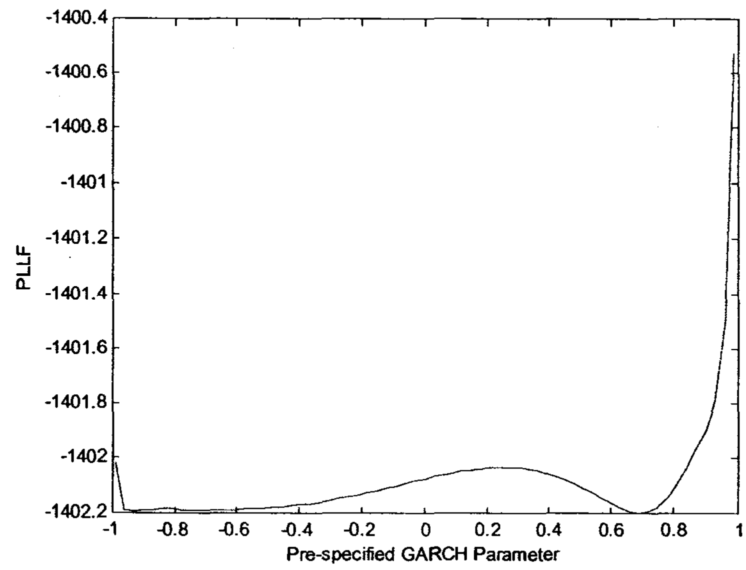


Figure 1.2: Scatter Plot of Estimated S.E. of $\hat{\beta}$ against $\hat{\beta}$ from the MC Experiment of GARCH(1,1) for $\omega = 1, \alpha = 0.01, \beta = 0, T = 1000$



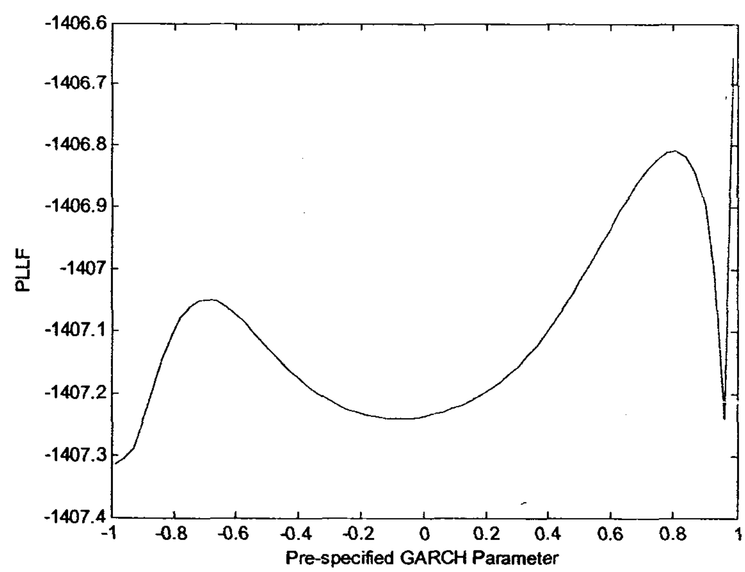


Figure 1.3: Two Examples of Profile LLF from the MC Experiment of GARCH(1,1) for $\omega = 1, \alpha = 0.01, \beta = 0, T = 1000$

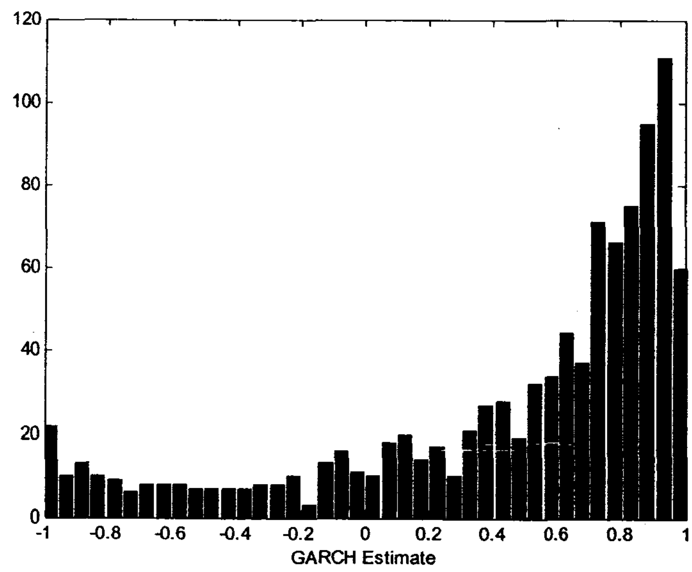


Figure 1.4: Histogram of $\hat{\beta}$ from the MC Experiment of GARCH(1,1) for $\omega = 1, \alpha = 0.01, \beta = 0.5, T = 1000$

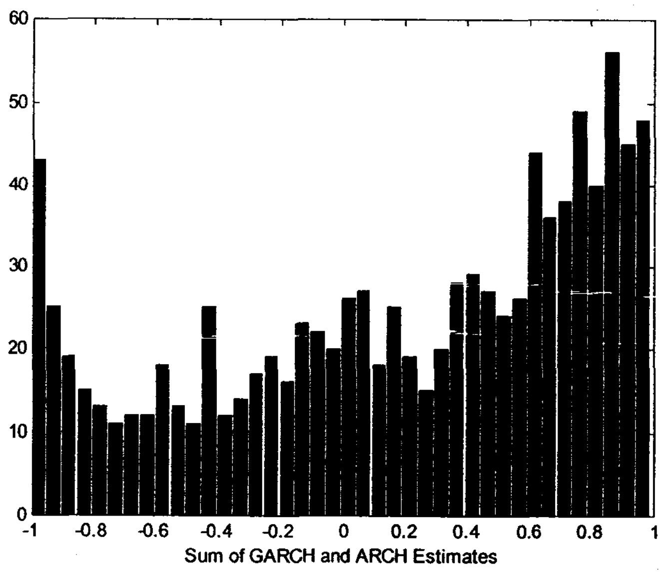


Figure 1.5: Histogram of $(\hat{\alpha} + \hat{\beta})$ from the MC Experiment of GARCH(1,1) for $\omega = 1, \alpha = 0.01, \beta = 0, T = 1000$

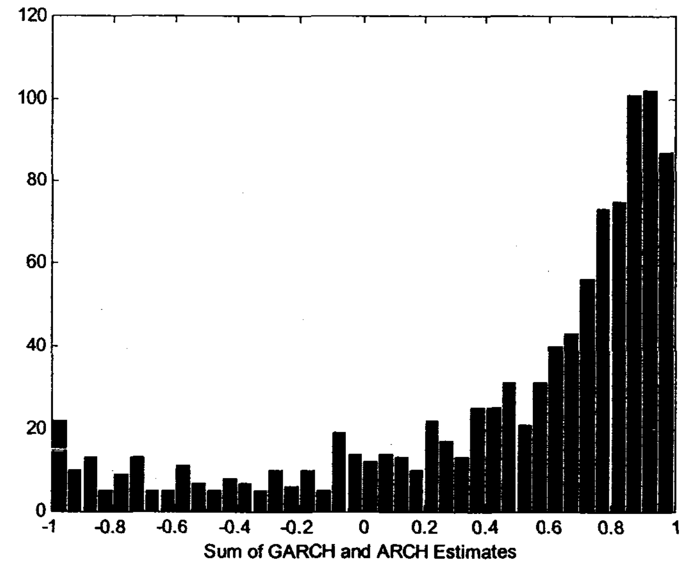


Figure 1.6: Histogram of $(\hat{\alpha}+\hat{\beta})$ from the MC Experiment of GARCH(1,1) for $\omega=1,\alpha=0.01,\beta=0.5,T=1000$

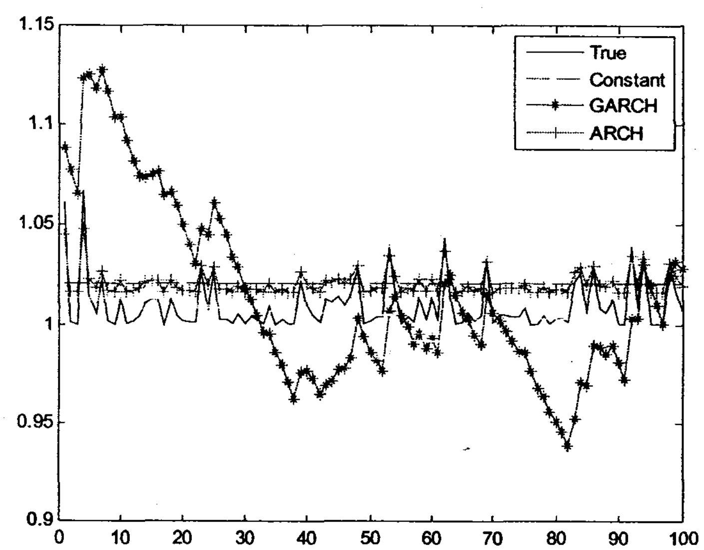


Figure 1.7: A Typical Comparison of the In-Sample Volatility Forecast from the MC Experiment of GARCH(1,1) for $\omega=1,\alpha=0.01,\beta=0,T=1000$

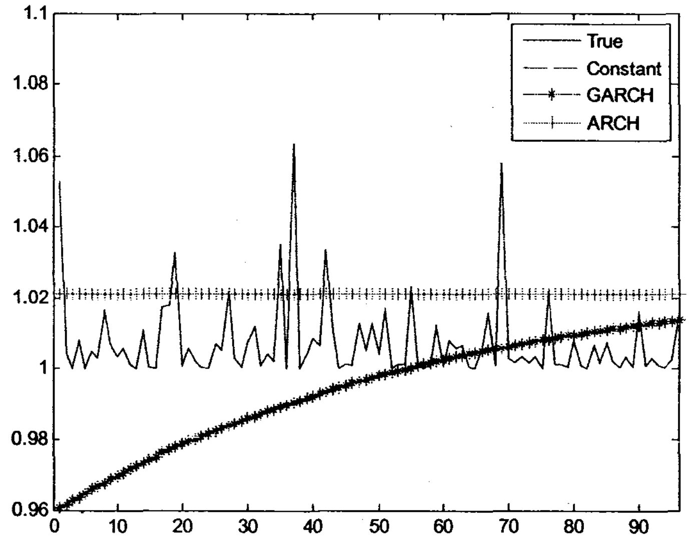


Figure 1.8: A Typical Comparison of the Out-of-Sample Volatility Forecast from the MC Experiment of GARCH(1,1) for $\omega = 1, \alpha = 0.01, \beta = 0, T = 1000$

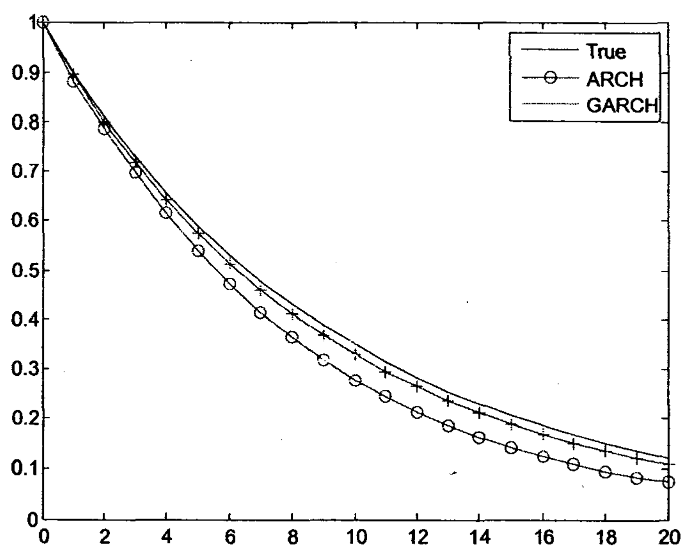


Figure 1.9: The ACF of the Conditional Volatility from the MC Experiment of GARCH(1,1) for $\alpha = 0.3, \beta = 0.6, T = 1000$

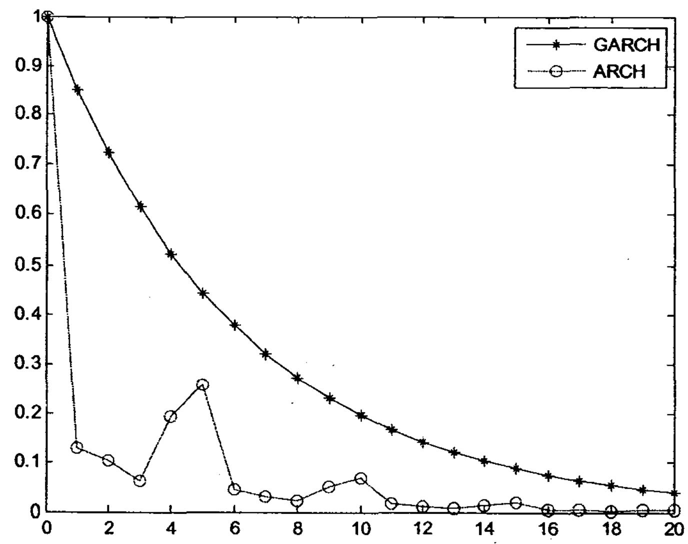


Figure 1.10: The ACF of the Conditional Volatility Implied by GARCH(1,1) and ARCH(5) Estimates for S&P 500 Index Return Data

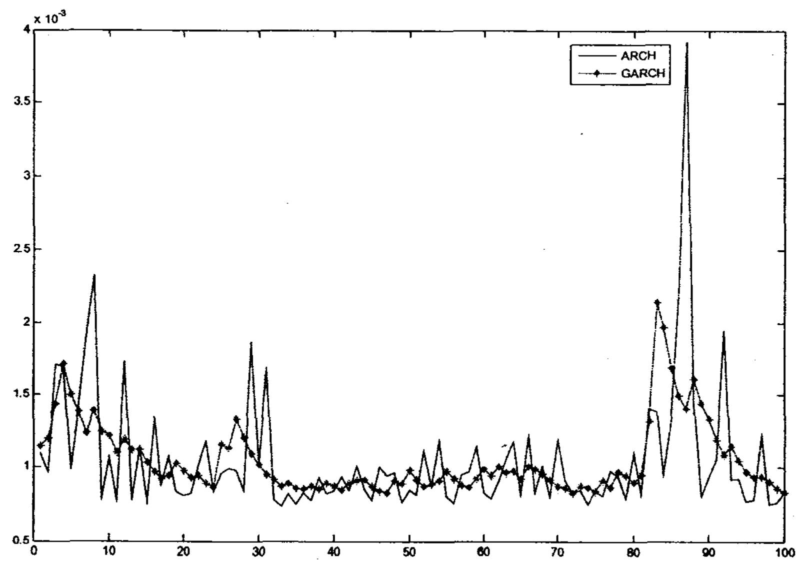


Figure 1.11: The Estimated Conditional Volatility from GARCH(1,1) and ARCH(5) Estimation for S&P 500 Index Return Data
(A typical sub-sample is presented here to facilitate the visualization.)

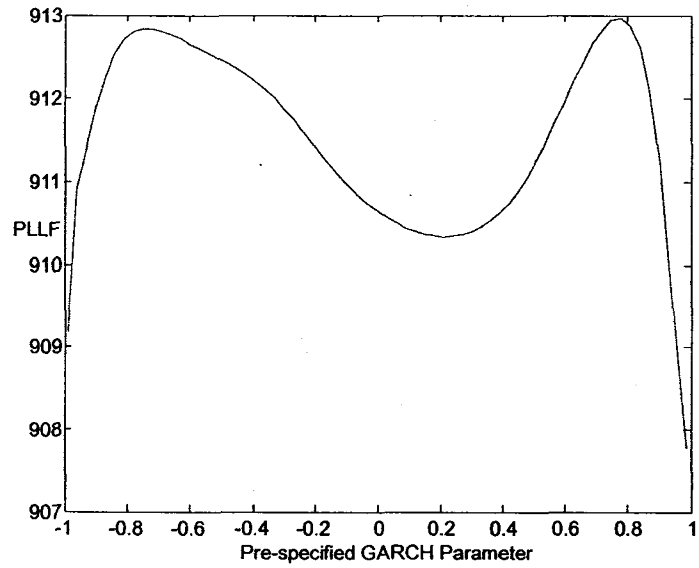


Figure 1.12: Profile LLF of GARCH(1,1) Estimation for S&P 500 Index Return Data

Chapter 2: A Closed-Form Asymptotic Variance-Covariance Matrix for the Maximum Likelihood Estimator of the GARCH(1,1) Model

2.1 Introduction

The GARCH(1,1) model has become a benchmark in modeling time-varying volatility since its introduction by Bollerslev (1986). However, its estimation is usually implemented by numerically maximizing the log-likelihood function and the involved nonlinearity makes both MLE and its asymptotic variance unavailable in a closed form. As a result, different numerical optimization procedures often lead to significantly different values for both MLE and its standard error. Even for the same numerical value of MLE, much different standard errors are frequently returned by various software packages. In Brooks, Burke and Persand (2001) they show that the GARCH(1,1) estimation of daily German mark/British pound exchange rate data returns 0.0725 in MATLAB as the estimated standard error for the GARCH estimate, in a sharp contrast to 0.0166, the one from TSP, given exactly the same GARCH point estimates. Furthermore, the lack of a closed-form asymptotic variance makes it difficult to study the property of GARCH MLE.

This paper works out a closed-form asymptotic variance matrix for GARCH(1,1) MLE in terms of only model parameters so as to provide an analytical formula to compute the standard error for GARCH estimates. This asymptotic formula has the advantage of agreeing upon the values of standard errors given the same point estimates. More importantly, the resulting analytic asymptotic variance shows clearly that the inference of the GARCH parameter depends functionally on the ARCH parameter, approaching zero continuously as the ARCH parameter goes to zero. This means that the GARCH(1,1) model satisfies the Zero-Information-Limit Condition formulated by Nelson and Startz (2007), which makes the inference questionable when the ARCH parameter is small. Ma, Nelson and Startz (2007) (chapter 1 of this dissertation) carefully investigate this issue.

The consistency and asymptotic normality of the GARCH(1,1) MLE have been well established by Bollerslev and Wooldridge (1992), Lee and Hansen (1994), and Lumsdaine (1996), which are discussed in section 2.2. A local approximation is taken in the information matrix to avoid taking the expectation of a nonlinear form. Consequently, the derivation breaks down to the derivations of the auto-covariance and cross-covariance structures for the squared GARCH process, which are derived in section 2.3. This results in a closed-form information matrix in terms of only model parameters. The asymptotic variance-covariance matrix is readily available by taking the inverse of this information matrix. In section 2.4, I carry out a Monte Carlo experiment and show that this formula works very well. Section 2.5 concludes.

2.2 The Asymptotics of GARCH(1,1) MLE

An archetype GARCH(1,1) model may be written as[6]:

$$\varepsilon_t = \sqrt{h_t} \cdot \xi_t, \xi_t \sim N(0,1) \tag{2.1}$$

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

Here, I assume ξ_t is independently drawn from a standard normal distribution. Assuming normality allows me to work on the MLE and have known expressions for the higher moments, which are required for the derivations.

Write up the log-likelihood function:

$$L_T(\theta) = T^{-1} \sum_{t=1}^{T} l_t(\theta) \tag{2.3}$$

$$l_t(\theta) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log h_t - \frac{1}{2}\frac{\varepsilon_t^2}{h_t} \tag{2.4}$$

[6] The mean of equation (1) is set to be 0 without loss of generality since the information matrix is block-diagonal, as shown by Bollerslev (1986).

Where, $\theta = (\omega, \alpha, \beta)'$ and $\hat{\theta}_T$ maximizes the log-likelihood function for a given sample data $\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T\}$. In practice, the evaluation of (2.3) and (2.4) conditions upon an initial assignment of h_0. In spite of various choices of h_0 the difference disappears asymptotically as long as the underlying process is stationary and ergodic.

The Gradient at each time t of the log-likelihood function is:

$$s_t(\theta) = \frac{\partial l_t}{\partial \theta} = \frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}\left(\frac{\varepsilon_t^2}{h_t} - 1\right) \tag{2.5}$$

By law of iterated expectation we have:

$$E[s_t(\theta)] = 0 \tag{2.6}$$

The Hessian at each time t is given by:

$$H_t(\theta) = \frac{\partial^2 l_t}{\partial \theta \partial \theta'} = \left(\frac{\varepsilon_t^2}{h_t} - 1\right)\frac{\partial}{\partial \theta'}\left[\frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}\right] - \frac{1}{2h_t^2}\frac{\partial h_t}{\partial \theta}\frac{\partial h_t}{\partial \theta'}\frac{\varepsilon_t^2}{h_t} \tag{2.7}$$

Again by law of iterated expectation, we have:

$$E[H_t(\theta)] = -\frac{1}{2h_t^2}\frac{\partial h_t}{\partial \theta}\frac{\partial h_t}{\partial \theta'} \tag{2.8}$$

Lumsdaine (1996) proves the consistency and asymptotic normality of the quasi-MLE by assuming a compact and convex parameter space along with a strict stationarity and ergodicity condition for the GARCH(1,1) model, derived in D. Nelson (1990):

$$E[\ln(\beta + \alpha \cdot \xi_t^2)] < 0 \tag{2.9}$$

To have a well-defined finite second moment, I further impose a stronger restriction as in Bollerslev (1986):

$$\alpha + \beta < 1 \tag{2.10}$$

One can easily verify that condition (2.10) along with the normality assumption of ξ_t is sufficient to derive condition (2.9) via Jensen's inequality.

Since ξ_t is normal, the log-likelihood function is correctly specified, which implies the asymptotic result:

$$T^{1/2}(\hat{\theta}_T - \theta_0) \overset{A}{\sim} N(0, I_0^{-1}) \tag{2.11}$$

Where θ_0 is true parameter value; I_0 is the information matrix evaluated at θ_0:

$$I_0 = -E[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}]|_{\theta_0} \tag{2.12}$$

By recursion and assuming the process extends infinitely far into the past, we have the analytical result[7]:

$$\frac{\partial h_t}{\partial \theta} = \begin{bmatrix} \sum_{i=1}^{\infty} \beta^{i-1} \\ \sum_{i=1}^{\infty} \beta^{i-1} \varepsilon_{t-i}^{\ 2} \\ \sum_{i=1}^{\infty} \beta^{i-1} h_{t-i} \end{bmatrix} \tag{2.13}$$

Combine result (2.8), (2.12) and (2.13) to get the symmetric and positive definitive information matrix:

[7]Fiorentini, Calzolari and Panattoni (1996) derive the first and second derivatives in GARCH models, which include this specific result.

$$I=\frac{1}{2}\begin{bmatrix} E\left[\frac{(\sum_{i=1}^{\infty}\beta^{i-1})^2}{h_t^2}\right] & E\left[\frac{\sum_{i=1}^{\infty}\beta^{i-1}\sum_{i=1}^{\infty}\beta^{i-1}\varepsilon_{t-i}^{\ 2}}{h_t^{\ 2}}\right] & E\left[\frac{\sum_{i=1}^{\infty}\beta^{i-1}\sum_{i=1}^{\infty}\beta^{i-1}h_{t-i}}{h_t^{\ 2}}\right] \\ -- & E\left[\frac{(\sum_{i=1}^{\infty}\beta^{i-1}\varepsilon_{t-i}^{\ 2})^2}{h_t^2}\right] & E\left[\frac{\sum_{i=1}^{\infty}\beta^{i-1}\varepsilon_{t-i}^{\ 2}\sum_{i=1}^{\infty}\beta^{i-1}h_{t-i}}{h_t^{\ 2}}\right] \\ -- & -- & E\left[\frac{(\sum_{i=1}^{\infty}\beta^{i-1}h_{t-i})^2}{h_t^2}\right] \end{bmatrix} \tag{2.14}$$

To avoid computing the expectation of a nonlinear form for each element in the information matrix, I take a local approximation in the neighborhood around $\alpha=0$ so as to take the denominator out[8]. Note here the interchange of the limit and expectation operations is valid supposing each element is bounded on the parameter space. I take the element $I(1,1)$ for illustration purpose:

$$\lim_{\alpha\to 0}E\left[\frac{(\sum_{i=1}^{\infty}\beta^{i-1})^2}{h_t^2}\right]=E\left[\frac{\lim_{\alpha\to 0}(\sum_{i=1}^{\infty}\beta^{i-1})^2}{\lim_{\alpha\to 0}h_t^2}\right]=(\lim_{\alpha\to 0}\frac{(1-\alpha-\beta)^2}{\omega^2})\cdot(\lim_{\alpha\to 0}E[(\sum_{i=1}^{\infty}\beta^{i-1})^2]) \tag{2.15}$$

Intuitively, when α is very small, h_t can be approximated by $\omega/_{1-\alpha-\beta}$. In this way, we can deal with only the linear part on the numerators. It is easy to derive the analytical expressions for the numerators of $I(1,1), I(1,2)$ and $I(1,3)$. Next, I show how to derive those of $I(2,2)$, $I(3,3)$ and $I(2,3)$.

---

[8] It is not unusual to see a small $\hat{\alpha}$ in empirical work, which makes this approximation empirically relevant.

2.3 The Derivation of A Closed from Information Matrix

To derive analytical expressions for the numerators of $I(2,2)$, $I(2,3)$ and $I(3,3)$, we need to work out the auto-covariance and cross-covariance structures for $\{\varepsilon_t^2\},\{h_t\}$. To see why, express out the terms:

$$E\left(\sum_{i=1}^{\infty}\beta^{i-1}\varepsilon_{t-i}{}^2\right)^2 = E\left(\sum_{i=1}^{\infty}\beta^{2(i-1)}\varepsilon_{t-i}^4 + 2\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\beta^{i+j-2}\varepsilon_{t-i}^2\varepsilon_{t-j}^2\right) \tag{2.16}$$

$$E\left(\sum_{i=1}^{\infty}\beta^{i-1}h_{t-i}\right)^2 = E\left(\sum_{i=1}^{\infty}\beta^{2(i-1)}h_{t-i}^2 + 2\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\beta^{i+j-2}h_{t-i}h_{t-j}\right) \tag{2.17}$$

$$E\left(\sum_{i=1}^{\infty}\beta^{i-1}\varepsilon_{t-i}{}^2\sum_{i=1}^{\infty}\beta^{i-1}h_{t-i}\right) = E\left[\sum_{i=1}^{\infty}\beta^{2(i-1)}\varepsilon_{t-i}^2h_{t-i} + \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\beta^{i+j-2}(\varepsilon_{t-i}^2h_{t-j} + h_{t-i}\varepsilon_{t-j}^2)\right] \tag{2.18}$$

The derivation of the auto-covariance structures for $\{\varepsilon_t^2\},\{h_t\}$ starts from the ARMA(1,1) representation for the GARCH(1,1) model[9]:

$$\varepsilon_t{}^2 = \omega + (\alpha+\beta)\cdot\varepsilon_{t-1}{}^2 + w_t - \beta w_{t-1} \tag{2.19}$$

Where innovation $w_t = h_t(\xi_t^2 - 1)$ is a Martingale Difference Sequence (MDS):

$$E[w_t \mid I_{t-1}] = 0 \tag{2.20}$$

Where $\{I_t\}$ denotes the information filtration and w_t is adapted to I_t.

Furthermore, notice that $\{h_t\}$ has an AR(1) representation:

$$h_t = \omega + (\alpha+\beta)\cdot h_{t-1} + \alpha w_{t-1} \tag{2.21}$$

[9] Refer to Hamilton (1994) for a standard treatment.

Note that h_t is adapted to I_{t-1}.

The assumption (2.10) implies a finite unconditional variance for ε_t:

$$E[\varepsilon_t^2] = E[h_t] = \frac{\omega}{1-\alpha-\beta} < \infty, \tag{2.22}$$

However, for the existence of a finite fourth moment, we impose one more restriction for parameters as derived in Bollerslev (1988):

$$3\alpha^2 + 2\alpha\beta + \beta^2 < 1 \tag{2.23}$$

Under this restriction, we have:

$$E[\varepsilon_t^4] = 3E[h_t^2] = \frac{3\omega^2(1+\alpha+\beta)}{(1-3\alpha^2-2\alpha\beta-\beta^2)(1-\alpha-\beta)} \tag{2.24}$$

And the following autocorrelations for both $\{\varepsilon_t^2\}$ and $\{h_t\}$:

$$\rho_1^{\varepsilon 2} = \frac{\alpha(1-\alpha\beta-\beta^2)}{1-2\alpha\beta-\beta^2}, \text{ and } \rho_i^{\varepsilon 2} = (\alpha+\beta)^{i-1}\rho_1, i = 2,3,\cdots \tag{2.25}$$

$$\rho_i^h = (\alpha+\beta)^i, i = 1,2,\cdots \tag{2.26}$$

These autocorrelations have also been independently derived in Bollerslev (1988) and Kristensen and Linton (2006). One can derive them following Harvey (1993, Chapter 1). He and Terasvirta (1999) also work out the general fourth moment structure of a squared GARCH process.

Manipulating by the standard formulas for expectation, covariance and variance, summing up the geometric series, and plugging in the expressions for the second and fourth moments, I obtained the closed forms of (2.16) and (2.17) (See Appendix B for details):

$$E\left(\sum_{i=1}^{\infty}\beta^{i-1}\varepsilon_{t-i}^{2}\right)^{2}=\frac{\omega^{2}}{(1-2\alpha\beta-\beta^{2})(1-\alpha-\beta)}\left[\frac{3(1+\alpha+\beta)}{1-3\alpha^{2}-2\alpha\beta-\beta^{2}}+\frac{2\beta}{(1-\beta)^{2}}\right] \tag{2.27}$$

$$E\left(\sum_{i=1}^{\infty}\beta^{i-1}h_{t}\right)^{2}=\frac{\omega^{2}}{(1-\beta^{2})(1-\alpha\beta-\beta^{2})(1-\alpha-\beta)}\left[\frac{(1+\alpha\beta+\beta^{2})(1+\alpha+\beta)}{1-3\alpha^{2}-2\alpha\beta-\beta^{2}}+\frac{2\beta}{1-\beta}\right] \tag{2.28}$$

Lastly, to derive the analytical expression for (2.18), I transform the cross-covariance between $\{\varepsilon_t^2\}$ and $\{h_t\}$ to known auto-covariance of $\{\varepsilon_t^2\}$ and $\{h_t\}$ by taking advantage of the MDS property of w_t (See Appendix for details).

These results, along with the previous work, allow us to derive the following closed form expression for the numerator of $I(2, 3)$:

$$\begin{aligned}
&E(\sum_{i=1}^{\infty}\beta^{i-1}\varepsilon_{t-i}^{2}\sum_{i=1}^{\infty}\beta^{i-1}h_{t-i})\\
&=\frac{\omega^{2}(1+\alpha+\beta)}{(1-\alpha-\beta)(1-3\alpha^{2}-2\alpha\beta-\beta^{2})(1-\beta^{2})}\left(\frac{1}{1-\alpha\beta-\beta^{2}}+\frac{3\alpha\beta}{1-2\alpha\beta-\beta^{2}}\right)\\
&+\frac{\omega^{2}\beta}{(1-\alpha-\beta)^{2}(1-\beta^{2})}\left(\frac{2}{1-\beta}-\frac{\alpha+\beta}{1-\alpha\beta-\beta^{2}}-\frac{\alpha}{1-2\alpha\beta-\beta^{2}}\right)
\end{aligned} \tag{2.29}$$

To finish this section, I list the result for numerators of $I(1,1)$, $I(1,2)$ and $I(1,3)$:

$$E[(\sum_{i=1}^{\infty}\beta^{i-1})^{2}]=\frac{1}{(1-\beta)^{2}} \tag{2.30}$$

$$E[\sum_{i=1}^{\infty}\beta^{i-1}\sum_{i=1}^{\infty}\beta^{i-1}\varepsilon_{t-i}^{2}]=\frac{1}{(1-\beta)^{2}}\cdot\frac{\omega}{1-\alpha-\beta} \tag{2.31}$$

$$E[\sum_{i=1}^{\infty}\beta^{i-1}\sum_{i=1}^{\infty}\beta^{i-1}h_{t-i}]=\frac{1}{(1-\beta)^2}\cdot\frac{\omega}{1-\alpha-\beta} \tag{2.32}$$

The above derivations result in a closed-form information matrix, taking the inverse of which, I obtain the asymptotic variance-covariance matrix[10].

2.4 MONTE CARLO SIMULATION EXPERIMENTS

To evaluate how well this closed-form expression works, I carry out a sequence of Monte Carlo simulation experiments. The sample size is fixed at $T = 1000$ and nine sets of parameter values have been chosen for empirical interests:

$$\begin{pmatrix}\omega\\ \alpha\\ \beta\end{pmatrix}=\begin{pmatrix}1\\ 0.05\\ 0\end{pmatrix},\begin{pmatrix}1\\ 0.10\\ 0\end{pmatrix},\begin{pmatrix}1\\ 0.05\\ 0.2\end{pmatrix},\begin{pmatrix}1\\ 0.10\\ 0.2\end{pmatrix},\begin{pmatrix}1\\ 0.05\\ 0.5\end{pmatrix},\begin{pmatrix}1\\ 0.10\\ 0.5\end{pmatrix},\begin{pmatrix}1\\ 0.05\\ 0.8\end{pmatrix},\begin{pmatrix}1\\ 0.10\\ 0.8\end{pmatrix}$$

The information matrix based upon the simulations is computed by averaging the realized information matrices across the simulated data paths for each set of parameter values. The number of simulations is set to be 100,000. Therefore, resulting variance matrix is almost the true one with little MC variation. Table 2.1 gives the comparison for the variance-covariance matrix and show that my analytical formula works fairly well, especially when α is small, regardless of the magnitude of β.

2.5 CONCLUSION

In this paper, I gave an analytical formula to compute the standard errors for GARCH estimates. The Monte Carlo simulation experiments demonstrate that this formula works well, especially when α is small. This formula can be used in practice to settle any discrepancy of estimated standard errors from various software packages. It also shows analytically that the GARCH(1,1) model satisfies

[10] Due to the lengthy algebra, the result is not displayed here but is available from the author upon request.

Zero-Information-Limit Condition defined by Nelson and Startz (2007), i.e., the information for the GARCH estimate $\hat{\beta}$ approaches zero as the ARCH parameter α goes to zero. ZILC implies that the information of $\hat{\beta}$ tends to be overestimated when α is small and the routine test will fail to report a correct size, as shown to be true in Ma, Nelson and Startz (2007).

Table 2.1. Comparison of the Asymptotic Variance-Covariance Matrix (Total difference is the sum of the absolute difference)

Parameter Values		Closed-Form Expression Variance Matrix			S.E.	Numerical Evaluation Variance Matrix			S.E.	Difference of S.E. in %	Total Difference
ω	1	0.3990	0.0000	-0.3772	**0.6317**	0.4418	-0.0003	-0.4177	**0.6647**	5.23%	
α	0.05		0.0009	-0.0009	**0.0300**		0.0014	-0.0009	**0.0370**	23.32%	33.78%
β	0			0.3592	**0.5993**			0.3977	**0.6307**	5.23%	
ω	1	0.0990	0.0000	-0.0873	**0.3146**	0.1224	-0.0004	-0.1080	**0.3499**	11.20%	
α	0.1		0.0008	-0.0008	**0.0282**		0.0017	-0.0009	**0.0412**	46.11%	68.49%
β	0			0.0794	**0.2817**			0.0981	**0.3132**	11.19%	
ω	1	0.6007	0.0047	-0.4536	**0.7750**	0.6427	0.0062	-0.4868	**0.8017**	3.44%	
α	0.05		0.0010	-0.0045	**0.0316**		0.0014	-0.0058	**0.0369**	16.77%	23.94%
β	0.2			0.3447	**0.5871**			0.3709	**0.6090**	3.73%	
ω	1	0.1631	0.0024	-0.1147	**0.4039**	0.1855	0.0037	-0.1321	**0.4307**	6.64%	
α	0.1		0.0010	-0.0026	**0.0316**		0.0017	-0.0039	**0.0412**	30.26%	44.64%
β	0.2			0.0830	**0.2881**			0.0963	**0.3104**	7.74%	

Table 2.1 Continued

ω	1	0.7215	0.0112	-0.3347	**0.8494**	0.8180	0.0146	-0.3818	**0.9044**	6.48%	
α	0.05		0.0009	-0.0059	**0.0300**		0.0013	-0.0077	**0.0354**	17.93%	31.47%
β	0.5			0.1566	**0.3957**			0.1795	**0.4237**	7.06%	
ω	1	0.1930	0.0054	-0.0814	**0.4393**	0.2403	0.0085	-0.1038	**0.4902**	11.57%	
α	0.1		0.0009	-0.0031	**0.0300**		0.0015	-0.0047	**0.0391**	30.26%	55.74%
β	0.5			0.0356	**0.1887**			0.0462	**0.2149**	13.91%	
ω	1	0.4996	0.0093	-0.0837	**0.7068**	0.5741	0.0118	-0.0976	**0.7577**	7.19%	
α	0.05		0.0005	-0.0019	**0.0230**		0.0007	-0.0024	**0.0266**	15.93%	31.71%
β	0.8			0.0145	**0.1203**			0.0171	**0.1307**	8.59%	
ω	1	0.1413	0.0038	-0.0171	**0.3759**	0.1795	0.0060	-0.0239	**0.4236**	12.71%	
α	0.1		0.0004	-0.0008	**0.0208**		0.0008	-0.0013	**0.0280**	34.85%	68.21%
β	0.8			0.0025	**0.0503**			0.0037	**0.0606**	20.65%	

Chapter 3: Valid Inference under Weak Identification in Models Where the Zero-Information-Limit-Condition Holds

By Jun Ma and Charles R. Nelson

3.1 Introduction

A number of econometric models of importance in practice have representations of the form

$$y_i = \gamma \bullet g(\beta, x_i) + \lambda' z + \varepsilon_i, \gamma \neq 0 \tag{3.1}$$

where γ and β are scalar parameters, *y*, *x* and **z** are data, $\boldsymbol{\lambda}$ is a vector of regression coefficients, and ε_i is a homoskedastic error with variance σ^2. Examples include certain non-linear regression models such as production functions, the Phillips curve model of Staiger, Stock and Watson (1997), and, perhaps less obviously, ARMA and GARCH models, and Unobserved Component models used to separate trend and cycle. Note that γ controls the amount of information that the data contain about β that is identified only if γ is non-zero. Models of this form satisfy the 'Zero Information Limit Condition' (hereafter, ZILC) of Nelson and Startz (2007) which requires that the reciprocal of the asymptotic variance of ML estimator $\hat{\beta}$, denoted by $I_{\hat{\beta}}$, approach zero as γ approaches a critical value. Suppressing **z** for simplicity, one readily obtains the following expression for the information that the data contain about

$$I_{\hat{\beta}} = V_{\beta}^{-1} = \frac{\gamma^2}{\sigma^2} \bullet \frac{[\sum g_i^2 \bullet \sum (g_i^{'})^2 - (\sum g_i \bullet g_i^{'})^2]}{\sum g_i^2}; \gamma \neq 0 \tag{3.2}$$

where g_i denotes $g(\beta, x_i)$ and g_i' its first derivative with respect to β. Clearly, $I_{\hat{\beta}}$ goes to zero as γ approaches zero, and thus ZILC holds for (1.1). NS show that when ZILC holds inference based on the Wald statistic $t = (\hat{\beta} - \beta) \bullet \hat{I}_{\hat{\beta}}^{0.5}$ is problematic for two reasons. One is that estimated information tends to be upward biased, the relative bias being larger the closer is γ to the ZILC point. The other is that the numerator and denominator of the *t*-statistic are not independent as in classical regression, but instead are functionally related. Consequently, although estimated standard errors for $\hat{\beta}$ are typically too small, the associated *t*-statistic may be either too large or too small, depending on the correlation between g_i and g_i'. In this paper we consider only models that are identified; nevertheless, spurious inference is a problem for values of γ that are economically far from zero.

The purpose of this paper is to study a strategy for obtaining a valid test statistic for β based on the expansion of $g(\beta, x_i)$ around $\beta = \beta_*$ which gives

$$y_i = \gamma \bullet (g(\beta_*, x_i) + (\beta - \beta_*) \bullet g'(\beta_*, x_i)) + e_i \qquad (3.3)$$

where e_i may incorporate a remainder term. In some models of practical importance $g(.)$ and $g'(.)$ are simply data, for example the Phillips curve of Staiger, Stock and Watson (1997) where y is the change in inflation, $g = (x_i + \beta)$ where x is the unemployment rate and β is the unknown natural rate, and g' is simply one. To illustrate the potential for spurious inference, a simulated Phillips curve with independent standard normal x_i and ε_i, 100 observations, and $\gamma = .01$, the estimated standard error for $\hat{\beta}$ from the non-linear regression routine in EViews™ has a median of about 2 compared to the asymptotic value of 10. Paradoxically, the associated '*t*-statistics' are also too small, indeed no rejections at a nominal .05 level occur in 1000 replications! As mentioned above, NS traced this apparent paradox to the dependence between the estimated standard error and the estimation error $(\hat{\beta}-\beta)$ in models where ZILC holds.

In the Phillips curve context an exact test of the null hypothesis $\beta=\beta_0$ is easily obtained by making the substitution $\beta=\beta_0+\delta$ in the model to obtain the linear regression $y_i=\gamma\bullet(x_i+\beta_0)+6+\varepsilon_i$ where $6=\gamma\bullet\delta$. Since the null hypothesis implies that $\theta=0$ we test that null hypothesis and obtain a classical *t*-test with exact size. More generally, we may expand g(.) around the value $\beta_*=\beta_0$ and obtain the linear approximation:

$$y_i=\gamma\bullet g(\beta_0,x_i)+(\gamma\bullet(\beta-\beta_0))\bullet g'(\beta_0,x_i)+e_i \qquad (3.4)$$

Thus, a test of the null hypotheses $\beta=\beta_0$ is obtained by regressing *y* on *g*(.) and its first derivative (evaluated under the null hypothesis), then testing whether the coefficient of the latter differs significantly from zero. The intuition is that if the null is true then the first term captures the entire influence of β on the data and the second term should be irrelevant, but if the null is wrong the second term gives an indication of how influential on the model are departures from the null. In general (3.4) is only an approximation - as for non-linear regression, Unobserved Components model, ARMA and GARCH models - so the actual size of the *t*-test is a question to be investigated.

Section 3.2 presents the results of this investigation for a non-linear regression model, and for ARMA (1,1), GARCH(1,1), and an Unobserved Components model. Section 3.3 concludes.

3.2. Valid Inference in Four Weakly Identified Models

3.2.1. A Nonlinear Regression Model

We have shown above, using the Phillips Curve example, that the expansion of *g*(.) is exact and so is the proposed test when the functional *g*(.) is linear in parameters. There are, however, other models which take a direct form of (3.1) with, generally, nonlinear *g*(.), such as the Hicks-neutral Cobb-Douglas production function we consider here:

$$y_i = \gamma \cdot x_i^{\beta} + \varepsilon_i ; \gamma \neq 0 \tag{3.5}$$

Where, y_i and x_i are per capita output and per capita capital input respectively; γ, in its economic term, represents the technology component or, more broadly, the so-called Total-factor productivity (TFP); econometrically, γ controls the amount of information data contains about β, the capital share of output.

Apparently this model satisfies the ZILC condition in NS and as a result, the inference based on the Wald *t*-statistic is not correct due to both its understated denominator and the dependence between its numerator and denominator. To illustrate the spurious inference in this model we generate a path of x_i from the log-normal distribution and pair it with 1000 paths of simulated standard normal ε_i, both of 100 observations to compute y_i with parameter values $\gamma = 0.01, \beta = 0.5$. The estimated standard error for $\hat{\beta}$ from the non-linear regression routine in EViews™ has a median of about 0.98 compared to the asymptotic value of 6.47. This underestimated standard error together with the dependence between the numerator and denominator of the Wald *t*-statistic gives about 0.11 rejection frequency for β at the nominal level of 0.05. This is just another example where ZILC results in a failure of the Wald *t*-statistic.

To construct a valid test we may expand *g*(.), which is x_i^{β} in this case, around the null $\beta = \beta_0$, giving rise to the following regression:

$$y_t = \gamma \bullet x_i^{\beta_0} + \lambda \bullet x_i^{\beta_0} \log(x_i) + e_i \tag{3.6}$$

Where $\lambda = \gamma \bullet (\beta - \beta_0)$. Here the expansion is not exact but only approximate to the first order as g(.) is not linear in β, and e_i may include a remainder term. However, the (3.6) is a classical regression once the regressors $[x_i^{\beta_0}, x_i^{\beta_0} \log(x_i)]$ are evaluated at the null β_0 and the *t*-statistic for λ is straightforward to compute and the inference of it no longer functionally depends upon any other unknown parameter. Using the same simulated dataset in the above MC experiment, we find that the *t*-statistic for testing the null $\lambda = 0$ gives an almost perfect empirical size of 0.051 at a nominal level of 0.5.

3.2.2. The ARMA Model with Near Cancellation

Among models satisfying the ZILC condition there are some taking the form (3.1) in a less obvious way. One simple example is the ARMA(1,1) model, known as, perhaps, the most parsimonious way to capture the serial correlation:

$$\begin{aligned} &(1-\phi L)y_t = (1-\theta L)\varepsilon_t; t = 1,\dots,T \\ &\varepsilon_t \sim i.i.d.N(0,\sigma_\varepsilon^2), |\phi| < 1, |\theta| < 1 \end{aligned} \tag{3.7}$$

Where ϕ is the AR coefficient, θ is the MA coefficient. To relate the ARMA(1,1) model with form (1.1), we may multiply both sides by $(1-\theta L)^{-1}$ and expand terms out to get:

$$y_t = \gamma \bullet g(\theta, \bar{y}_{t-1}) + \varepsilon_t \tag{3.8}$$

Where, $\gamma = \phi - \theta$, $g(\theta, \vec{y}_{t-1}) = \sum_{i=1}^{\infty} \theta^{i-1} y_{t-i}$ and $\vec{y}_{t-1} = (y_{t-1}, y_{t-2}, \ldots)$. NS show that when γ is small the standard error for either $\hat{\phi}$ or $\hat{\theta}$ is severely underestimated and this downward bias of the estimated standard error is fairly strong even after taking into account the stationarity and invertibility requirement, resulting in a much overstated size of the Wald *t*-statistic. The other important driving force of size distortion is the dependence between the point estimate and its standard error, as shown for $\hat{\theta}$ here.

The concern of parameter redundancy in ARMA models traces back to Box and Jenkins (1970) in which they urge people to scrutinize the necessity of higher ARMA terms since obviously the order of an ARMA model can be increased arbitrarily by adding the AR and MA terms with roughly the same root without violating the restrictions imposed by real data. Although this concern was usually relieved by the observation that this type of parameter redundancy typically makes little difference as far as the forecasting is concerned, the significant difference will instead be made in terms of its economic implications. For example, denote the GDP growth rate at time *t* by y_t and assume y_t follows an ARMA(1,1) process:

$$y_t = \mu + \phi \bullet y_{t-1} + \varepsilon_t - \theta \bullet \varepsilon_{t-1} \tag{3.9}$$

Where, μ is the average growth rate. The implied expectation for the growth rate will then be governed by an AR(1) process:

$$E[y_{t+1} \mid I_t] = \mu + \phi \bullet E[y_t \mid I_{t-1}] + \gamma \bullet \varepsilon_t \tag{3.10}$$

Where $\gamma = \phi - \theta$ determines the size of underlying shock to the expectation process, and a higher ϕ implies any current shock to the economy lasts longer, increasing the risk level of economy. The reliance on such a high level of persistence, so called long run risk, to generate interesting results in a general equilibrium just starts to become popular recently, for example see Bansal and Lundblad (2002), Bansal and Yaron (2004), and Gavin, Keen, and Pakko (2006). However, from a statistical point of view, when the underlying shock has a small size as is true in many cases, one might need to be very careful to draw any conclusion for the persistence level solely based on the, likely wrong, routine inference.

To investigate how the proposed valid test performs for, say, θ in this model, expand *g*(.) in (3.8) around the null θ_0:

$$y_t = \gamma \bullet g(\theta_0, \bar{y}_{t-1}) + \lambda \bullet g_\theta(\theta_0, \bar{y}_{t-1}) + e_t \tag{3.11}$$

Where $g_\theta(\theta, \bar{y}_{t-1}) = \frac{\partial g(\theta, \bar{y}_{t-1})}{\partial \theta} = \sum_{i=2}^{\infty} (i-1) \bullet \theta^{i-2} y_{t-i}$, and $\lambda = \gamma \bullet (\theta - \theta_0)$. Testing the null $\lambda = 0$ will be equivalent to testing the null $\theta = \theta_0$.

In practice, to evaluate the regressors $[\, g(\theta_0, \bar{y}_{t-1}), g_\theta(\theta_0, \bar{y}_{t-1})\,]$, we cannot observe y_t back into the infinite past, a standard technique is to set $y_t = 0$ for all $t \le 0$. This makes negligible difference as long as the stationarity and invertibility conditions are satisfied. Again (3.11) mimics a classical regression and the inference for λ is exact in finite sample given normal shocks and can be well approximated otherwise. To see how this test corrects the spurious inference resulting otherwise from the routine Wald *t*-test, we simulate 1000 data paths each of sample size T = 100 with $\phi = 0.01, \theta = 0, \sigma_\varepsilon = 1$ in Eviews™. Although the usual inference based on the Wald *t*-test for θ fails to cover the true value about 50.1% of the time, our proposed linear test based on (3.11) does correct for this spurious inference and gives 4.9% rejection frequency, roughly its nominal level 5%. Notice here since $\theta = 0$ testing $\lambda = 0$ is simply testing the second lag in the AR(2) model, which is equivalent to the Q-test with one lag for the residuals from the AR(1) estimation. In general, this equivalence is not necessarily true though.

Despite the brevity of testing MA root θ, a valid test for $\phi = \phi_0$ requires some extra efforts. To see why, duality gives a representation of the ARMA(1,1) similar to (3.8) but parameterized in terms of γ and ϕ:

$$y_t = \gamma \bullet g(\phi, \vec{\varepsilon}_{t-1}) + \varepsilon_t \tag{3.12}$$

Where the functional form $g(.)$ remains the same but the argument becomes the unobserved variable ε instead of data y : $g(\phi, \vec{\varepsilon}_{t-1}) = \sum_{i=1}^{\infty} \phi^{i-1} \varepsilon_{t-i}$ and $\vec{\varepsilon}_{t-1} = (\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$.

However, as long as a consistent estimate for ε can be found, the valid test for $\lambda = 0$ in the expansion of (3.13) is still feasible to compute in the following regression:

$$y_t = \gamma \bullet g(\phi_0, \vec{\varepsilon}_{t-1}) + \lambda \bullet g_\phi(\phi_0, \vec{\varepsilon}_{t-1}) + e_t \tag{3.13}$$

Where $g_\phi(\phi, \vec{\varepsilon}_{t-1}) = \dfrac{\partial g(\phi, \vec{\varepsilon}_{t-1})}{\partial \phi} = \sum_{i=2}^{\infty} (i-1) \bullet \phi^{i-2} \varepsilon_{t-i}$, and $\lambda = \gamma \bullet (\phi - \phi_0)$.

One way to consistently estimate ε so as to evaluate the regressors in (3.13) is to estimate the model by imposing the null $\phi = \phi_0$ first, and then compute $\tilde{\varepsilon}$ using the estimates. If the null is true, $\tilde{\varepsilon}$ will be a good approximate for ε. The other equivalent way to implement the test is to replace $\tilde{\varepsilon}$ with restricted estimates and data y via inherent transformation. Several steps of manipulation results in (see Appendix

C.1 for details):

$$y_t = \gamma \bullet g(\tilde{\theta}, \bar{y}_{t-1}) + \lambda \bullet [\tilde{\gamma}^{-1} g(\phi_0, \bar{y}_{t-1}) - \tilde{\gamma}^{-1} g(\tilde{\theta}, \bar{y}_{t-1})] + e_t \tag{3.14}$$

Where, $\tilde{\gamma}$ and $\tilde{\theta}$ are the estimates from ARMA(1,1) subject to the restriction $\phi = \phi_0$.

Both approaches give the same numerical test statistic as confirmed in the MC experiment implemented. Data is generated in Eviews™ with true parameter values $\gamma = 0.01, \phi = 0, \sigma_\varepsilon = 1$ and sample size T = 100, and the proposed test statistic rejects the null 54 times in 1000 replications, roughly at its nominal level 0.05.

This valid test based on the expansion of functional $g(.)$ is straightforward to be generalized to deal with a potential spurious inference in estimating an ARMA model of higher order with near-identical roots. With multiple AR and MA roots an ARMA model is weakly identified if any pair of AR and MA roots is sufficiently close to each other.

For illustration purpose, consider an ARMA(2,1) model with real roots:

$$\begin{aligned} &(1-\phi_1 L)(1-\phi_2 L) y_t = (1-\theta L)\varepsilon_t ; t = 1,\ldots,T \\ &\varepsilon_t \sim i.i.d.N(0,\sigma_\varepsilon^2), |\phi_1| < 1, |\phi_2| < 1, |\theta| < 1 \end{aligned} \tag{3.15}$$

When either one of the two AR roots is close to the MA root or both are, this model is weakly identified and the usual Wald t-test is problematic. For example, if data is generated with parameter values $\phi_1 = 0.5, \phi_2 = 0.01, \theta = 0, \sigma_\varepsilon = 1$ of sample size T = 100, t-test for the MA root from routine estimation has an empirical size as large as 52.4% at a nominal level 5% with 1000 replications in Eviews™.

To implement the valid test strategy, first write (3.15) into the general form:

$$(1-\phi_1 L)y_t = (1-\phi_1 L)[\gamma \bullet g(\theta, \vec{y}_{t-1})] + \varepsilon_t \tag{3.16}$$

Where $g(.)$ is defined as above and $\gamma = \phi_2 - \theta$. Take an expansion of $g(.)$ around the null to get, after several steps of manipulation:

$$y_t = \alpha_1 \bullet [(1-\theta_0 L)^{-1} y_{t-1}] + \alpha_2 \bullet [(1-\theta_0 L)^{-1} y_{t-2}] + \lambda \bullet [(1-\theta_0 L)^{-2} y_{t-3}] + e_t \tag{3.17}$$

Where, $\alpha_1 = \gamma + \phi_1$, $\alpha_2 = -\phi_1 \bullet (\theta_0 + \gamma)$, and $\lambda = \phi_1 \bullet \gamma \bullet (\theta - \theta_0)$. Again the valid test statistic for the null $\theta = \theta_0$ is the *t*-stat for λ in this linear regression (3.17). Just to find out how this test performs in finite sample, we simulate data of sample size T = 100 with $\phi_1 = 0.5, \phi_2 = 0.01, \theta = 0, \sigma_\varepsilon = 1$ in Eviews™. Our proposed linear test based on (3.17) gives 5.2% rejection frequency for testing the null $\theta = 0$ in contrast to the large size distortion based on the usual Wald *t*-stat. Notice here since $\theta = 0$ testing $\lambda = 0$ is simply testing the third lag in the AR(3) model, as is true for the simple ARMA(1,1) model. Likewise, to test the AR root, take advantage of duality and use the restricted estimate $\tilde{\varepsilon}$, just as have shown for the simple ARMA(1,1) model.

3.2.3. The Unobserved Component Model for Trend and Cycle Decomposition

The decomposition of real output into trend and cycle has been of a great interest. However, in literature two widely used parametric methods, namely Beveridge-Nelson decomposition (hereafter BN) by Beveridge and Nelson (1981) and the Unobserved Component model (hereafter UC) proposed by Harvey (1985) and Clark (1987), produce surprisingly different estimates for trend and cycle. BN decomposition typically attributes most output variation to trend but UC model usually concludes with much larger amplitude of cycle. Morley, Nelson and Zivot (2002) provide an insightful reconciliation of this seemingly difference. Here we provide another potential resolution of this puzzle based upon our ZILC findings.

In a UC model, real output is explicitly expressed as a sum of the trend and cycle components which are treated as state variables, the Kalman filter is adopted to obtain the maximum likelihood estimate and the trend and cycle can be extracted afterwards based on the estimated hyper-parameters. Consider the following simple UC setting-up for modeling the real GNP y_t :

$$y_t = \tau_t + c_t \tag{3.18}$$

$$\tau_t = \tau_{t-1} + \mu + \eta_t, \eta_t \sim i.i.dN(0, \sigma_\eta^2) \tag{3.19}$$

$$(1 - \phi L)c_t = \varepsilon_t, \varepsilon_t \sim i.i.dN(0, \sigma_\varepsilon^2) \tag{3.20}$$

Where τ_t and c_t are unobserved trend and cycle, respectively. Trend is simply a random walk with a constant drift (drift itself may be random as well) and the cycle is AR(1). Nelson (1988), in a MC experiment, shows that even if there is no cycle variation ($\sigma_\varepsilon^2 = 0$), the UC model assigning most variation to the cycle appears to fit better, which is just another example of the Dickey-Fuller (1979) Phenomenon.

Here, we are concerned with the case when cycle variation is not strictly zero but small, i.e., $\sigma_\varepsilon^2 > 0$ but small. We show below that the inference of parameter estimate $\hat{\phi}$ depends on σ_ε^2 and is problematic when σ_ε^2 is close to zero.

The above UC model implies a reduced-form ARIMA(1,1,1) for y_t:

$$(1-\phi L)\Delta y_t = \mu + (1-\phi L)\eta_t + \varepsilon_t - \varepsilon_{t-1} = \mu + u_t - \theta u_{t-1} \quad (3.21)$$

Where, the identification maps the above parameters in the following way:

$$\gamma_0 = (1+\phi^2)\sigma_\eta^2 + 2\sigma_\varepsilon^2 + (1+\phi)\sigma_{\eta\varepsilon} = (1+\theta^2)\sigma_u^2 \quad (3.22)$$

$$\gamma_1 = -\phi\sigma_\eta^2 - \sigma_\varepsilon^2 - (1+\phi)\sigma_{\eta\varepsilon} = -\theta\sigma_u^2 \quad (3.23)$$

As clear from the above, there are more structural parameters than in the reduced form. Therefore one restriction is required to identify the model. Typically it is assumed that the trend innovation and cycle innovation are uncorrelated ($\sigma_{\eta\varepsilon} = 0$). Imposing this restriction and along with the invertibility requirement, we solve for the unique θ:

$$\theta = \frac{(1+\phi^2)\sigma_\eta^2 + 2\sigma_\varepsilon^2 - \sqrt{[(1+\phi)^2\sigma_\eta^2 + 4\sigma_\varepsilon^2]\bullet[(1-\phi)^2\sigma_\eta^2]}}{2\phi\sigma_\eta^2 + 2\sigma_\varepsilon^2} \quad (3.24)$$

It is straightforward to see that $\phi - \theta \to 0$ as $\sigma_\varepsilon^2 \to 0$. Comparing this to the ARMA case discussed in the above section well indicates that ZILC holds in this model with the variance of cycle innovation controlling the amount of information data contains for cycle amplitude estimate $\hat{\phi}$. If σ_ε^2 is indeed small, little information is available for estimating ϕ which may lead to a wide confidence interval and the resulting uncertainty of filtered cycle estimate should be fairly large if one accounts for both the filter uncertainty and the parameter uncertainty as in Hamilton (1986). However, our ZILC findings suggest that in this situation the standard error for $\hat{\phi}$ may be severely underestimated and the resulting uncertainty band might be severely underestimated.

To illustrate a potential spurious inference in the scenario as discussed above, we design a MC experiment with parameters $\mu = 0.8, \phi = 0, \sigma_\eta^2 = 0.95, \sigma_\varepsilon^2 = 0.05$ and sample size $T = 200$. Notice here the cycle innovation is small ($\sigma_\varepsilon^2 = 0.05$) and the true cycle amplitude is even zero, so most of the resulting output variation is due to the trend component. We generate 1000 replications from the model (3.18)-(3.20) with the above parameter values. Using routine estimation algorithm based on the Kalman filter, the hyper-parameters are estimated and reported along with their estimated standard error. Similar to the ARMA case, a standard t-test for ϕ rejects the null about 47.5% of the time and there is even an unusual upward bias for $\hat{\phi}$ (see Figure 3.1).

ZILC does not provide an explanation for this upward bias; but if one believes continuity is not a bad assumption here, the Dickey-Fuller phenomenon (see Nelson (1988)) might provide a good intuition. ZILC, however, indeed predicts the underestimated standard error for $\hat{\phi}$: the median of estimated standard error for $\hat{\phi}$ in the MC experiment is merely 0.30 compared with its asymptotic value 1.48. The upward bias of $\hat{\phi}$ and its underestimated standard error may lead people to believe the cycle amplitude is large with quite confidence while actually it is zero.

This paper strives to provide a general approach to constructing a valid inference in such a situation. The essential idea is to avoid the functional dependence of the information for $\hat{\phi}$ on the other parameter, say σ_ε^2. Although the estimation of a UC model involves more complications, the comparison with the ARMA model suggests a shortcut of implementing the general valid test strategy in the following steps: first impose the null $\phi = \phi_0$ and estimate all other parameters; secondly, take advantage of the inherent restriction (3.24) to compute the implied restricted estimate $\tilde{\theta}$; lastly, resort to (3.14) based on the reduced representation (3.21) of the UC component model to test the restriction $\phi = \phi_0$. Using the same simulated dataset as in the above MC experiment, this valid test strategy rejects the null about 44 times in the 1000 replications, roughly at the nominal level 0.05.

3.2.4. The GARCH(1,1) Model with a Small ARCH Effect

Ma, Nelson and Startz (2007) show that ZILC issue is also contagious to the GARCH (Generalized-Autoregressive-Conditional-Heteroskedasticity) model, the first order of which being perhaps one of the most popular approaches in capturing time-varying volatility for time series data. Despite a great deal of rich extensions, the archetypal GARCH(1,1) model may be written:

$$\varepsilon_t = \sqrt{h_t} \cdot \xi_t, \xi_t \sim i.i.d.N(0,1) \tag{3.25}$$

$$h_t = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot h_{t-1} \tag{3.26}$$

Although quite a few papers have established the consistency and asymptotic normality for the GARCH MLE (see Bollerslev and Wooldredge (1992), Lee and Hansen (1994), Lumsdaine (1996) etc.), there has not yet been a closed-form asymptotic variance matrix available in literature until Ma (2007) derived one based upon a local approximation. Ma's result demonstrates that ZILC holds in the GARCH(1,1) model and the inference of β, the so-called GARCH effect, depends on parameter α and is problematic when α is small. An ARMA(1,1) representation of the GARCH(1,1) model proves helpful:

$$\varepsilon_t^2 = \omega + (\alpha + \beta) \cdot \varepsilon_{t-1}^2 + w_t - \beta \cdot w_{t-1} \tag{3.27}$$

where the innovation $w_t = \varepsilon_t^{\,2} - h_t = h_t(\xi_t^2 - 1)$ is a Martingale Difference Sequence (MDS) with a time-varying volatilty. Ma, Nelson and Startz (2007) find that when α is small, corresponding to a near-cancellation of AR and MA roots, the standard error for $\hat{\beta}$ is underestimated and the dependency between the point estimate and estimated standard error reinforces the delusion of a significantly strong GARCH effect even when there is very little, and furthermore that there is an unusual upward bias of $\hat{\beta}$ indicating a concentration of $\hat{\beta}$ around a value greater than the true β, a robust result after carefully dealing with potential bimodality of the likelihood function. Their investigation into the real data sets reveals that this issue is fairly relevant and justifies further endeavors for an approach to correct for this type of spurious inference.

The valid test strategy proposed above may be extended to this not quite standard context where the expansion has to be made in the variance term. For example, to test the null hypothesis $\beta = \beta_0$, first realize (3.26) implies:

$$h_t = \frac{\omega}{1-\beta} + \alpha \cdot g(\beta, \vec{\varepsilon}_{t-1}^{\,2}) \tag{3.28}$$

Where $g(\beta, \vec{\varepsilon}_{t-1}^{\,2}) = \sum_{i=1}^{\infty} \beta^{i-1} \varepsilon_{t-i}^2$ and $\vec{\varepsilon}_{t-1}^{\,2} = (\varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots)$. Take an expansion of $g(.)$ around the null and the complete model is given:

$$\varepsilon_t = \sqrt{h_t} \cdot \xi_t, \xi_t \sim i.i.d. N(0,1) \tag{3.25}$$

$$h_t = c + \alpha \cdot g(\beta_0, \vec{\varepsilon}_{t-1}^{\,2}) + \lambda \cdot g_\beta(\beta_0, \vec{\varepsilon}_{t-1}^{\,2}) + \text{remainder} \tag{3.29}$$

where, $c = \frac{\omega}{1-\beta}$, $\lambda = \alpha \bullet (\beta - \beta_0)$, and $g_\beta(\beta, \vec{\varepsilon}_{t-1}^{\,2}) = \sum_{i=2}^{\infty} (i-1) \cdot \beta^{i-2} \varepsilon_{t-i}^2$. As a result, the t-stat for $\lambda = 0$ in (3.29) is equivalently to test the original null $\beta = \beta_0$. To see how this test makes correction, we simulate data of sample size T = 1000 with $\omega = 1, \alpha = 0.01, \beta = 0$ and the t-stat for $\lambda = 0$ reports 76 rejections among 1000 replications in EViews™, in contrast to as large as 598 times of rejections based on a routine t-stat for β. Notice since the expansion is on the variance, the t-stat for λ is no longer in a classical linear regression context as for ARMA case, which results in a notable deterioration of its size but well within tolerance of usual standard.

Very often it is the sum $\alpha + \beta$ that is of a great practical interest, see e.g., Bansal and Yaron (2004), since the implied volatility h_t is governed by a particular AR(1) process, with $\alpha + \beta$ being the persistence measure:

$$h_t = \omega + (\alpha + \beta) \cdot h_{t-1} + \alpha \cdot w_{t-1} \tag{3.30}$$

Unfortunately, when α is small the estimated persistence $\hat{\alpha} + \hat{\beta}$ shares the same problem as that of $\hat{\beta}$, that is, an upward bias along with an underestimated standard error, leading to an artificially persistent pattern for h_t fluctuating a great deal more than the true one.

Since $\alpha + \beta$ corresponds to the AR root in (3.27) a comparison with ARMA model reveals that a valid test for it inevitably requires one extra step due to the fact that the expansion will be based on an evaluation using the unobserved variable. Denote $\alpha + \beta$ by ρ and (3.26) may be equivalently written as:

$$h_t = \frac{\omega}{1-\rho} + \alpha \bullet g(\rho, \vec{w}_{t-1}) \tag{3.31}$$

Where, $\vec{w}_{t-1} = (w_{t-1}, w_{t-2},...)$. And an expansion of g(.) around the null $\rho = \rho_0$ is:

$$h_t = \frac{\omega}{1-\rho} + \alpha \cdot g(\rho_0, \vec{w}_{t-1}) + \lambda^* \cdot g_\rho(\rho_0, \vec{w}_{t-1}) + \text{remainder} \tag{3.32}$$

Where, $\lambda^* = \alpha \bullet (\rho - \rho_0)$. Again to compute the test statistic for $\lambda^* = 0$, one can first estimate the GARCH(1,1) model with the restriction $\alpha + \beta = \rho_0$ and extract $\widetilde{h}_t$ using the restricted estimates to obtain a restricted estimate of w_t. Alternatively, we can re-parameterize (3.32) using the inherent transformation to get (Appendix C.2 gives details):

$$\varepsilon_t = \sqrt{h_t} \cdot \xi_t, \xi_t \sim i.i.d.N(0,1) \tag{3.25}$$

$$h_t = c^* + \alpha \bullet g(\widetilde{\beta}, \vec{\varepsilon}^{\,2}_{t-1}) + \lambda^* \bullet [\widetilde{\alpha}^{-1} g(\rho_0, \vec{\varepsilon}^{\,2}_{t-1}) - \widetilde{\alpha}^{-1} g(\widetilde{\beta}, \vec{\varepsilon}^{\,2}_{t-1})] + \text{remainder} \tag{3.33}$$

To test the null hypothesis $\rho = \rho_0$, one just need to evaluate the pseudo-regressors in (3.33) using restricted estimates $\widetilde{\alpha}, \widetilde{\beta}$ along with the null ρ_0 and re-estimate the model to compute a *t*-stat for $\lambda^* = 0$. We simulate data of sample size $T = 1000$ with $\omega = 1, \alpha = 0.01, \beta = 0$ and the *t*-stat for $\lambda^* = 0$ reports 69 rejections among 1000 replications, in contrast to 42.7%, the empirical size based on the routine *t*-test.

3.3 Conclusion

This paper shows that a number of important econometric models, seemingly unrelated, share a common issue in testing a parameter of interest, derived from the fact that the inference of it depends functionally upon another identifying parameter, the ZILC phenomenon. In all cases where ZILC applies, the information is overestimated and the resulting standard t-test fails to report the correct size. We suggest a general approach to obtaining a valid test in a linear approximation of the original model. By showing how to implement this valid test strategy in various cases, we present the evidence that this test performs fairly well generally, able to correct for the spurious inference which would appear otherwise if based upon the standard t-test.

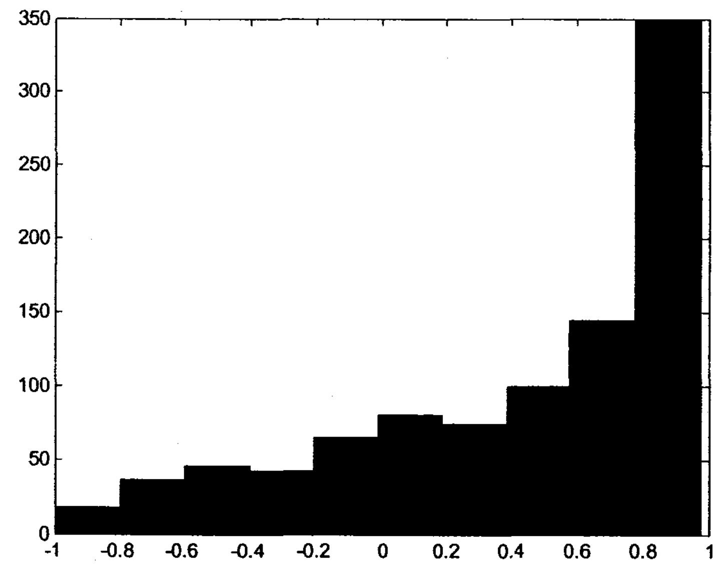


Figure 3.1 Histogram of $\hat{\phi}$ from the MC experiment of UC Component Model for $\mu = 0.8, \phi = 0, \sigma_\eta^2 = 0.95, \sigma_\varepsilon^2 = 0.05$

Chapter 4: Consumption Persistence and the Equity Premium Puzzle: A Resolution or Not?

4.1 Introduction

It has remained a major challenge to address too high observed equity premium given consumption properties, leading to so-called "equity premium puzzle" as pointed out by Mehra and Prescott (1985). Recently, Bansal and Yaron (2000, 2004) propose a resolution based on a small but highly persistent component in consumption named the "long-run risk", along with the Epstein-Zin (1989, 1991) recursive utility function. Furthermore, Bansal and Lundblad (2002) show that this long-run risk can also successfully resolve the "volatility puzzle", a phenomenon found by Shiller (1981) and LeRoy and Porter (1981) that equity prices are much more volatile than their dividends. Despite success in explaining these asset pricing anomalies, the long-run risk in consumption is hard to detect empirically as made clear by Hansen, Heaton and Li (2005). Cochrane (2006) expresses his concern by stating "...without strong direct evidence for the required long run properties of consumption growth, the conclusions will always be a bit shaky..." This paper aims to evaluate this resolution of the equity premium puzzle by directly addressing Cochrane's concern if there is "strong direct evidence" for the highly persistent component in consumption, in light of a recent finding about the potential spurious inference in weakly identified models by Nelson and Startz (2007), Ma, Nelson and Startz(2007), and Ma and Nelson (2007). It will be shown that the model adopted to identify the high level of persistence in consumption growth expectation and volatility processes is weakly identified. The resulting spurious inference might account for the apparent resolution.

Numerous efforts have been made to tackle the equity premium puzzle, for example, the habit-formation utility function by Constantinides (1990), Abel (1990), Campbell and Cochrane (1999), etc. However, Mehra and Prescott (2003) point out the "effective" risk aversion in these resolutions are still unreasonably high. Weil (1989) finds that the Epstein-Zin utility function is not much helpful to address the equity premium puzzle given very little predictability in consumption growth[11]. Note that Weil (1989) models consumption growth using a two-state Markov-switching (MS) process which implies only one period momentum in consumption predictability, in contrast to the continuous ARMA-GARCH framework in Bansal and Yaron (2000, 2004) which has a potential of capturing a small but persistent predictable component and the claimed high level of persistence in consumption growth expectation and its volatility processes is the key to their resolution.

In Section 4.2, I restate that a high level of persistence in consumption, the long-run risk, has a potential of resolving the equity premium puzzle and it appears to be empirically robust based on the routine estimation method. In section 4.3, I show that the relevant model is weakly identified and the resulting spurious inference might over-estimate the persistence size with too tight a confidence interval. With the help of a valid inference strategy proposed by Ma and Nelson (2007) (Chapter 3 of this dissertation), I correct the confidence interval for persistence level in this scenario and my result casts a doubt on the existence of such long-run risk. Section 4.4 concludes.

4.2 How Persistence in Consumption Resolves the Puzzle

The analytical demonstration in this section is largely based on Bansal and Yaron (2000), in which they model the consumption growth process as an ARMA(1,1) process:

$$(g_{t+1} - \mu) = \phi \cdot (g_t - \mu) + \eta_{t+1} - \upsilon \cdot \eta_t \tag{4.1}$$

where g_t is real per capita consumption growth (continuously compounding) and η_t is

[11] One applauded merit of Epstein-Zin utility function is that it allows for a separation of the risk aversion and IES (Intertemporal Elasticity of Substitution) which are controlled by one single parameter in a typical CRRA (Constant Relative Risk Aversion) utility function.

a serially uncorrelated normal innovation but with time-varying conditional volatility governed by GARCH(1,1) (see Bollerslev (1986)):

$$\begin{aligned} &\eta_{t+1} \sim N(0, h_t) \\ &h_t = \omega + \alpha \cdot \eta_t^2 + \beta \cdot h_{t-1} \end{aligned} \tag{4.2}$$

It is helpful to write out the implied state-space form of this specification:

$$\begin{aligned} &(x_t - \mu) = \phi \cdot (x_{t-1} - \mu) + \tau \cdot \eta_t \\ &g_{t+1} = x_t + \eta_{t+1} \sim N(0, h_t) \end{aligned} \tag{4.3}$$

$$\begin{aligned} &h_t = \omega + \rho \cdot h_{t-1} + \alpha \cdot w_t \\ &\eta_{t+1}^2 = h_t + w_{t+1} \end{aligned} \tag{4.4}$$

where $x_t = E_t[g_{t+1}]$ is the expectation of the next period's growth rate conditional upon the available information at time *t*. Clearly, the AR root ϕ determines how long any current shock to the conditional expectation persists, and $\tau = \phi - \upsilon$ determines the size of the underlying shock. Furthermore, the conditional volatility is allowed to be time-varying, as opposed to Mehra and Prescott (1985), Weil (1989) and Cecchetti, Lam and Mark (1990). Apparently, $\rho = \alpha + \beta$ measures how long any current shock to the volatility process lasts and α determines the size of the underlying shock $w_t = \eta_t^2 - h_{t-1}$. Using dividend data, Bansal and Yaron (2000) estimate both persistence measures ϕ and ρ to be very large with a fairly tight confidence interval. Bansal and Yaron (2004) calibrate a similar model to consumption and dividend data by assigning very large values to both ϕ and ρ. Along with the Epstein-Zin recursive utility, the model generates a sizable equity premium consistent with the historical level with risk aversion parameter no higher than 10, the upper bound imposed by Mehra and Prescott (1985).

To continue the illustration, I lay out the analytical solution here but leave a sketch of their derivations in the appendix D[12]:

$$E[r_m - r_f] = [(1-\theta)P^2 + \frac{\theta}{\psi}P]\cdot\sigma_\eta^2 + [(1-\theta)\kappa_1^2 Q^2]\cdot\alpha^2\sigma_w^2 - 0.5\cdot\sigma_m^2 \qquad (4.5)$$

where ψ denotes the IES, γ the risk aversion, and $\theta = \frac{1-\gamma}{1-1/\psi}$; constants $P = 1+\kappa_1\tau\cdot\frac{1-1/\psi}{1-\kappa_1\phi}$ and $Q = \frac{\theta(1-1/\psi)^2(1-\kappa_1(\phi-\tau))^2}{2(1-\kappa_1\phi)^2(1-\kappa_1\rho)}$; κ_1 is an approximation constant smaller than but close to 1. Here, r_m is the market portfolio return and r_f is the risk free rate. From (4.5), the equity premium has three contributions: compensation for the risk involving σ_η^2, the unconditional variance of the innovation in consumption growth level, compensation for the risk associated with σ_w^2, the unconditional variance of the innovation in consumption volatility, and lastly the convexity adjustment term involving σ_m^2, the unconditional variance of the market portfolio return which can be further given by:

$$\sigma_m^2 = P^2\cdot\sigma_\eta^2 + \kappa_1^2 Q^2\cdot\alpha^2\sigma_w^2 \qquad (4.6)$$

Equation (4.5) allows us to study the impact of persistent components to the equity premium. The appendix D demonstrates that the multiplier term before σ_η^2 is increasing in P for the considered range of preference parameters. As a result, by just

---

[12] Readers interested in more detailed derivations are referred to Bansal and Yaron (2000, 2004). They solve the model both numerically and analytically but show there is no significant difference.

increasing ϕ one can easily match any equity premium level. Similarly, increasing the persistence level ρ in the volatility process helps increase the premium as well.

There are two points worth emphasizing. First, the Epstein-Zin recursive utility specification is necessary to allow the persistence to have a significant impact. In the standard CRRA specification, where $\gamma \cdot \psi = 1$ or $\theta = 1$, the time-variation of volatility would not be priced and no level equity premium comparable with the historical level can be generated with a reasonably low risk aversion[13]. Second, if the level of persistence is not high enough, the Epstein-Zin utility specification generates no much higher equity premium than the standard CRRA utility specification, which is what Weil (1989) concluded. In the extreme case when consumption growth is *i.i.d*, the equity premium is simply: $\gamma \cdot \sigma_\eta^2$, exactly the same result as would be with the CRRA specification. Therefore, correctly estimating persistence measures is critical. From a statistical point of view, this is more about a correct inference than one single point estimate.

Next, let us turn to a study of consumption data in search for an empirical evidence of the long-run risk. There are a few reasons why I want to focus on the consumption data instead of equity cash flows: first, note most classical literature on the equity premium puzzle does so, see Mehra and Prescott (1985) and Weil (1989) etc.; secondly, for the model considered in this resolution, consumption and dividends share the same persistent component and the differentiation of them considered by Bansal and Yaron (2004) is not essential to the resolution; lastly, dividends are not accurately measured since observed dividends typically include only cash dividends and have to ignore the other forms of payments. Furthermore, it is not trivial to apply an appropriate seasonal adjustment procedure to raw dividends data without introducing artificial serial correlation.

The quarterly real per capita personal consumption expenditure of nondurable goods and service from 1947I to 2005IV is obtained from the Bureau of Economic Analysis (BEA). The frequency of quarter is chosen to have a larger sample size than that of the

[13] See Appendix D for algebraic details.

annual data, while avoiding the seasonality issue of monthly data, as discussed by Wilcox (1992). The consumption growth rate is annualized by continuously compounding. The average growth rate is 2.1% and the annualized standard deviation is 4.0%, both slightly higher than 1.8% and 3.6%, the numbers in Mehra and Prescott (1985) for the sample period of 1889 – 1978.

First an ARMA(1,1) is estimated and the point estimates along with their White Heteroskedasticity-Consistent standard errors are reported below:

$$\hat{\mu} = 0.021(0.002) \qquad \hat{\phi} = 0.78(0.15) \qquad \hat{\tau} = 0.15(0.06)$$

Notice that τ seems to be significantly greater than 0. Formally, the hypothesis test for $H_0 : \tau = 0$ is nonstandard since the AR and MA roots are not identified under the null. I implement the sup*LR* test proposed by Andrews and Ploberger (1996) and obtain a test statistic of 16.37, far exceeding the 5% critical value. This is well expected since the first autocorrelation for consumption growth is 0.23, significantly positive. More importantly, the persistence measure has a point estimate 0.78 with a 95% confidence interval [0.58, 1.08], implying a fairly high level of persistence.

Fitting the residuals from the above estimation into the GARCH(1,1) model, I report the point estimates together with the Bollerslev-Wooldrige robust standard errors[14]:

$$\hat{\omega} = 1.43\times10^{-5}(1.40\times10^{-5}) \qquad \hat{\rho} = 0.96(0.04) \qquad \hat{\alpha} = 0.11(0.05)$$

First, there is clearly some serial dependency in the conditional volatility and the first autocorrelation of the squared residuals is 0.22. Since a test for the hypothesis $\alpha = 0$ is nonstandard, I implement the sup*LR* test proposed by Andrews (2001) and the null is

[14] I separately estimate the level and volatility processes to facilitate ensuing analyses and this procedure is robust to a potential misspecification of GARCH as we will challenge its validity in the next section.

rejected significantly at the 5% level. Secondly, the persistence measure ρ has a point estimate as high as 0.96 with a 95% confidence interval [0.88, 1.04]. So there also seems to be a large and significant GARCH effect, resulting in an extremely high persistence level in volatility.

Given that the confidence interval for persistence level identified above implies the existence of a highly persistent component in consumption, the so-called long-run risk, it is not surprising to find that a relatively high premium is requested by the agent. Using the above point estimates, an equity premium of 6.2%, comparable with the historical level, can be generated by assigning the following values of preference parameters: $\gamma = 10, \psi = 2$ [15]. Each Risk involving level risk and volatility risk contributes about half respectively. Furthermore, a risk free rate as low as about 0.8% can also be simultaneously replicated by setting the discount factor $\delta = 0.966$.

Table 4.1 describes the variation of equity premium at different persistence levels, for fixed preference parametersat at $\gamma = 10, \psi = 2$. Two features are sparkling: first, equity premium is increasing in persistence level in a much faster pace as one moves into the area of higher level of persistence: in the area of low level of persistence, even an increase as large as 0.6 in both level and volatility persistence increases the equity premium only marginally (by merely 0.53%); by contrast in the area of high level of persistence any small increase in persistence triggers a fairly large jump in the equity premium. This observation may well be aligned with the analytical solution (4.5) where the equity premium is not only increasing but convex in both persistence measures. Secondly, the equity premium tends to explode if either the level or volatility turns out to be non-stationary.

From another perspective, Table 4.2 shows values of risk aversion parameter required to match the historical equity premium level at about 6% corresponding to

[15] The risk aversion parameter is set at the upper bound imposed by Mehra and Prescott (1985); IES is set at a value well within the reasonable range as estimated by Hansen and Singleton (1984). Note, however, that there has been extensive debates on estimating IES, see e.g. a recent study by Yogo (2004) which claims that IES is rather far smaller than 1. It is important to note that the resolution discussed here depends critically upon that the IES is greater than 1.

various persistence levels within the 95% confidence interval identified above. If there is no persistence at all, γ needs to be as high as 40. But even for the lower bound here, γ is reduced greatly to 21.

These two exercises illustrate that a very high level of persistence helps resolve the equity premium puzzle and, if the confidence intervals for ϕ and ρ identified above are correct, the puzzle would not seem as firm as before. Unfortunately, as we scrutinize the above ARMA(1,1) and GARCH(1,1) estimation result, both $\hat{\tau}$ and $\hat{\alpha}$ are small, which suggests that the models are weakly identified. In the next section, I will show this may result in spurious inference for ϕ and ρ and we need to address this issue using a valid inference strategy.

4.3 Obtaining a Valid Inference for Persistence Measures

Nelson and Startz (2007) show that when identification of one parameter is conditional on another inference for the former will be misleading if the Zero-Information-Limit Condition (ZILC) holds. The ARMA(1,1) is illustrated as one example where the standard error for either ϕ or υ tends to be understated when the identifying parameter $\tau = \phi - \upsilon$ is small. Ma, Nelson and Startz (2007) further find that ZILC holds in GARCH(1,1) as well and the standard error for either $\rho = \alpha + \beta$ or β tends to be underestimated when the identifying parameter α is small. What is even worse is that the estimated standard error is typically negatively correlated with the point estimate, which might reinforce the illusion of a very high level of persistence while actually there is very little.

To give a fresh picture about how misleading the routine estimation could be when models are weakly identified, I present two simple Monte Carlo (MC) experiments here, for ARMA(1,1) and GARCH(1,1) respectively. Interesting readers are advised to obtain more details in Nelson and Startz (2007), and Ma, Nelson and Startz (2007). In the first experiment, I simulate 1000 paths of data $\{g_t\}$ according to equation (4.1) with parameter values $\phi = 0, \upsilon = -0.15$ [16], implying no persistence at all in the growth expectation, with a simplification of standard normal η_t and $\mu = 0$. The sample size $T = 236$ is chosen to match that of the quarterly consumption data. The coverage probability of 95% confidence intervals for ϕ constructed from the standard estimation algorithm turns out to be only 72.6%, with a MC variation of 1.4%. Similarly, I carry out the second experiment for GARCH(1,1). 1000 paths of data $\{\eta_t\}$ are simulated according to equation (4.2) with sample size T = 236 and parameters $\omega = 1, \alpha = 0.11, \beta = 0$, implying there is only a modest persistence ($\rho = 0.11$). The coverage probability of 95% confidence intervals for ρ is 49.6% with a MC variation of 1.6%, even much worse than the ARMA case. Furthermore, the point estimate $\hat{\rho}$ has a large mass near unity and its median is as high as 0.7, the phenomenon of an upward bias which seems unique to GARCH model.

The results from MC experiments show that, in presence of weak identification, even if the underlying true process does not embody a high level of persistence or the long-run risk, the routine estimation algorithm, however, might find one and tends to overstate its significance. A pertinent question is how to obtain a valid inference for the persistence measures in the presence of weak identification? Built upon Nelson and Startz's (2007) work, Ma and Nelson (2007) suggest taking a linear approximation of the original nonlinear model and implementing a *t*-test in it as a valid inference strategy when the model is weakly identified. Their work discusses in detail why and how this works nicely for a general class of models including ARMA and GARCH. As we will see shortly, this test can indeed obtain a correct size while a routine *t*-test fails to do so when the model is weakly identified.

---

[16] Nelson and Startz (2006), and Ma, Nelson and Startz (2006) find that the identifying parameter is not subject to ZILC and thus is well estimated. Therefore, we use the estimated value as a good approximation of the underlying true identifying parameter to generate the data.

To implement this test strategy for the persistence measure ϕ, write out a general form for ARMA(1,1):

$$\tilde{g}_t = \tau \cdot f(\phi, \vec{\eta}_{t-1}) + \eta_t \tag{4.7}$$

where $\tilde{g}_t = g_t - \mu$, $f(\phi, \vec{\eta}_{t-1}) = \sum_{i=1}^{\infty} \phi^{i-1} \eta_{t-i}$, $\vec{\eta}_{t-1} = (\eta_{t-1}, \eta_{t-2}, \ldots)$, and η_t is uncorrelated but allows for a possible but unknown heteroskedastic structure. In the first step, the ARMA(1,1) is estimated by imposing the null $H_0 : \phi = \phi_0$, resulting in restricted estimates $\tilde{\upsilon}$ and $\tilde{\tau} = \phi_0 - \tilde{\upsilon}$. In the second step, a linear approximation is taken, leading to a linear regression with transformed regressand and regressors evaluated at $\tilde{\upsilon}$ and ϕ_0:

$$\tilde{g}_t = \tau \bullet f(\tilde{\upsilon}, \vec{g}_{t-1}) + \lambda \bullet [f(\phi_0, \vec{g}_{t-1}) - f(\tilde{\upsilon}, \vec{g}_{t-1})] + res \tag{4.8}$$

Where, $\lambda = \tau \bullet (\phi - \phi_0)$, $\vec{g}_{t-1} = (g_{t-1}, g_{t-2}, \ldots)$. The resulting valid test for the null hypothesis $H_0 : \phi = \phi_0$ is a classical t-test for $\lambda = 0$ in the linear regression (4.8). Intuitively, if the null is true the first term in (4.8) should be enough to summarize the serial correlation in data.

To see if this test can correct for the spurious inference issue, I compute the test statistic for each sample data in the first MC experiment. The coverage probability of 95% confidence intervals for the GNR t-test is 95.3%, very close to its nominal level but in sharp contrast to the severe size distortion of the routine t-test.

I apply this valid inference strategy to consumption data and compute the test robust to unknown heteroskedasticity structure by adopting White's (1980) consistent variance-covariance matrix. Inverting this test numerically, I obtain Figure 4.1 which gives a valid 95% confidence interval for the persistence measure $\phi \in$ (-1, 0.98), much wider than the one obtained in the above section through routine estimation: [0.58, 1). This finding is consistent with Dufour's (1997) econometric insight that Wald-based test statistic is problematic in weakly identified model and for locally almost unidentified (LAU) parameters valid confidence intervals must have a non-zero probability of being unbounded. In particular, the valid interval includes 0 but significantly excludes very high level of persistence. Note that in the consumption growth data only the first auto-correlation seems significant and large in magnitude, being about 0.2, fairly close to the "Working" effect identified by Working (1960), i.e., the artificial serial correlation due to aggregation,. This suggests that the small momentum in the aggregate consumption data might be nothing but a statistical artifact.

Likewise, we can implement this valid test strategy for the persistence measure ρ in the GARCH(1,1) model. In the first step, GARCH(1,1) is estimated by imposing the null $\rho = \alpha + \beta = \rho_0$, resulting in the restricted estimates $\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}$. A general representation for h_t may be written as:

$$h_t = \frac{\omega}{1-\rho} + \alpha \bullet f(\rho, \vec{w}_t) \tag{4.9}$$

where $f(\rho, \vec{w}_t) = \sum_{i=1}^{\infty} \rho^{i-1} w_{t+1-i}$ and $\vec{w}_t = (w_t, w_{t-1}, \ldots)$. A linear approximation is taken on the variance term to get, after algebraic manipulations:

$$h_t = c + \alpha \bullet f(\tilde{\beta}, \vec{\eta}_t^2) + \lambda \bullet [f(\rho_0, \vec{\eta}_t^2) - f(\tilde{\beta}, \vec{\eta}_t^2)] \tag{4.10}$$

where $\lambda = \alpha \bullet (\rho - \rho_0)$. The valid test for the null hypothesis $H_0 : \rho = \rho_0$ is the test for $\lambda = 0$ in (4.10). Again, if the null is true, the third term in (4.10) cannot contribute to the variance significantly.

In MC experiment this valid test is computed and shown to cover the null 83.3% of the time with 95% confidence interval, not perfect but much better than the routine test which gives 49.6% coverage probability. Next I apply this test to residuals from consumption data after eliminating the "Working" effect in level. Figure 4.2 gives a 95% confidence interval by numerically inverting the test statistics, robust to a potential misspecification based on the Bollerslev-Wooldridge robust variance-covariance matrix (Bollerslev and Wooldridge (1992)). The resulting confidence interval for ρ is [0, 0.98], not only much wider than the interval [0.88, 1.04] from routine inference (again consistent with Dufour (1997)) but clearly includes 0 persistence.

Given corrected confidence intervals, it seems impossible to further pin down the value ranges for persistence measures ϕ and ρ beyond the valid but wide ones above. There does not seem to be direct empirical evidence in support of the existence of the so-called long-run risk which is supposedly able to increase the equity premium requested by economic agents. Instead, my finding that we cannot reject zero persistence in both level and volatility supports the random walk assumption for consumption, whose theoretical foundation traces at least back to Hall (1978) and a large body of literature dealing with asset pricing is built upon this assumption, see Abel (1990), Campbell and Cochrane (1999) etc.

To check the robustness, I did a study and report briefly my findings about annual and monthly consumption growth data. The source of data is BEA. The sample period for annual data is from 1929 to 2005 and for the monthly data is from 1959M01 to 2005M12. First, due to fewer sampling points for annual data, even the routine estimation gives a quite uninformative confidence interval for ϕ [0.11, 1.17] at the 95% significant level (but with a point estimate as large as 0.64) which, however, still disagrees with the confidence interval [-0.78, 1) given by the valid test. Regarding the volatility process, the routine test gives for ρ a much tighter confidence interval [0.85, 1.07], which differs much from [0.06, 1) the much wider one identified by the valid test. For monthly data, what is odd and significantly different from both the annual and quarterly data is that the first order auto-correlation turns out to be significantly negative. And even the routine estimation fails to find a high level of persistence for growth expectation with resulting point estimate for ϕ being -0.13, with a 95% confidence interval [-0.51, 0.25]. The valid test, however, gives three disjoint intervals (-1, -0.90]U[-0.29,0.06]U[0.90,1), in which only the second part agrees roughly with the routine one. For volatility persistence, routine estimation still gives an extremely tight 95% confidence interval [0.98, 0.99], which implies a very high level of persistence but the valid one [0.02, 0.97] reveals that there is no such evidence.

Despite much evidence provided here and elsewhere to support that consumption might simply a random walk as the classical Permanent Income Hypothesis implies, the consumption growth expectation might be integrated as Barsky and Delong (1993) suggest and interesting asset pricing implications can be derived from this. The valid test would not be appropriate if the underlying growth expectation is truly integrated ($\phi = 1$) due to Dickey-Fuller's (1979) work. To scrutinize this possibility I impose that $\phi = 1$ and resort to the Median Unbiased Estimator (MUE) of τ proposed by Stock and Watson (1998) to test the null $\tau = 0$. The MUE is needed since Shepard and Harvey (1990) fount that τ is biased downward and has a large mass at 0 if one relies on usual estimation method when the true ϕ is 1 but τ is small. To carry out Stock and Watson's (1998) procedure, we need to impose a technical restriction that the innovations to the expectation process and the cyclical shock are serially and mutually independent. So we may rewrite equation (4.3) as:

$$\begin{aligned} \tilde{g}_t &= x_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2) \\ x_t &= x_{t-1} + \tau \cdot \eta_t, \eta_t \sim N(0, \sigma^2) \end{aligned} \quad (4.11)$$

Table 4.3 summarizes the MUE of τ with its 95% confidence interval. These results unanimously fail to reject the null $H_0 : \tau = 0$. Finally, even if the argument that these tests may have low power against a very small τ leaves a possibility of the integrated expectation alive, this would, however, explode the equity premium (see Table 4.1) for any fixed level of preference parameters. From Table 4.2, it does seem that an integrated expectation ($\phi = 1$), in the limit, allows the risk aversion to approach 1. In this scenario, however, the volatility would not be priced ($\theta \to 0$). More importantly, the equity premium would be extremely sensitive to any small change of ψ, leading to an "on-the-edge" solution.

4.4. Conclusion

In this work I show that a recent resolution of the equity premium puzzle based on the long-run risk, a highly persistent component in consumption growth expectation and volatility, depends critically upon a correct statistical inference of the persistence level. The routine estimation method is not reliable since the employed model in this resolution is weakly identified and this may result in spurious evidence for a high level of persistence. A valid inference strategy is applied to this scenario and the valid confidence interval reveals very little empirical evidence in support of this resolution.

Table 4.1: Resulting Equity Premium (%) for Various Levels of Persistence
Preference parameters are fixed at: $\gamma = 10, \psi = 2, \delta = 0.966$

		$\rho = \alpha + \beta$				
		0	0.60	0.88	0.96	0.98
	0	1.72	1.73	1.79	2.47	6.81
	0.60	2.24	2.25	2.38	3.74	12.4
ϕ	0.90	5.43	5.52	6.65	18.8	96.6
	0.95	10.8	11.2	16.8	76.2	456.4
	0.99	58.5	74.7	279	2472	16490

Table 4.2: Required Values of γ to Match the Observed Equity Premium (about 6%) for Various Level of Persistence; other preference parameters: $\psi = 2, \delta = 0.966$.

		$\rho = \alpha + \beta$			
		0.88	0.92	0.96	0.98
	0.60	21	18	13	8
ϕ	0.80	15	13	10	6
	0.90	10	9	6.5	4.2
	0.99	1.8	1.7	1.6	1.4

Table 4.3: Testing the Integrated Expectation Based on Median Unbiased Estimator at the 95% Confidence Intervals: Real Per Capita Consumption Growth Data

	MUE of τ (p-value)	95% Confidence Interval
	Frequency: Annually	
QLR	0.059 (0.21)	[0, 0.328]
MW	0.000 (0.70)	[0, 0.163]
EW	0.013 (0.46)	[0, 0.239]
	Frequency: Quarterly	
QLR	0.010 (0.37)	[0, 0.086]
MW	0.000 (0.89)	[0, 0.029]
EW	0.000 (0.51)	[0, 0.071]
	Frequency: Monthly	
QLR	0.010 (0.14)	[0, 0.049]
MW	0.008 (0.21)	[0, 0.049]
EW	0.009 (0.16)	[0, 0.049]

Note: *QLR* represents the maximum F_T statistic, see e.g., Quandt (1960), Andrews (1993).
MW and *EW* denote, respectively, the mean and exponential Wald statistic in Andrews and Ploberger (1994).

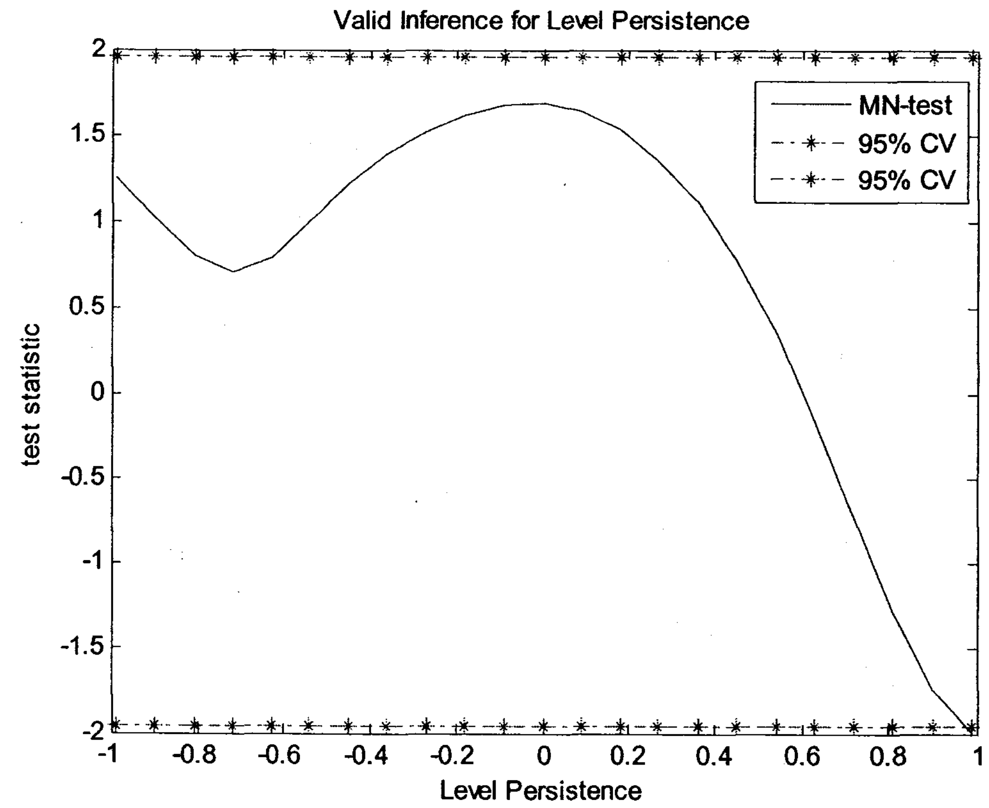


Figure 4.1: The 95% Confidence Interval for Persistence Measure of Level Based on the Valid Test

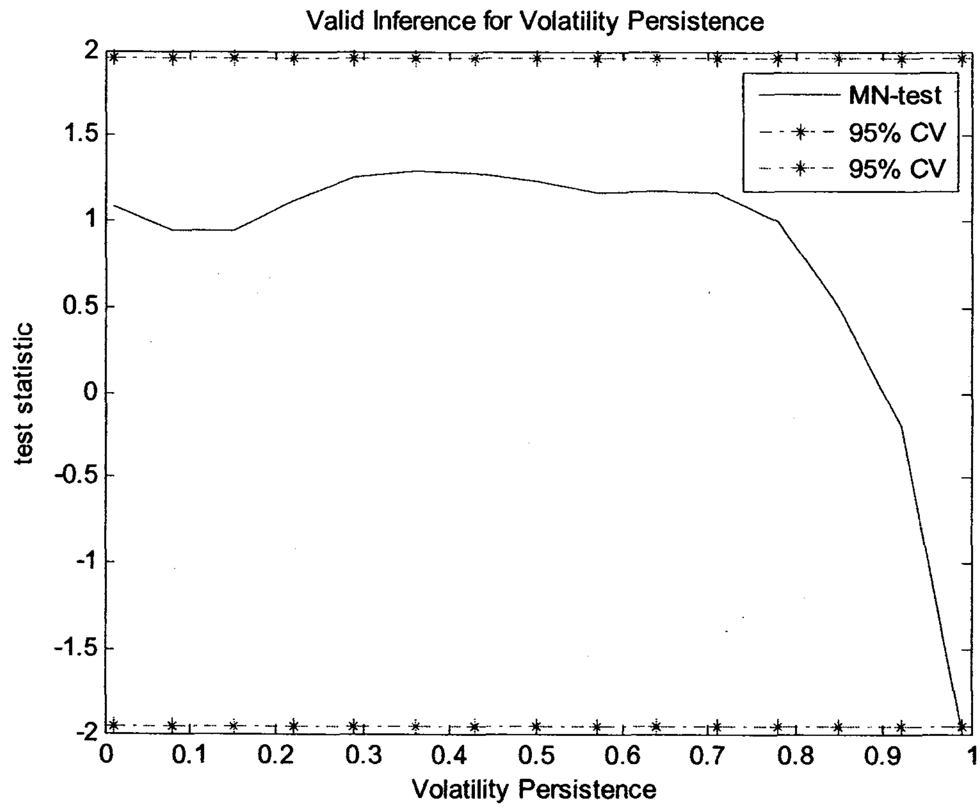


Figure 4.2: The 95% Confidence Interval for Persistence Measure of Volatility Based on the Valid Test

Bibliography

Abel, A., 1990, Asset Prices under Habit Formation and Catching up with the Joneses, *American Economic Review*, 80, 38-42.

Anderson, T.W., and H. Rubin, 1949, Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations, *The Annals of Mathematical Statistics*, 20, 46-63.

Andrews, D. W. K., 1993, Tests for Parameter Instability and Structural Change with Unknown Change Point, *Econometrica*, 61, 821-856.

Andrews, D. W. K., and W. Ploberger, 1994, Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative, *Econometrica*, 62, 1383-1414.

Andrews, D. W. K., and W. Ploberger, 1996, Testing for Serial Correlation against an ARMA(1,1) Process, *Journal of the American Statistical Association*, 91, 1331-1342.

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

Andrews, D., M. J. Moreira, and J. H. Stock, 2006, Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression, *Econometrica*, 74, 715-752.

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.

Bansal, R., and C. Lundblad, 2002, Market Efficiency, Asset Returns, and the Size of The Risk Premium in Global Equity Markets, *Journal of Econometrics*, 109, 195-237.

Bansal, R., and A. Yaron, 2004, Risks for the Long Run: a Potential Resolution of Asset Pricing Puzzles, *Journal of Finance*, LIX, 1481-1509.

Barsky, R., and B. J. DeLong, 1993, Why Does the Stock Market Fluctuate?, *Quarterly Journal of Economics*, 108, 291-312.

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.

Beveridge, S. and C. R. Nelson, 1981, A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle, *Journal of Monetary Economics*, 7, 151-174.

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.

Breusch, T. S., and A. R. Pagan, 1980, The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics, *The Review of Economic Studies*, 47, 239-53.

Brooks, C., S. Burke, and G. Persand, 2001, Benchmarks and the accuracy of GARCH model estimation, *International Journal of Forecasting*, 17, 45-56.

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

Campbell, J. Y. and R. Shiller, 1988, The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factor, *Review of Financial Studies*, 1, 195-227.

Campbell, J. Y., and J. Cochrane, 1999, By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior, *Journal of Political Economy*, 107, 205-251.

Campbell, J. Y., 2000, Asset Pricing at the Millennium, *The Journal of Finance*, Vol. LV, 1515-1567.

Cecchetti, S., P. S. Lam, and N. Mark, 1990, Mean Reversion in Equilibrium Asset Prices, *American Economic Review*, 80, 398-419.

Clark, P. K., 1987, The Cyclical Component of U.S. Economic Activity, *The Quarterly Journal of Economics*, 102, 797-814.

Cochrane, J. H., 2006, Financial Markets and the Real Economy, working paper, University of Chicago.

Constantinides, G. M., 1990, Habit Formation: A Resolution of the Equity Premium Puzzle, *Journal of Political Economy*, 98, 519-543.

Davidson, R., and J. G. MacKinnon (1993): *Estimation and Inference in Econometrics*, Oxford University Press.

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.

Dickey, D. A. and W. A. Fuller, 1979, Distributions of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, 74, 427-431.

Dufour, J. M., 1997, Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models, *Econometrica*, 65, 1365-88.

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., 1984, Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics, *Handbook of Econometrics*, Volume II, Edited by Z. Griliches and M. D. Intriligator.

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.

Epstein, L. G., and S. Zin, 1989, Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework, *Econometrica*, 57, 937-969.

Epstein, L. G., and S. Zin, 1991, Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns ii: An Empirical Analysis, *Journal of Political Economy* 99, 1263-1286.

Fieller, E. C., 1932, The Distribution of the Index in a Normal Bivariate Population, *Biometrika*, 24, 428-440.

Fieller, E. C., 1954, Some Problems in Interval Estimation, *Journal of the Royal Statistical Society. Series B (Methodological)*, 16, 175-85.

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.

Gavin, W.T., Keen, B.D., and Pakko, M.R., 2006, Inflation Risk and Optimal Monetary Policy, working paper 2006-035A, Federal Reserve Bank of St. Louis.

Hall, R.E., 1978, Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence, *Journal of Political Economy*, 86, 971-987.

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

Hamilton, J., 1996, A Standard Error for the Estimated State Vector of a State-Space Model, *Journal of Econometrics*, 33, 387-397.

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.

Hansen, L. P., and K. Singleton, 1984, Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models: Errata, *Econometrica*, 52, 267-268.

Hansen, L.P., J. C. Heaton and N. Li, 2005, Consumption Strikes Back?, Manuscript, University of Chicago.

Harvey, A. C., 1993, Time Series Models, 2nd Edition, London, Harvester Wheatsheaf.

Harvey, A. C., 1995, Trends and Cycles in Macroeconomic Time Series, *Journal of Business and Economics Statistics*, 3, 216-227.

He, C. and T. Terasvirta, 1999, Fourth moment structure of the GARCH(*p,q*) process, *Econometric Theory*, 15, 824-846.

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.

Kleibergen, F., 2002, Pivotal Statistics for Testing Structural Parameters in Instrumental Variable Regression, *Econometrica*, 70, 1781-1803.

Kocherlakota, N. R., 1996, The Equity Premium: It's Still a Puzzle, *Journal of Economic Literature*, 34, 42-71.

Kristensen, D. and O. Linton, 2006, A closed-form estimator for the GARCH(1,1)-model, *Econometric Theory*, 22, 323-337.

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

LeRoy, S. F., and R. D. Porter, 1981, The Present Value Relation: Tests Based on Implied Variance Bounds, *Econometrica*, 49, 555-574.

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., C. R. Nelson., and R. Startz, 2007, Spurious in the GARCH(1,1) Model When It Is Weakly Identified, *Studies in Nonlinear Dynamics & Econometrics*, Vol.11, No. 1, Article 1.

Ma, J., 2007, 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. and C. R. Nelson, 2007, Valid Inference under Weak Identification in Models Where the Zero-Information-Limit-Condition Holds, working paper, Department of Economics, University of Washington.

Mehra, R., and E. C. Prescott, 1985, The Equity Premium: A Puzzle, *Journal of Monetary Economics*, 15, 145-161.

Mehra, R., and E. C. Prescott, 2003, The Equity Premium in Retrospect, *Handbook of the Economics of Finance*, edited by G. M. Contantinides, M. Harris and R. Stulz, North Holland, Amsterdam.

Morley, J. C., C. R. Nelson, and E. Zivot, 2002, Why Are Beveridge-Nelson and Unobserved-Component Decompositions of GDP So Different?, *Review of Economics and Statistics*, 85, 235-243.

Nelson, C. R., 1988, Spurious Trend and Cycle in the State Space Decomposition of a Time Series with a Unit Root, *Journal of Economic Dynamics & Control*, 12, 475-488.

Nelson, C. R., and R. Startz, 1990a, The Distribution of the Instrumental Variables Estimator and Its *t*-Ratio When the Instrument Is a Poor One, *The Journal of Business*, 63, 125-40.

Nelson, C. R., and R. Startz, 1990b, Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator, *Econometrica*, 58, 967-76.

Nelson, C. R., and R. Startz, 2007, The Zero-Information-Limit Condition and Spurious Inference in Weakly Identified Models, *Journal of Econometrics*, 138, 47-62.

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

Quandt, R. E., 1960, Tests of the Hypothesis That a Linear Regression System obeys Two Separate Regimes, *Journal of American Statistical Association*, 55, 324-330.

Shephard, N. G, and A. C. Harvey, 1990, On the Probability of Estimating a Deterministic Component in the Local Level Model, *Journal of Time Series Analysis*, 11, 339-347.

Shiller, R., 1981, Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends? *American Economic Review*, 71, 421-436.

Stock, J. H., and M. W. Watson, 1998, Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model, *Journal of the American Statistical Association*, 93, 349-358.

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.

Staiger, D., and J. Stock, 1997, Instrumental Variables Regression with Weak Instruments, *Econometrica*, 65, 1997, 557-586.

Staiger, D., J. H. Stock., and M. W. Watson, 1997, The NAIRU, Unemployment and Monetary Policy, *The Journal of Economic Perspectives*, 11, 33-49.

Startz, R., E. Zivot., and C. R. Nelson, 2006, Improved Inference in Weakly Identified Instrumental Variables Regression, *Frontiers in Analysis and Applied Research: Essays in Honor of Peter C.B. Phillips*, Cambridge University Press.

Stock, J., J. H. Wright., and M. Yogo, 2002, A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business & Economic Statistics*, 20, 518-29.

Wang, J.H., and E. Zivot, 1998, Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments, *Econometrica*, 66, 1389-1404.

Weil, P., 1989, The Equity Premium Puzzle and the Risk Free Rate Puzzle, *Journal of Monetary Economics*, 24, 401-421.

Weiss, A., 1986, Asymptotic theory for ARCH models: Estimation and Testing, *Econometric Theory*, 2, 107-131.

White, H., 1980, A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, 817-838.

Wilcox, D. W., 1992, The Construction of U.S. Consumption Data: Some Facts and Their Implications for Empirical Work, *The American Economic Review* 82, 922-41.

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

Yogo, M, 2004, Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak, *The Review of Economics and Statistics*, 86, 797-810.

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

A.1 A General Proof of ZILC

Ma (2007) (or see the second chapter), gives an analytical information matrix for the GARCH(1,1) estimator $(\hat{\omega},\hat{\alpha},\hat{\beta})$:

$$I=\frac{(1-\alpha-\beta)^2}{2\omega^2}\begin{pmatrix} A & B & B \\ B & C & D \\ B & D & E \end{pmatrix} \tag{A.1.1}$$

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

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

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

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

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

The 'information' measure for $\hat{\beta}$, defined to be the inverse of its variance by Nelson and Startz (2007), 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

$$T \cdot \frac{(1-\alpha-\beta)^2}{2\omega^2} \to T \cdot \frac{(1-\beta)^2}{2\omega^2} \neq 0 \text{ as } \alpha \to 0 \qquad \text{(A.1.2)}$$

$$AC - B^2 \to \frac{2\omega^2}{(1-\beta)^6(1+\beta)} \neq 0 \text{ as } \alpha \to 0 \qquad \text{(A.1.3)}$$

However,

$$B^2(2D-C-E)+A(CE-D^2) \to 0 \text{ as } \alpha \to 0 \qquad \text{(A.1.4)}$$

This completes the proof of (1.4).

A.2 A Few Implications When There Is No GARCH Effect

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

$$Asy.Var\begin{pmatrix}\hat{\omega}\\ \hat{\alpha}\\ \hat{\beta}\end{pmatrix} = \begin{bmatrix} \frac{\omega^2(1+\alpha)}{\alpha^2(1-\alpha)} & 0 & -\frac{\omega(1-3\alpha^2)}{\alpha^2(1-\alpha)} \\ -- & \frac{(1-3\alpha^2)}{(1+\alpha)(1-\alpha)} & -\frac{(1-3\alpha^2)}{(1+\alpha)(1-\alpha)} \\ -- & -- & \frac{(1-3\alpha^2)}{\alpha^2(1+\alpha)(1-\alpha)} \end{bmatrix} \quad \text{(A.2.1)}$$

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

$$I_{\hat{\beta}}(\omega,\alpha,\beta) = \frac{\alpha^2(1+\alpha)(1-\alpha)T}{(1-3\alpha^2)} \to 0, \text{as } \alpha \to 0 \quad \text{(A.2.2)}$$

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

$$I_{\hat{\omega}}(\omega,\alpha,\beta) = \frac{\alpha^2(1-\alpha)T}{\omega^2(1+\alpha)} \to 0, \text{as } \alpha \to 0 \quad \text{(A.2.3)}$$

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

$$Asy.Corr(\hat{\omega},\hat{\beta}) = -\sqrt{1-3\alpha^2} \to -1, \text{as } \alpha \to 0 \quad \text{(A.2.4)}$$

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

$$I_{\hat{\alpha}}(\omega,\alpha,\beta) = \frac{(1+\alpha)(1-\alpha)T}{(1-3\alpha^2)} \to T \neq 0, \text{as } \alpha \to 0 \quad \text{(A.2.5)}$$

Appendix B Derivations of Covariances

B.1 The Derivation of (2.27) and (2.28)

Using the standard formulas for expectation, covariance and variance, we have:

$$\begin{aligned} E[\varepsilon_t^2 \varepsilon_{t-i}^2] &= Cov(\varepsilon_t^2, \varepsilon_{t-i}^2) + E[\varepsilon_t^2]E[\varepsilon_{t-i}^2] \\ &= \rho_i Var(\varepsilon_t^2) + (E[\varepsilon_t^2])^2 \\ &= \rho_i (E[\varepsilon_t^4] - (E[\varepsilon_t^2])^2) + (E[\varepsilon_t^2])^2 \end{aligned} \tag{B.1}$$

Plug (B.1) into (2.16) and Sum up the infinite geometric series to get:

$$\begin{aligned} E\left(\sum_{i=1}^{\infty} \beta^{i-1} \varepsilon_{t-i}^2\right)^2 &= \frac{E[\varepsilon_t^4]}{1-\beta^2} + \frac{2\beta}{1-\beta^2}(E[\varepsilon_t^2\varepsilon_{t-1}^2] + \beta E[\varepsilon_t^2 \varepsilon_{t-2}^2] + \cdots) \\ &= \frac{E[\varepsilon_t^4]}{1-\beta^2} + \frac{2\beta}{1-\beta^2}\sum_{i=1}^{\infty}\beta^{i-1}E[\varepsilon_t^2\varepsilon_{t-i}^2] \\ &= \frac{E[\varepsilon_t^4]}{1-\beta^2} + \frac{2\beta}{1-\beta^2}\{\sum_{i=1}^{\infty}\beta^{i-1}((\alpha+\beta)^{i-1}\rho_1^{\varepsilon 2}(E[\varepsilon_t^4] - (E[\varepsilon_t^2])^2) + (E[\varepsilon_t^2])^2)\} \\ &= \frac{E[\varepsilon_t^4]}{1-\beta^2} + \frac{2\beta}{1-\beta^2}\{E[\varepsilon_t^4]\frac{\rho_1^{\varepsilon 2}}{1-(\alpha+\beta)\beta} - (E[\varepsilon_t^2])^2\frac{\rho_1^{\varepsilon 2}}{1-(\alpha+\beta)\beta} + (E[\varepsilon_t^2])^2\frac{1}{1-\beta}\} \end{aligned} \tag{B.2}$$

Plug in the expressions for $E[\varepsilon_t^4]$, $\rho_1^{\varepsilon 2}$, and $E[\varepsilon_t^2]$, one can end up with the result (2.27). Likewise, one can derive the result (2.28).

B.2 The Derivation of Cross-Covariance

By the MDS property of the innovation w_t, we have:

$$E[w_{t-1} \cdot h_{t-1}] = 0 \Rightarrow E[(\varepsilon_{t-1}^2 - h_{t-1})h_{t-1}] = 0 \Rightarrow E[\varepsilon_{t-1}^2 h_{t-1}] = E[h_{t-1}^2] \tag{B.2.1}$$

Here, one needs to notice that $h_{t-1} = \omega + \alpha \cdot \varepsilon_{t-2}^2 + \beta \cdot h_{t-2} = f(I_{t-2})$, thus is adapted to the information set at time *t-2*. Since $\{I_t\}$ is an filtration associated with w_t, we have: $\cdots \supset I_t \supset I_{t-1} \supset \cdots$. Apply the law of iterative expectation to get: $E[w_{t-1} \mid I_{t-i}] = E[[w_{t-1} \mid I_{t-2}] \mid I_{t-i}] = 0, i = 3,4,\cdots$.

Therefore, we have result:

$$E[w_{t-1} \cdot h_{t-i}] = 0 \Rightarrow E[(\varepsilon_{t-1}^2 - h_{t-1})h_{t-i}] = 0 \Rightarrow E[\varepsilon_{t-1}^2 h_{t-i}] = E[h_{t-1}h_{t-i}], i = 2,3\cdots$$

In the same way, one can derive the result:

$$E[w_{t-1}\varepsilon_{t-i}^2] = 0 \Rightarrow E[(\varepsilon_{t-1}^2 - h_{t-1})\varepsilon_{t-i}^2] = 0 \Rightarrow E[h_{t-1}\varepsilon_{t-i}^2] = E[\varepsilon_{t-1}^2 \varepsilon_{t-i}^2], i = 2,3,\cdots$$

Appendix C Transformation in the ARMA and GARCH Models

Appendix C.1 The Derivation of (3.14)

(3.13) is simply, in terms of lag operators:

$$y_t = \gamma \bullet (1-\phi_0 L)^{-1} L\tilde{\varepsilon}_t + \lambda \bullet (1-\phi_0 L)^{-2} L^2 \tilde{\varepsilon}_t + e_t \qquad \text{(C.1.1)}$$

Substituting $\tilde{\varepsilon}_t = (1-\phi_0 L) \bullet (1-\tilde{\theta}L)^{-1} y_t$ in to get:

$$y_t = \gamma \bullet (1-\tilde{\theta}L)^{-1} L y_t + \lambda \bullet (1-\phi_0 L)^{-1}(1-\tilde{\theta}L)^{-1} L^2 y_t + e_t \qquad \text{(C.1.2)}$$

By using the identity:

$$(1-\phi_0 L)^{-1}(1-\tilde{\theta}L)^{-1} = [(1-\phi_0 L)^{-1} - (1-\tilde{\theta}L)^{-1}] \bullet (\phi_0 - \tilde{\theta})^{-1} L^{-1} \qquad \text{(C.1.3)}$$

We have:

$$y_t = \gamma \bullet (1-\tilde{\theta}L)^{-1} L y_t + \lambda \bullet [\tilde{\gamma}^{-1}(1-\phi_0 L)^{-1} L y_t - \tilde{\gamma}^{-1}(1-\tilde{\theta}L)^{-1} L y_t] + e_t \qquad \text{(C.1.4)}$$

Appendix C.2 The Derivation of (3.33)

(3.32) may be equivalently written in terms of lag operators:

$$h_t = c + \alpha \bullet (1-\rho_0 L)^{-1} L\widetilde{w}_t + \lambda^* (1-\rho_0 L)^{-2} L^2 \widetilde{w}_t + \text{remainder} \qquad \text{(C.2.1)}$$

Where, $\widetilde{w}_t = (1-\rho_0 L)(1-\widetilde{\beta} L)^{-1} \bullet \varepsilon_t^2 - \omega(1-\widetilde{\beta})^{-1}$ and plug in to get:

$$h_t = c^* + \alpha \bullet (1-\widetilde{\beta} L)^{-1} L\varepsilon_t^2 + \lambda^* \bullet (1-\rho_0 L)^{-1}(1-\widetilde{\beta} L)^{-1} L^2 \varepsilon_t^2 + \text{remainder} \quad \text{(C.2.2)}$$

Using the identity similar to (C.1.3) to get:

$$h_t = c^* + \alpha \bullet (1-\widetilde{\beta} L)^{-1} L\varepsilon_t^2 + \lambda^* \bullet [\widetilde{\alpha}^{-1}(1-\rho_0 L)^{-1} L\varepsilon_t^2 - \widetilde{\alpha}^{-1}(1-\widetilde{\beta} L)^{-1} L\varepsilon_t^2] + \text{remainder} \qquad \text{(C.2.3)}$$

Appendix D Derivation of Equity Premium

The derivation of (4.5) starts with the Euler equation:

$$E_t[\exp\{\theta \ln \delta - \frac{\theta}{\psi} g_{t+1} + \theta r_{m,t+1}\}] = 1 \tag{D.1}$$

The marginal rate of substitution (MRS) is given by:

$$m_{t+1} = \theta \ln \delta - \frac{\theta}{\psi} g_{t+1} - (1-\theta) r_{m,t+1} \tag{D.2}$$

Adopting Campbell and Shiller's (1988) approximation:

$$r_{m,t+1} = \kappa_0 + \kappa_1 z_{t+1} - z_t + g_{t+1} \tag{D.3}$$

Where z_t is the price-dividend (consumption) ratio; κ_0, κ_1 are approximation constants. The market portfolio return $r_{m,t+1}$ is solved with the assumption of log-normality. Equation (4.5) is then derived by following the classical formula in Campbell (2000) and taking unconditional expectations:

$$E_t(r_{m,t+1} - r_{f,t+1}) = -\mathrm{cov}_t(r_{m,t+1}, m_{t+1}) - 0.5 \cdot \mathrm{var}_t(r_{m,t+1}) \tag{D.4}$$

The risk free rate is given by:

$$E(r_f) = -\ln \delta + {E(g)}/{\psi} + 0.5 \cdot [(\theta - 1)P^2 - {\theta}/{\psi^2}] \cdot \sigma_\eta^2 + 0.5 \cdot (\theta - 1)\kappa_1^2 Q^2 \alpha^2 \sigma_w^2 \tag{D.5}$$

The multiplier term before σ_η^2 in (4.5) is a parabola in *P*:

$$(1-\theta)[(P+\frac{\theta}{2(1-\theta)\psi})^2-\frac{\theta^2}{4(1-\theta)^2\psi^2}] \qquad \text{(D.6)}$$

As in Bansal and Yaron (2000, 2004), I consider the following region of preference parameters: $1<\gamma<10$, $\psi>1$, and $\gamma\psi>1$ implying $\theta<0$. For this range of preferences, $P\geq 1$ with $\tau\geq 0$, and (D.6) is monotonically increasing in *P*.

For standard CRRA case, when $\theta=1$, (4.5) becomes:

$$E(r_m-r_f)=-0.5\cdot[S^2-2\gamma S]\sigma_\eta^2 \qquad \text{(D.7)}$$

where $S=1+a(1-\gamma)$ with $a=\frac{\tau}{1/\kappa_1-\phi}$.

The multiplier term for σ_η^2 becomes $-0.5\cdot(S-\gamma)^2+0.5\cdot\gamma^2$, a parabola in *S*. Since for $\gamma>1$, the case I consider here, $S\leq 1$. For any given level of $\gamma>1$, the maximum equity premium is achieved when $S=1$, i.e. when $\tau=0$, the case of an *i.i.d* consumption process. As a result, GEU specification with $\theta<0$ is important to have persistence play a significant role in generating the sizable equity premium.

VITA

Jun Ma was born in Jiangsu Province, People's Republic of China. He received a Bachelor of Arts in Economics from the Nanjing University in June 2001 and a Master of Arts in Economics from the University of Washington in March 2006. He has published on the Berkeley Electronic Journal *Studies in Nonlinear Dynamics & Econometrics*.