A Long Memory Model with Mixed Normal GARCH for US Inflation Data

Yin-Wong Cheung and Sang-Kuck Chung *

We introduce a time series model that captures both long memory and conditional heteroskedasticity and assess its ability to describe the US inflation data. Specifically, the model allows for long memory in the conditional mean formulation and uses a normal mixture GARCH process to characterize conditional heteroskedasticity. We find that the proposed model yields a good description of the salient features, including skewness and heteroskedasticity, of the US inflation data. Further, the performance of the proposed model compares quite favorably with, for example, ARMA and ARFIMA models with GARCH errors characterized by normal, symmetric and skewed Student-*t* distributions.

Keywords: Conditional heteroskedasticity, Skewness, Inflation, Long memory, Normal mixture

JEL Classification: C22, C51, C52, E31

I. Introduction

In this study, we consider a time series model that features both long memory and conditional heteroskedasticity and assess its ability to describe the U.S. inflation data. In the empirical literature, long

* Professor, Department of Economics, E2, University of California, 1156 High Street, Santa Cruz, CA 95064, USA, (Tel) +831-459-4247, (Fax) +831-459-5077, (E-mail) cheung@ucsc.edu; Corresponding Author, Associate Professor, Department of Economics, Center for Research on Northeast Asian Economy, Inje University, Obang-dong 607, Kimhae, Kyungnam 621-749, Korea, (Tel) +82-55-320-3124, (Fax) +82-55-337-2902, (E-mail) tradcsk@inje.ac.kr, respectively. The authors would like to thank Haas Markus and Juri Marcucci for their codes that were incorporated into the computer programs for our exercise, and the referees and Bong-Han Kim for their constructive comments and suggestions. Chung acknowledges the financial support from the Inje Research Grant in 2008.
[**Seoul Journal of Economics** 2009, Vol. 22, No. 3]

memory in time series data is quite commonly modeled using the autoregressive fractionally integrated moving average (ARFIMA) specification (Granger 1980; Granger and Joyeux 1980; Hosking 1981). Under the usual ARMA regime, data are classified as either, say, integrated of order 0, I(0) or integrated of 1, I(1). ARFIMA models avoid the knife-edge choice between I(0) stationarity and I(1) unit-root persistence by allowing the order of integration to assume a real value. Arguably, since its introduction in the 1980s, the ARFIMA model offers the most popular framework for characterizing long-memory data persistence.[1]

Conditional heteroskedasticity is an important attribute of economic data. The basic autoregressive conditional heteroskedasticity (ARCH) model was introduced in the seminal work of Engle (1982). Bollerslev (1986) generalizes the model to the generalized ARCH (GARCH) specification, which is the workhorse of analyzing time-varying (conditional) volatility. Various modifications of the basic GARCH model have been proposed.[2]

Recently, Alexander and Lazar (2004, 2006) and Haas *et al.* (2002, 2004a, 2004b) advance a model with a normal mixture of GARCH (NM-GARCH) processes. Essentially, the mixture model is designed to describe a conditional volatility process that is driven by a linear combination of GARCH processes. In addition to its flexibility in analyzing volatility, the use of the normal mixture specification allows the NM-GARCH model to describe skewness in both conditional and unconditional distributions.

The model considered in this study augments a long memory model with the recently proposed NM-GARCH specification. We call the augmented model an ARFIMA-NM-GARCH model. By design, the proposed model is apt for modeling data that display both long memory and GARCH behavior and with the normal mixture feature, it could capture both conditional skewness and conditional heteroskedasticity in data.

To illustrate its empirical relevance, we apply the ARFIMA-NM-GARCH model to the U.S. inflation data.[3] We should point out that both ARFIMA and GARCH models have been used to describe the U.S. inflation data. Thus, the ARFIMA-NM-GARCH specification is a natural

[1] Some early applications include Cheung (1993), Cheung and Lai (1993, 2001), and Diebold and Rudebusch (1991).

[2] Interested readers are referred to a recent survey Bauwens *et al.* (2006).

[3] Arguably, inflation is an important macroeconomic variable in designing policy rules; see, for example, Seo and Kim (2007).

extension for modeling the U.S. inflation data.

The rest of the paper is structured as follows. Section II describes the ARFIMA-NM-GARCH model and the maximum likelihood estimation procedure. Section III presents the results of modeling the U.S. inflation data. Section IV contains the summary.

II. The ARFIMA-NM-GARCH Model

A. The Model

Baillie *et al.* (1996), for example, consider models incorporating both ARFIMA and GARCH effects. The model we proposed here, thus, can be viewed as a follow-up of this line of research. Specifically, an ARFIMA(p, q)-NM(k)-GARCH(r, s) is given by

$$(1-\phi(B))(1-B)^d\ (y_t-\mu)=(1+\theta(B))\,\varepsilon_t, \tag{1}$$

$$\varepsilon_t\,|\,\Omega_{t-1} \sim MN(p_1, \ldots, p_k;\ \lambda_1, \ldots, \lambda_k;\ \sigma^2_{1t}, \ldots, \sigma^2_{kt}), \text{ and} \tag{2}$$

$$\sigma^2_t=\omega+\textstyle\sum^s_{i=1} \alpha_i\, \varepsilon^2_{t-i}+\sum^r_{i=1} \beta_i\, \sigma^2_{t-i}. \tag{3}$$

The long memory property is characterized by the fractional differencing operation $(1-B)^d$ in Equation (1). The $\phi(B)$ and $\theta(B)$ are the standard *p-th* order autoregressive and *q-th* order moving-average polynomials. The process is said to display long memory when $0<d<0.5$.[4] If the two lag polynomials $\phi(B)$ and $\theta(B)$ have roots outside the unit circle, the process is stationary and invertable for $-0.5<d<0.5$.

The innovation term ε_t is assumed to follow a mixture of k normal distributions, conditional on an information set Ω_{t-1}, with the mixing parameters $p_i \in (0, 1)$, $i=1, \ldots, k$ and $\sum^k_{i=1} p_i=1$. The means and variances of these normal distributions are denoted by λ_i and σ^2_{it}; $i=1, \ldots, k$. Following the practice in literature, we set $\lambda_i=-\sum^{i-1}_{j=1} (p_j/p_i)\,\lambda_i$. Equation (3) gives a general representation of the normal mixture GARCH process for which individual conditional variances evolve according to a GARCH process that depends on the lags of squared residuals and conditional variances. Specifically, $\sigma^2_t=[\sigma^2_{1t}, \sigma^2_{2t}, \ldots, \sigma^2_{kt}]^T$,

[4] Note that $(1-B)^d=\sum^\infty_{k=0} \Gamma(k-d)B^k/[\Gamma(-d)\Gamma(k+1)]$, where $\Gamma(.)$ is the gamma function. The basic properties of fractionally differenced series are discussed in, for example, Granger and Joyeux (1980) and Hosking (1981).

$\omega=[\omega_1, \ldots, \omega_k]^T$, $\alpha_i=[\alpha_{i1}, \ldots, \alpha_{ik}]^T$; $i=1, \cdots, s$; and β_i is a $k\times k$ coefficient matrix $[\beta_{i,mn}]_{m,n=1,\ldots,k}$, $i=1, \ldots, r$.

By construction, a NM-GARCH process is a linear combination of individual GARCH processes. It inherits all the salient features of a GARCH process; including the ability to model volatility clustering and volatility persistence. In addition, the NM-GARCH process has the flexibility to accommodate the possibility that the data heteroskedasticity generating process is driven by more than one GARCH factors. The flexibility, in turn, allows the NM-GARCH process to capture time-varying skewness, in addition to kurtosis. It is noted that a generic GARCH process is symmetric and could not be used to model skewness.

For a NM-GARCH process, its degree of skewness is zero when the means of all the component normal processes are zero; that is, $\lambda_1=\lambda_2=\ldots=\lambda_k=0$; see, for example, Alexander and Lazar (2006) and Haas *et al.* (2004a). Thus, the process can be symmetric or asymmetric depending on parameter configuration. In the following, we label the model with the restriction $\lambda_1=\lambda_2=\ldots=\lambda_k=0$ a symmetric model and the one without the restriction an asymmetric model.

B. Estimation and Statistical Inference

In this subsection, we outline the maximum likelihood (ML) estimation procedure and briefly discuss its performance. Suppose $\mu=0$ and let $\gamma=(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, d)^T$, $\delta=(p_1, \ldots, p_{k-1}; \lambda_1, \ldots, \lambda_{k-1}; \kappa_1, \ldots, \kappa_k)^T$, where $\kappa_i=(\omega_i, \alpha_i, \beta_i)^T$ for $i=1, 2, \ldots, k$, and $\psi=(\gamma^T, \delta^T)^T$. Note that p_k and λ_k can be derived from p_j and λ_j; $j<k$. For simplicity, we consider the GARCH(1, 1) case while recognizing the possibility of generalizing it to a GARCH(r, s) process. We assume the true parameter vector $\psi_0=(\gamma_0{}^T, \delta^T)^T$ is in the interior of a compact set ψ. The ML estimator $\hat{\psi}_n$ maximizes the conditional log-likelihood $L(\psi)=\Sigma^n_{t=1} l_t(\psi)+(n/2)\ln(2\pi)$, where the point log likelihood is given by

$$l_t(\psi)=\ln\{\Sigma^k_{j=1} p_j, f_{jt}\} \tag{4}$$

and f_{jt} is the density of the *j-th* component of the normal mixture, *i.e.*,

$$f_{jt}=(\sqrt{2\pi}\sigma_{j,t})^{-1}\exp\left\{-\frac{1}{2}\left(\frac{\varepsilon_t-\lambda_j}{\sigma_{j,t}}\right)^2\right\} \quad \text{for} \quad j=1, \ldots, k.$$

Under the following conditions: a) $|d|<1/2$ and that all roots of $\phi(B)$ and $\theta(B)$ are outside the unit circle, b) $p_i\in(0, 1)$, $i=1$, ..., k and $\sum_{i=1}^{k} p_i=1$, c) $\alpha_i\geq 0$ and $1>\beta_i\geq 0$; $i=1$, ..., k, d) $\omega_i+\alpha_i m/w>0$; $i=1$, ..., k, $m=\sum_{i=1}^{k} p_i\lambda_i^2+\sum_{i=1}^{k} p_i\omega_i/(1-\beta_i)>0$, and $w=\sum_{i=1}^{k} p_i(1-\alpha_i-\beta_i)/(1-\beta_i)>0$, and e) $E(\varepsilon_t^4)<\infty$, it can be shown that

$$-\frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix}\dfrac{\partial^2 l_t}{\partial\gamma\partial\gamma^T} & \dfrac{\partial^2 l_t}{\partial\gamma\partial\delta^T}\\ \dfrac{\partial^2 l_t}{\partial\delta\partial\gamma^T} & \dfrac{\partial^2 l_t}{\partial\delta\partial\delta^T}\end{bmatrix}\xrightarrow{a.s.}\begin{pmatrix}\Sigma_\gamma & 0\\ 0 & \Sigma_\delta\end{pmatrix};n\to\infty, \tag{5}$$

where "$\xrightarrow{a.s.}$" denotes almost sure convergence, and Σ_γ and Σ_δ are positive matrices.[5] Under (5), there exists a MLE $\hat{\psi}_n$ such that it satisfies $\partial l(\hat{\psi}_n)/\partial\psi=0$, $\hat{\psi}_n \xrightarrow{probability} \psi_0$ as $n\to\infty$, and $\sqrt{n}\,(\hat{\psi}_n-\psi_0) \xrightarrow{Distribution} N\,(0,\ \Sigma_0^{-1})$ as $n\to\infty$, where $\Sigma_0=diag\,(\Sigma_{\gamma 0},\ \Sigma_{\delta 0})$, and $\Sigma_{\gamma 0}$ and $\Sigma_{\delta 0}$ are values of Σ_γ and Σ_δ at $\psi=\psi_0$. Note that Σ_0 is block diagonal and, thus, the parameter vectors γ and δ can be estimated separately without loss in asymptotic efficiency.

Cheung and Chung (2007) assess the performance of the ML estimator using the Monte Carlo approach. An ARFIMA-NM-GARCH model with two GARCH component processes in the normal mixture formulation was considered. These authors documented some encouraging evidence on estimating an ARFIMA-NM-GARCH model. For instance, while the sampling uncertainty is inversely related to the sample size, the biases of parameter estimates are not statistically significantly even for a sample size of 100.

Their simulation exercise found that the biases of the estimated ARCH and GARCH effects are affected by the mixing parameter p_1. Specifically, the absolute values of these biases are in general inversely related to their associated mixing parameter values. For instance, the absolute values of the β_i-biases are negatively related to the values of p_1. Similarly, the intercept in the conditional variance equation also displays a bias that is in general inversely related to the values of the corresponding mixing parameter.

The simulated variations of the conditional variance parameter

[5] See, for example, Ling and Li (1997) and Alexander and Lazar (2004, 2006) for the use of these conditions and the related results to derive a formal proof of the convergence results.

estimates are, in general, larger than those of other parameter estimates. On the other hand, the fractional parameter, compared with other model parameters, could be quite precisely estimated. Even for the sample size of 100, the estimated bias is only in the order of -0.03 and it declines to 0.001 when the sample size is 1,000.

III. The US Inflation Dynamics

Both ARFIMA and GARCH models have been used to describe the U.S. inflation dynamics.[6] Thus, in addition to results pertaining to the proposed ARFIMA-NM-GARCH models, we also present estimates from ARMA-GARCH and ARFIMA-GARCH models for comparison purposes. For these two GARCH-type models, we consider three innovation distributions; namely normal, Student-*t* and skewed Student-*t*.[7] For completeness, we present the Student-*t* and skewed Student-*t* distributions in the Appendix A. Note that, similar to the normal distribution, the Student-*t* distribution is symmetric. On the other hand, the skewed Student-*t* distribution, as its name implies, is asymmetric.

A. Preliminary Analyses

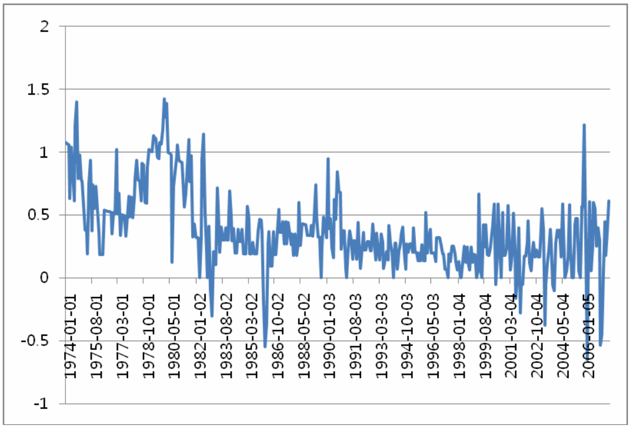
The U.S. inflation data were retrieved from the IMF database. The sample contains 399 monthly observations measured as 100 times the first differences of the logarithms of the CPI index from January 1974 to March 2007. Figure 1 presents the data and their autocorrelation coefficient estimates. The data plot displays a considerable degree of volatility clustering and is also suggestive of skewness and kurtosis. The persistence in inflation data is quite well illustrated by the slowly decaying correlogram pattern revealed in Figure 1B. It is noted that, at least for the first 70 autocorrelation coefficients, the estimates are statistically significant and outside the usual two-standard-error band (Bartlett 1946). The slowly decaying autocorrelation pattern is typical of data experiencing fractional integration; a property that we will investigate in the next sub-section.

Table 1A presents some descriptive statistics. These sample statistics

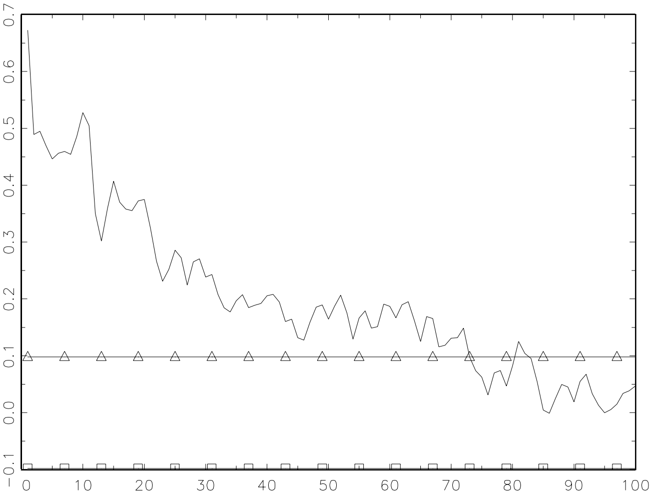
[6] For more detail see Baillie *et al*. (1992, 1996), Hassler and Wolters (1995), Doornik and Ooms (2004) among others.

[7] On the use of Student-*t* errors, see, for example, Baillie and Bollerslev (1989), Bollerslev (1987), and Palm and Vlaar (1997). On the use of skewed Student-*t* innovations, see Lambert and Laurent (2001a, 2001b).

A. Inflation: January 1974 to March 2007



B. Autocorrelation coefficients for the lag of 100



Note: The two-standard errors band is (-0.1, 0.1).

FIGURE 1
MONTHLY INFLATION AND AUTOCORRELATION FUNCTIONS

affirm that inflation data are skewed and leptokurtic. The statistics also suggest that the inflation data do not have a normal distribution and display serial correlation in both their levels and squares.

As part of the preliminary data analysis, the augmented Dickey-

TABLE 1
PRELIMINARY DATA ANALYSES

		Inflation
A. Summary Statistics		
Mean		0.3713
Skewness		0.6304
Kurtosis		4.0871
$Q(10)$		1011.1768*
$Q^2(10)$		1307.9220*
JB		45.9593*
B. Persistence Tests		
ADF	CNT	-2.7034* (11)
	CT	-3.1345* (11)
KPSS	CNT	0.4909* (14)
	CT	0.4626* (17)
GPH (υ=0.55)	d	0.6709**

Note: The $Q(10)$ and $Q^2(10)$ give the Ljung-Box statistics that include serial correlation in the first ten lags of residuals and their squares, respectively. JB gives the Jarque-Bera statistics. ADF gives the augmented Dickey-Fuller test statistics for models with a) a constant but not a trend (CNT), and b) with both a constant and a trend (CT). KPSS gives the Kwiatkowski-Phillips-Schmidt-Shin statistics for testing the null hypothesis of stationarity. GPH gives the Geweke and Porter-Hudak statistics for fractional integration with υ=0.55 and the number of periodograms used to general these statistics is given by T^{υ}. "*" indicates significance at a level of 10% or lower. "**" indicates the result is in favor of the alternatives of $0<d<1$, where d is the order of integration.

Fuller test, the Kwiatkowski, Phillips, Schmidt, and Shin (1992) test, and the Geweke and Porter-Hudak (1983) test are used to assess the integration property of the inflation data. The test results are presented in Table 1B. While the Dickey-Fuller test rejects the I(1) null hypothesis, the Kwiatkowski-Phillips-Schmidt-Shin test rejects the stationary I(0) null.[8] That is, the two tests do not agree on whether the inflation data follow an I(1) or an I(0) process. The Geweke-Porter-Hudak test, on the other hand, suggests that the data are fractionally integrated. Specifically, the test results indicate the differencing parameter is between zero and one. The presence of fractional

[8] The lag parameters are chosen to eliminate serial correlation in estimated residuals. Indeed, varying the lag parameter from 1 to 20 yields test results that are qualitatively the same as those reported.

differencing is consistent with the inclusive evidence obtained from the augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin tests that are designed to discriminate an I(0) process from an I(1) process. It is also in accordance with the slowly decaying autocorrelation pattern depicted in Figure 1B.

B. Estimation Results

Table 2 presents the results of fitting both symmetric and asymmetric ARFIMA-NM-GARCH models to the U.S. inflation data. For comparison purposes, the results of fitting ARMA-GARCH and ARFIMA-GARCH models to the data are also reported in the Table.[9,10] The respective model specifications are determined based on information criteria.

Note that the ARFIMA-NM-GARCH models we fitted to the inflation data have two components in the normal mixture GARCH formulation; that is, $k=2$. Further, the off-diagonal elements of β_i s are set to zeros. Alexander and Lazar (2006) and others find that substantial estimation biases appear when $k>2$ and allowing for non-zero off-diagonal elements in β_i s does not improve the empirical performance of NM-GARCH models they considered. Thus, for brevity, we consider the simplified version in the empirical part of our exercise. For the GARCH component, we focus on the GARCH(1,1) specification because it is known that the specification offers a very reasonable description to economic data in general. That is, (3) is simplified to $\sigma^2_{it}=\omega_i+\alpha_i\varepsilon^2_{t-1}+\beta_i\sigma^2_{it-1}$; $i=1, 2$.

The parameter estimates of the selected ARMA-GARCH models show that the inflation data are quite persistent and display strong GARCH effects. The result echoes those reported in previous studies that use inflation data to illustrate GARCH effects. The estimates of the degree of freedom parameter (υ) suggest that the innovation process is not likely to be normal. Indeed, the degree of freedom estimates are quite small and are about 5 — a value that makes the underlying Student-*t* distribution quite far away for the normal one. There is also evidence that the innovation is skewed — the estimate of asymmetric parameter

[9] The parameters are estimated by maximizing the log-likelihood function in (4) using CML and MAXLIK procedures in GAUSS.

[10] Preliminary analyses indicated that data on the core CPI inflation rate also display ARFIMA-NM-GARCH effects. We focused on the inflation data because these are the data examined in most of the studies that our exercise is compared with. Thus, to conserve space, we did not include the core inflation data in our paper.

TABLE 2

PARAMETER ESTIMATES FOR U.S. INFLATION

Model	ARMA-GARCH Models			ARFIMA-GARCH Models			ARFIMA-NM-GARCH	
	Normal	Student-t	Skewed-t	Normal	Student-t	Skewed-t	Symmetric	Asymmetric
μ	0.8946 (0.1306)	0.9504 (0.1139)	0.9359 (0.1033)	0.5629 (0.0781)	0.4613 (0.0966)	0.5331 (0.1052)	0.5703 (0.0884)	0.6118 (0.0980)
ϕ_1	0.9978 (0.0033)	0.9989 (0.0031)	0.9976 (0.0034)	-	-	-	-	-
θ_1	0.7876 (0.0326)	0.7751 (0.0314)	0.7706 (0.0321)	-	-	-	-	-
d	-	-	-	0.4582 (0.0357)	0.4706 (0.0378)	0.4623 (0.0367)	0.4935 (0.0380)	0.4786 (0.0381)
ω_1	0.0029 (0.0008)	0.0024 (0.0013)	0.0025 (0.0013)	0.0028 (0.0010)	0.0022 (0.0014)	0.0028 (0.0016)	0.0186 (0.0354)	0.0101 (0.0138)
α_1	0.1571 (0.0312)	0.1747 (0.0583)	0.1902 (0.0644)	0.1128 (0.0280)	0.1390 (0.0528)	0.1707 (0.0644)	0.2322 (0.3702)	0.1969 (0.2107)
β_1	0.7992 (0.0328)	0.8018 (0.0553)	0.7900 (0.0560)	0.8360 (0.0391)	0.8279 (0.0591)	0.8031 (0.0617)	0.8005 (0.2910)	0.8447 (0.1625)
$\alpha_1+\beta_1$	0.9563	0.9765	0.9802	0.9488	0.9669	0.9738	1.0327	1.0416
p_1	-	-	-	-	-	-	0.1444 (0.0896)	0.2210 (0.1043)
λ_1	-	-	-	-	-	-	-	0.0714 (0.0546)
υ	-	5.0451 (1.3554)	4.9414 (1.3083)	-	5.4006 (1.5379)	4.7522 (1.2616)	-	-
ξ	-	-	1.0741 (0.0730)	-	-	1.1881 (0.0810)	-	-
ω_2	-	-	-	-	-	-	0.0008 (0.0008)	0.0007 (0.0008)
α_2	-	-	-	-	-	-	0.0853 (0.0348)	0.0980 (0.0437)
β_2	-	-	-	-	-	-	0.8509 (0.0579)	0.8159 (0.0740)
$\alpha_2+\beta_2$	-	-	-	-	-	-	0.9362	0.9139
p_2	-	-	-	-	-	-	0.8556	0.7790
λ_2	-	-	-	-	-	-	-	-0.0203
LLK	48.3167	64.1311	64.5567	51.5864	65.3002	67.9212	68.7334	70.1706

Note: The table present ML estimates. Standard errors are given in parentheses. The models' log likelihood values are given in the row labeled "LLK."

(ξ) is positively significant — indicating that the distribution is positively skewed.

It is interesting to note that modifying the distributional assumption from normal to Student-t or to skewed Student-t does not noticeably affect the ARMA and GARCH estimates. In comparing the log likelihood values, the model with normal errors delivers the worst performance while the one with the skewed Student-t distribution is marginally better than the one with the Student-t distribution.

There is long memory in inflation data. The fractional parameter estimates are significant under each of the three distributional assumptions and all are less than 0.5 — implying considerable long-term persistence in the U.S. inflation data. The inclusion of a fractional parameter makes the ARMA coefficients insignificant but does not have a large impact on the conditional variance equation estimates.[11] Overall, the introduction of long memory improves the model's goodness-of-fit. The log likelihood values of the three ARFIMA specifications are larger than those of the corresponding ARMA specifications.

Modifying the conditional variance specification to normal mixtures yields a noticeable improvement in performance. Specifically, the log likelihood values of the selected ARFIMA-NM-GARCH models are quite large comparable with those of the selected ARMA-GARCH and ARFIMA-GARCH models with the normality assumption. Among the two selected ARFIMA-NM-GARCH models, the one restricting the means of the component normal processes to be zero (that is, $\lambda_1=\lambda_2=0$) garners a smaller log likelihood value. The result suggests the U.S. inflation data have an asymmetric distribution — a result that is in accordance with the skewed Student-t estimation results.

An astute observer will point out that a simple comparison of log likelihood values is not a vigorous way to select a specification among these different models. Because the models under examination are not all properly nested, there is no simple testing procedure to compare their degrees of goodness-of-fit. In the next sub-section, we will present a few additional model comparison measures.

The estimates of the mixing parameters are consistent with the presence of two GARCH processes driving the conditional volatility of inflation. The component GARCH process associated with a smaller mixing parameter estimate has a level of persistence, measured by the sum of ARCH and GARCH parameter estimates, similar to those estimated from the other simple GARCH processes. Also, the component GARCH process with a larger mixing parameter estimate is less persistence. Note that the standard errors of the estimates are higher

[11] Both the long memory and GARCH effects are quite comparable to those reported in, say Baillie *et al.* (1996). It is noted that the incorrect exclusion of long memory could lead to spuriously significant ARMA estimates because these estimates assume the serial correlation in data.

for the parameters of the component GARCH process that has a smaller mixing parameter estimate. That is, we have to interpret the persistence estimate given by the α_1- and β_1- estimates under the normal mixture specification with caution.

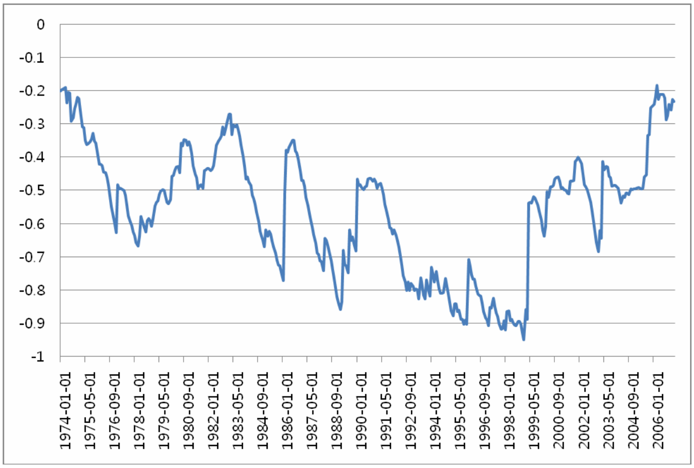
Since the mixing parameter estimates can be interpreted as the occurrence frequencies, we note that the process generating the U.S. inflation data includes two distinct volatility regimes — one has a higher occurrence frequency and relatively lower level of persistence. Anecdotal evidence suggests that the U.S. inflation has experienced some infrequent sharp movements induced by, say, changes in the monetary policy in the early 1980s and steep changes in commodity prices in the early 1970s and the 2000s. Our result is consistent with the interpretation of the presence of a high volatile GARCH process that induces some spikes in the U.S. inflation data. Further, without the restriction of $\lambda_1=\lambda_2=0$, the asymmetric specification offers a sharper estimate of the mixing parameter than the symmetric version that imposes the restriction.

The parameter estimates obtained from the symmetric and asymmetric ARFIMA-NM-GARCH models are quite similar. One subtle variation is the relative magnitude of the estimates across the two component GARCH processes. For the asymmetric model, both the ARCH (α_1) and GARCH (β_1) estimates are smaller for the component GARCH process that has a larger mixing parameter estimate. For the symmetric model, the ARCH effect is inversely related to and the GARCH effect, on the other hand, is positively related to the mixing parameter (*c.f.* Haas *et al.*, 2002 and 2004a, 2004b).

In Figure 2, we plot the conditional skewness and conditional kurtosis estimates extracted from the fitted asymmetric ARFIMA-NM-GARCH model. Both conditional skewness and conditional kurtosis estimates exhibit substantial time-variability. While the time variation in Kurtosis may be captured by other GARCH type models, the time-varying conditional skewness in the U.S. inflation data could present some challenge for these models.

In sum, the proposed model offers a good description of the inflation data. The results reported in Table 2 show that both the long memory feature and the normal mixture GARCH specification help improve the model performance.

A. Conditional skewness



B. Conditional kurtosis

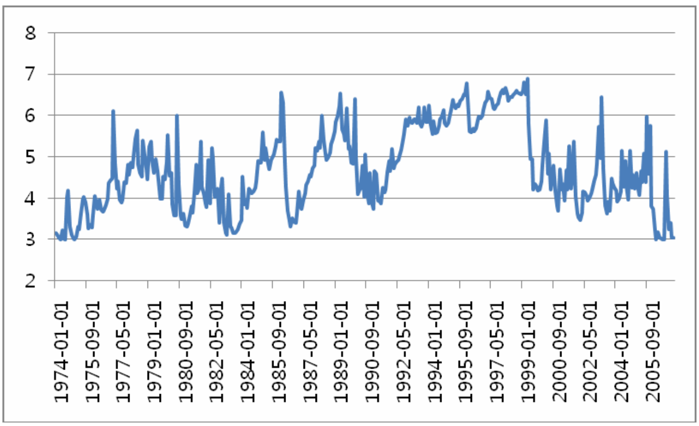


FIGURE 2

THE CONDITIONAL SKEWNESS AND KURTOSIS ETSIMTATES OF THE ASYMMETRIC ARFIMA-NM-GARCH

C. Model Selection and Some Diagnostics

In this sub-section, we offer a few measures that compare model performance. As pointed out earlier, the model specifications considered in Table 2 are not properly nested models. To further complicate the issue, the estimated standardized residuals of an ARFIMA-NM-GARCH model would not be identically distributed even if it is correctly specified. Thus, it is not appropriate to directly evaluate the distributional properties of is estimated residuals, $\hat{\varepsilon}_t$'s.

In Table 3, we first present the values of AIC and BIC. The AIC ranks the ARFIMA-GARCH with a skewed student-*t* distribution above the asymmetric ARFIMA-NM-GARCH model. The BIC, on the other hand, selects the ARFIMA-GARCH with a student-*t* distribution.

In the reminding part of the Table 3, we present some diagnostic results based on estimated residuals. For the standard GARCH type models, the usual methods are used to obtain their standardized residuals. For an ARFIMA-NM-GARCH model, we transform the estimated residuals such that they have a standard normal distribution under the null hypothesis of the model is correctly specified. Specifically, the residuals are transformed according to:

$$\hat{u}_t = \sum_{j=1}^{k} p_j \Phi_j(\hat{\varepsilon}_t), \ t=1, 2, \ldots, n, \tag{6}$$

where $k(=2)$ is the number of normal densities in the mixture, and Φ_k is the standard normal distribution function of the *j-th* element of the mixture. Under the null hypothesis, $\hat{u}_t$ will be independently and uniformly distributed and the inverse of the cumulative standard normal distribution of $\hat{u}_t$, given by $Z_t = \Phi^{-1}(\hat{u}_t)$, is distributed *iid* $N(0, 1)$ and does not exhibit any serial autocorrelation.

Following Alexander and Lazar (2006), Harvey and Siddique (1999), and Newey (1985), we implement a cumulative test to check the following conditions: $E(Z_t)=0$, $E(Z_t^2-1)=0$, $E(Z_t^3)=0$, $E(u_t^4-3)=0$, $E(Z_t Z_{t-j})=0$, $E(Z_t^2 Z_{t-j}^2)=0$, $E(Z_t^3 Z_{t-j}^3)=0$, and $E(Z_t^4 Z_{t-j}^4)=0$, for $j=1, 2, \ldots, 4$.

The results of the cumulative test show that none of the selected model passed all the moment restrictions. It should not be too alarming because it is well-known that the cumulative test is quite stringent for most practical applications. There is no specific rejection pattern revealed in the Table 3. We do not want to over-play it — however, it is comforting to observe that the asymmetric ARFIMA-NM-GARCH model gives the smallest number of rejection statistics in the Table. At the same time, recall that it is the same model specification yields the largest log likelihood value.

Next, we examine the skewness and kurtosis of the properly transformed residuals. The results presented in Table 3 suggest that the asymmetric ARFIMA-NM-GARCH model is the only model that yields

TABLE 3

SOME DIAGNOSTIC MEASURES

Model	ARMA-GARCH Models			ARFIMA-GARCH Models			ARFIMA-NM-GARCH	
	Normal	Student-*t*	Skewed-*t*	Normal	Student-*t*	Skewed-*t*	Symmetric	Asymmetric
AIC	-0.2126	-0.2871	-0.2842	-0.2341	-0.2980	-0.3061	-0.3002	-0.3024
BIC	-0.1525	-0.2170	-0.2041	-0.1840	-0.2379	-0.2360	-0.2100	-0.2022
$E[\hat{Z}_t]=0$	1.2359	0.4943	1.6395	4.5018*	1.3068	4.4295*	5.5300*	3.6892
$E[\hat{Z}_t^2-1]=0$	0.1272	283.65#	1.2296	2.9734	253.59#	10.707#	1.5839	6.7211#
$E[\hat{Z}_t^3]=0$	0.7238	0.1646	0.5368	1.4032	0.2798	1.2301	11.023#	0.8022
$E[\hat{Z}_t^4-3]=0$	11.187#	48.986#	15.154#	16.989#	73.933#	17.534#	0.4692	1.3429
$E[\hat{Z}_t\hat{Z}_{t-1}]=0$	30.335#	25.123#	24.197#	1.2359	0.7040	1.0398	0.5832	0.1804
$E[\hat{Z}_t\hat{Z}_{t-2}]=0$	7.3499#	7.8398#	8.1905#	8.9929#	10.570#	9.9119#	11.292#	11.981#
$E[\hat{Z}_t\hat{Z}_{t-3}]=0$	1.8772	2.1511	2.2299	0.0031	0.0120	0.0041	0.0406	0.0262
$E[\hat{Z}_t\hat{Z}_{t-4}]=0$	6.8101#	7.5196#	7.8123#	0.7385	1.1166	1.0152	0.8105	0.6175
$E[\hat{Z}_t^2\hat{Z}_{t-1}^2]=0$	0.0533	0.0652	0.3465	0.0031	0.1530	0.3193	0.7702	0.8267
$E[\hat{Z}_t^2\hat{Z}_{t-2}^2]=0$	0.7492	2.1409	0.4573	1.3803	2.7092	0.6323	1.8651	3.9185*
$E[\hat{Z}_t^2\hat{Z}_{t-3}^2]=0$	0.8020	0.8994	1.0999	0.2431	0.8724	0.9564	0.2684	0.5803
$E[\hat{Z}_t^2\hat{Z}_{t-4}^2]=0$	2.6644	0.0699	2.9051	2.2892	0.2323	3.5547	1.6169	1.0127
$E[\hat{Z}_t^3\hat{Z}_{t-1}^3]=0$	2.7720	2.2370	2.1273	1.1322	0.8803	0.9715	1.4143	0.7744
$E[\hat{Z}_t^3\hat{Z}_{t-2}^3]=0$	2.9799	2.9977	3.2708	3.2044	3.8336*	4.1108	7.1509#	6.9674#
$E[\hat{Z}_t^3\hat{Z}_{t-3}^3]=0$	0.7648	0.6016	0.6936	0.1932	0.1730	0.3766	0.0570	0.1582
$E[\hat{Z}_t^3\hat{Z}_{t-4}^3]=0$	8.6240#	7.4409#	7.5288#	5.5389*	5.3992*	5.5576	6.4586*	5.4998*
$E[\hat{Z}_t^4\hat{Z}_{t-1}^4]=0$	0.0289	0.4534	0.0002	0.1659	0.6721	0.0015	0.0326	0.0924
$E[\hat{Z}_t^4\hat{Z}_{t-2}^4]=0$	0.4662	1.6726	0.2439	0.6726	2.0682	0.1027	0.8976	3.7653
$E[\hat{Z}_t^4\hat{Z}_{t-3}^4]=0$	12.279#	5.2447*	13.667#	12.596#	4.2756*	14.212#	3.0450	2.7189
$E[\hat{Z}_t^4\hat{Z}_{t-4}^4]=0$	9.0653	0.8924	8.8420#	9.9091#	1.8575	11.713#	4.3614*	2.6672
Skewness	-0.2331*	-0.6642#	-1.6128#	-0.2891*	-0.4647#	-1.6149#	-0.2103*	-0.1646
Kurtosis	5.9158#	5.9473#	5.9577#	5.3671#	5.3849#	5.3703#	3.1567	3.3528
$Q(10)$	0.3625	0.3894	0.3821	0.0008	0.0048	0.0015	0.0016	0.0016
$Q^2(10)$	7.2114	7.1581	7.1475	7.2291	7.1844	7.1657	6.3861	6.4962
ACF	0.0352	0.0879	0.1185	0.0197	0.0431	0.0736	0.0190	0.0126

Note: AIC (BIC) gives the AIC (BIC) values of the models. The conditional moment tests of the selected models are reported under the $E[.]$ rows. "Skewness" denotes the skewness coefficient, γ_1 and "Kurtosis" the kurtosis coefficient, γ_2. Under normality, $T\gamma_1^2/6\sim\chi^2(1)$ and $T(\gamma_2-3)^2/24$ $\sim\chi^2(1)$ asymptotically. The $Q(10)$ and $Q^2(10)$ are the Ljung-Box statistics for first ten serial correlation coefficients of the residuals and their squares. Asterisks * and # indicate significance at the 5% and 1% levels, respectively. ACF gives the mean squared errors of the correlation coefficient estimates of the squared residuals.

insignificant skewness and kurtosis coefficient estimates. The symmetric ARFIMA-NM-GARCH model passes the kurtosis test but not the skewness one; indicating the relevance of the ability to model skewness. The estimated residual of all other models under consideration, including the ARFIMA-GARCH model with a skewed t-distribution, are found to have some significant degrees of skewness and kurtosis. That is, these models do not adequately describe the skewness and kurtosis in the U.S. inflation data.

The transformed residuals are used to calculate the Ljung and Box (1978) Q-statistic. Specifically, we calculate the Q-statistics based on the first ten autocorrelation coefficient estimates derived from the transformed residuals and their squares and label them $Q(10)$ and $Q^2(10)$ in the Table. For all the models under consideration, there is no significant temporal dependency in the residuals and their squares. That is, these models offer a reasonable a specification to describe the serial in the U.S. inflation data and their squares.

Last, but not the least, we assess the ability of ARFIMA-NM-GARCH models to capture the autocorrelations of the squared residuals. To this end, for each model, we compare its empirical and theoretical autocorrelation coefficients of the squared residuals.[12] In Table 3, the row labeled "ACF" reports the mean squared prediction errors for the first 250 lags of the squared residual autocorrelation.[13] It is evidence that the ARFIMA-NM-GARCH models yields the two smallest mean squared errors with the asymmetric version has the smallest error. While all the models under consideration offer a good description of conditional heteroskedasticity, the asymmetric ARFIMA-NM-GARCH model show the smallest deviation from the theoretically predicted conditional heteroskedasticity pattern.

The overall evidence from Table 3 and the log likelihood values in Table 2 is in favor of the asymmetric ARFIMA-NM-GARCH model, which has the flexibility to describe both the time-varying conditional heteroskedasticity and conditional skewness.

IV. Summary

In this exercise, we introduce a class of models that incorporates two interesting time series features; namely long memory and conditional heteroskedasticity given by a normal mixture GARCH specification. We label it an ARFIMA-NM-GARCH model. The long memory component

[12] The theoretical autocorrelation functions of the squared residuals of these models are given in the Appendix.

[13] The mean squared prediction errors derived from different numbers of correlation coefficient estimates give qualitatively similar results.

offers a flexible means to describe data persistence including stationary long-term persistence. The normal mixture GARCH component extends the standard GARCH framework and allows the conditional volatility to be determined by more than one GARCH processes. Also, a desirable property of the mixture process is its ability to model time variations in higher conditional moments including skewness and kurtosis.

The U.S. inflation data are used to illustrate the empirical relevance of the proposed model. The evidence suggests that the inflation data exhibit long memory persistence in levels and their (conditional) volatility is driven by two GARCH processes. The proposed ARFIMA-NM-GARCH model, indeed, compares quite favorably with some alternative ARFIMA and GARCH models used in the literature. Specifically, the asymmetric ARFIMA-NM-GARCH model is found to be the only model, amongst those considered, that captures the skewness and kurtosis in the data. The model also generated empirical correlation coefficients of the squared residuals that have the smallest deviation from their theoretical values. The empirical application highlights the potential benefits of integrating long memory and mixed normal GARCH in modeling economic data and the flexibility of modeling data asymmetry.

There are several ways to extend the current study. For instance, it is of interest to consider time-varying mixing parameters that depend on some relevant fundamental economic variables. The component GARCH process can also be modified to accommodate some specific volatility characteristics including differential effects of large and small shocks.[14] These extensions should be left for future studies.

(Received 23 June 2008; Revised 8 May 2009)

Appendix

A. Student-t and Skewed Student-t Distributions

The density function of a random variable z that follows a Student-t distribution (that is $z \sim t(\upsilon)$) is given by

[14] Some alternative specifications of normal mixture GARCH models are considered in, for example, Bai *et al.* (2003), Ding and Granger (1996), and Vlaar and Palm (1993).

$$g(z \mid v) = \Gamma\left(\frac{v+1}{2}\right)\left(1+\frac{z^2}{v-2}\right)^{-\frac{v+1}{2}} / \left[\sqrt{(v-1)\pi}\,\Gamma\left(\frac{v}{2}\right)\right], \tag{A1}$$

where v is the degree of freedom. The distribution approaches a normal distribution as v is approaching infinity.

If z follows a skewed Student-t distribution (that is, $z \sim Skewed \cdot t(\xi, v)$), then its density function is given by

$$f(z \mid \xi,\ v) = 2/(\xi+\xi^{-1})\, s\{g[\xi(sz+e_1) \mid v]\, I(z+e_1/s) + g[(sz+e_1)/\xi \mid v]\, I(z+e_1/s)\}, \tag{A2}$$

where $g[\cdot]$ is the density of the Student-t distributions in (A1), and I is an indicator function

$$I_t(z) = \begin{bmatrix} 1 \xleftarrow{if} z_t \geq -e_1/s \\ -1 \xleftarrow{if} z_t < -e_1/s \end{bmatrix},$$

$$e_r = e_r(v, \xi) = E(y^r \mid v, \xi) = M_r(v) \cdot \left[\xi^{r+1} + \frac{(-1)^r}{\xi^{r+1}}\right] / \left[\xi + \frac{1}{\xi}\right],$$

$$M_r(v) = \Gamma\left(\frac{v-r}{2}\right)\Gamma\left(\frac{1+r}{2}\right)(v-2)^{\frac{r}{2}} / \sqrt{\pi}\,\Gamma\left(\frac{v}{2}\right),$$

and

$$s = \sqrt{Var(y \mid \xi, v)} = E(y^2 \mid \xi, v) - [E(y \mid \xi, v)] = e_2 - e_1^2.$$

B. Autocorrelation Functions of the Squared Residuals

For the ARFIMA(0, d, 0)-NM(2)-GARCH(1, 1), the overall variance is $\sigma_t^2 = \sum_{i=1}^{k} p_i \sigma_{it}^2 + \sum_{i=1}^{k} p_i \lambda_i^2$, where $\sigma_{it}^2 = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{it-1}^2$ for $i=1$, 2 and the conditional skewness and kurtosis are then given by

$$\tau_t = \frac{E_{t-1}(\varepsilon_t^3)}{(\sigma_t^2)^{3/2}} = \frac{3\sum_{i=1}^{k} p_i \lambda_i \sigma_{it}^2 + \sum_{i=1}^{k} p_i \lambda_i^3}{(\sigma_t^2)^{3/2}}, \tag{B1}$$

$$\kappa_t = \frac{E_{t-1}(\varepsilon_t^4)}{(\sigma_t^2)^2} = \frac{3\sum_{i=1}^{k} p_i (\sigma_{it}^2)^2 + 6\sum_{i=1}^{k} p_i \lambda_i^2 \sigma_{it}^2 + \sum_{i=1}^{k} p_i \lambda_i^4}{(\sigma_t^2)^2}. \tag{B2}$$

Taking expectations of the overall and individual unconditional variances gives $x = E(\varepsilon_t^2) = E(\sigma_t^2) = \left[\sum_{i=1}^{k} p_i\lambda_i^2 + \sum_{i=1}^{k}\frac{p_i\omega_i}{(1-\beta_i)}\right] / \left[\sum_{i=1}^{k}\frac{p_i(1-\alpha_i-\beta_i)}{(1-\beta_i)}\right]$ and $y_j = E(\sigma_{jt}^2) = [\omega_j + \alpha_j x]/[1-\beta_j]$ for $i, j = 1, 2$ respectively. Then the unconditional skewness is given by

$$\tau = \frac{\sum_{i=1}^{k} p_i(3y_i\lambda_i + \lambda_i^3)}{(x)^{3/2}}. \tag{B3}$$

and the unconditional kurtosis is given by

$$\kappa = E(\varepsilon_t^4)/E(\varepsilon_t^2)^2 = z/x^2, \tag{B4}$$

where $z = \left(\frac{3P'B^{-1}f+s}{1-3P'B^{-1}g}\right)$, where $P = (p_1, p_2)'$, $s = \sum_{i=1}^{k} p_i(6\lambda_i^2 y_i^2 + \lambda_i^4)$, $q = \sum_{k=1}^{k} p_k\lambda_k^2$, $w_i = \omega_i^2 + 2\omega_i\alpha_i x + 2\omega_i\beta_i y_i$, $r_{ij} = w_i w_j + x(\omega_i\alpha_j + \omega_j\alpha_i) + \beta_i y_i\omega_j + \beta_j y_j\omega_i$ for $i, j = 1, 2$,

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1-\dfrac{p_2\beta_1\alpha_2}{1-\beta_1\beta_2} & -\dfrac{p_2\beta_1\alpha_2}{1-\beta_1\beta_2} \\ -\dfrac{p_2\beta_1\alpha_2}{1-\beta_1\beta_2} & 1-\dfrac{p_2\beta_1\alpha_2}{1-\beta_1\beta_2} \end{bmatrix},\quad c_i = \sum_{j=1}^{k} a_{ij}\left(\sum_{\substack{k=1\\k\neq j}}^{k}\frac{p_k r_{jk}}{(1-\beta_j\beta_k)} + y_j q\right),$$

$$d_i = \sum_{j=1}^{k} a_{ij}\left(\sum_{\substack{k=1\\k\neq j}}^{k}\frac{p_k\alpha_j\alpha_k}{(1-\beta_j\beta_k)}\right),\ e_{ij} = a_{ij}p_j,\ B = \begin{bmatrix} 1-\beta_1^2-2\alpha_1\beta_1 e_{11} & -2\alpha_1\beta_1 e_{12} \\ -2\alpha_2\beta_2 e_{21} & 1-\beta_2^2-2\alpha_2\beta_2 e_{22} \end{bmatrix},$$

$$f = \begin{bmatrix} w_1 + 2\alpha_1\beta_1 c_1 \\ w_2 + 2\alpha_2\beta_2 c_2 \end{bmatrix},\ \text{and}\ \ g = \begin{bmatrix} \alpha_1^2 + 2\alpha_1\beta_1 d_1 \\ \alpha_1^2 + 2\alpha_2\beta_2 d_2 \end{bmatrix}.$$

For the ARFIMA(0, d, 0)-NM(2)-GRACH(1, 1) model given by (1) to (3), the autocorrelation function of the squared residuals is

$$\rho_k = Corr(\varepsilon_t^2, \varepsilon_{t-k}^2) = \frac{Cov(\varepsilon_t^2, \varepsilon_{t-k}^2)}{Var(\varepsilon_t^2)} = \frac{E(\varepsilon_t^2, \varepsilon_{t-k}^2) - x^2}{E(\varepsilon_t^4) - x^2} = \frac{c_k - x^2}{z - x^2}, \tag{B5}$$

where $c_k = E(\varepsilon_t^2, \varepsilon_{t-k}^2) = x\sum_{i=1}^{k} p_i\lambda_i^2 + \sum_{i=1}^{k} p_i b_{ik}$ for $K = 1, 2$ with $b_{ik} = \omega_i x + \alpha_i c_{k-1} + \beta_i b_{ik-1}$, $c_0 = z$ and $b_{i0} = c_i + d_i z + e_i' B^{-1}(f + gz)$ for $i = 1, 2$.

The autocorrelation function of the squared residuals in other models is given by

$$\rho_k = Corr(\varepsilon_t^2,\ \varepsilon_{t-k}^2) = \frac{Cov(\varepsilon_t^2,\ \varepsilon_{t-k}^2)}{Var(\varepsilon_t^2)} = \frac{E(\varepsilon_t^2,\ \varepsilon_{t-k}^2) - x^2}{E(\varepsilon_t^4) - x^2}$$

$$= \left(\alpha + \frac{\alpha^2\beta}{1-2\alpha\beta-\beta^2}\right)(\alpha+\beta)^{k-1}. \qquad \text{(B6)}$$

References

Alexander, C., and Lazar, E. Normal Mixture GARCH(1,1): Applications to Exchange Rate Modelling. ISMA Centre Discussion Papers in Finance, UK: ISMA Centre, University of Reading, 2004.

__________. "Normal Mixture GARCH(1,1): Applications to Exchange Rate Modelling." *Journal of Applied Econometrics* 21 (No. 3 2006): 307-36.

Bai, X., Russell, J. R., and Tiao, G. C. "Kurtosis of GARCH and Stochastic Volatility Models with Non-normal Innovations." *Journal of Econometrics* 114 (No. 2 2003): 349-60.

Baillie, R. T., and Bollerslev, T. "The Message in Daily Exchange Rates: A Conditional-Variance Tale." *Journal of Business and Economic Statistics* 7 (No. 3 1989): 297-305.

Baillie, R. T., Chung, C. F., and Tieslau, M. A. The Long Memory and Variability of Inflation: A Reappraisal of the Friedman Hypothesis. Discussion Paper, Tilburg University, 1992.

__________. "Analyzing Inflation by the Fractionally Integrated ARFIMA-GARCH Model." *Journal of Applied Econometrics* 11 (No. 1 1996): 23-40.

Bartlett, M. S. "On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series." *Journal of the Royal Statistical Society* B 8 (No. 8 1946): 27-41.

Bauwens, L., Laurent, S., and Rombouts, J. V. K. "Multivariate GARCH Models: A Survey." *Journal of Applied Econometrics* 21 (No. 1 2006): 79-109.

Bollerslev, T. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31 (No. 3 1986): 307-27.

__________. "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return." *Review of Economics and Statistics* 69 (No. 3 1987): 542-47.

Cheung, Y. W. "Long Memory in Foreign Exchange Rates." *Journal of Business & Economic Statistics* 11 (No. 1 1993): 93-101.

Cheung, Y. W., and Chung, S. K. A Long Memory Model with Mixed Normal GARCH. Working Paper, UCSC and INJE University, 2007.

Cheung, Y. W., and Lai, K. S. "Do Gold Market Returns Have Long Memory?" *Financial Review* 28 (No. 2 1993): 181-202.

__________. "Long Memory and Nonlinear Mean Reversion in Japanese Yen-Based Real Exchange Rates." *Journal of International Money and Finance* 20 (No. 1 2001): 115-32.

Diebold, F. X., and Rudebusch, G. D. "Is Consumption Too Smooth? Long Memory and the Deaton Paradox." *Review of Economics and Statistics* 73 (No. 1 1991): 1-17.

Ding, Z., and Granger, C. W. J. "Modeling Volatility Persistence of Speculative Returns: A New Approach." *Journal of Econometrics* 73 (No. 1 1996): 185-215.

Doornik, J. A., and Ooms, M. "Inference and Forecasting for ARFIMA Models With an Application to US and UK Inflation." *Studies in Nonlinear Dynamics & Econometrics* 8 (No. 2 2004): 1-23.

Engle, R. F. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50 (No. 4 1982): 987-1007.

Geweke, J., and Porter-Hudak, S. "The Estimation and Application of Long Memory Time Series Models." *Journal of Time Series Analysis* 4 (No. 4 1983): 221-38.

Granger, C. W. J. "Long Memory Relationships and the Aggregation of Dynamic Models." *Journal of Econometrics* 14 (No. 2 1980): 227-38.

Granger, C. W. J., and Joyeux, R. "An Introduction to Long Memory Series Models and Fractional Differencing." *Journal of Time Series Analysis* 1 (No. 1 1980): 15-29.

Haas, M., Mittnik, S., and Paolella, M. S. Mixed Normal Conditional Heteroskedasticity. CFS Working Paper No. 10, 2002.

__________. "Mixed Normal Conditional Heteroskedasticity." *Journal of Financial Econometrics* 2 (No. 2 2004a): 211-50.

__________. "A New Approach to Markov Switching GARCH Models." *Journal of Financial Econometrics* 2 (No. 4 2004b): 493-530.

Harvey, C. R., and Siddique, A. "Autoregressive Conditional Skewness." *Journal of Financial and Quantitative Studies* 34 (No. 4 1999): 465-87.

Hassler, U., and Wolters, J. "Long Memory in Inflation Rates: International Evidence." *Journal of Business and Economic*

Statistics 13 (No. 1 1995): 37-45.

Hosking, J. R. M. "Fractional Differencing." *Biometrika* 68 (No. 1 1981): 165-76.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics* 54 (Nos. 1-3 1992): 159-78.

Lambert, P., and Laurent, S. Modelling Skewness Dynamics in Series of Financial Data Using Skewed Location-Scale Distributions. Institut de Statistique, Louvain-la-Neuve Discussion Paper 0119, 2001a.

__________. Modelling Financial Time Series Using GARCH-Type Models with a Skewed Student Distribution for the Innovations. Institut de Statistique, Louvain-la-Neuve Discussion Paper 0125, 2001b.

Ling, S., and Li, W. K. "On Fractionally Integrated Autoregressive Moving-Average Time Series Models with Conditional Heteroskedasticity." *Journal of the American Statistical Association* 92 (No. 439 1997): 1184-94.

Ljung, G. M., and Box, G. E. P. "On a Measure of Lack of Fit in Time Series Models." *Biometrika* 65 (No. 2 1978): 297-303.

Newey, W. K. "Generalized Method of Moments Specification Testing." *Journal of Econometrics* 29 (No. 3 1985): 229-56.

Palm, F., and Vlaar, P. J. G. "Simple Diagnostics Procedures for Modeling Financial Time Series." *Allgemeines Statistisches Archiv* 81 (No. 1 1997): 85-101.

Seo, B., and Kim, K. "Rational Expectations, Long-run Taylor Rule, and Forecasting Inflation." *Seoul Journal of Economics* 20 (No. 2 2007): 239-62.

Vlaar, P. J. G., and Palm, F. C. "The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps." *Journal of Business & Economic Statistics* 11 (No. 3 1993): 351-60.