Conditional Heteroskedasticity Modeling

in Problems of Multiperiod Prediction

Stanislav Anatolyev*

New Economic School, Moscow

Abstract

When the disturbance is a multiperiod prediction error and one wants to model its conditional variance within the ARCH framework, it is conventional to impose a moving average structure on the error term making the innovation a martingale difference sequence following ARCH. We argue that this is overly restrictive in several respects and suggest a maximum likelihood approach that does not entail this assumption. We apply this framework to two empirical problems. One is re-estimation of the relationship between the term structure of interest rates and future inflation at short horizons, and the other is a question of whether disagreement is an appropriate measure for forecast uncertainty.

JEL Classification Codes C22, C51, C53

Key Words: ARCH, serial correlation, prediction error, maximum likelihood

---

* Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Prospekt, 47, Moscow, 117418 Russia. Phone: 7-095-129-1700. Fax: 7-095-129-3722. E-mail: sanatoly@nes.ru. I thank seminar audiences at the New Economic School and the 2001 North American Summer Meeting of Econometric Society at the University of Maryland. Research for this paper was supported in part by a fellowship from the Economics Education and Research Consortium, with funds provided by the Government of Sweden through the Eurasia Foundation. The opinions expressed in this paper are those of the author and do not necessarily reflect the views of the Government of Sweden, the Eurasia Foundation, or any other member of the Economics Education and Research Consortium.

1 Introduction

Since Robert Engle's (1982) paper on autoregressive heteroskedasticity there has been an outburst of research that suggests various generalizations and extensions to the basic ARCH model. They have gone along several major directions. First, many researchers sought more flexible specifications for the conditional variance that would better fit the observed economic series – e.g., Generalized ARCH of Bollerslev (1986) and Nonlinear ARCH of Higgins and Bera (1992) – or provide an account for some important features of the data such as, for instance, the leverage effect – e.g., Exponential ARCH of Nelson (1991) and Threshold ARCH of Zakoïan (1994). Second, to explain high leptokurticity or skewness of the data, the original conditional normality assumption was reconsidered – e.g., GARCH-t of Bollerslev (1987) and GARCH-EGB2 of Wang, Fawson, Barrett and McDonald (2001). Third, to explain long memory in volatility, the conditional variance was modelled as Integrated GARCH of Engle and Bollerslev (1986) and Fractionally Integrated GARCH of Baillie, Bollerslev and Mikkelsen (1996). Finally, to explain time-varying risk premia, a feedback between the mean and variance was modeled as ARCH-in-mean of Engle, Lilien and Robins (1987). For excellent surveys from an econometric perspective, see Bera and Higgins (1993) and Bollerslev, Engle and Nelson (1994). The present paper explores another direction that has been ignored in the ARCH literature.

The ARCH-type modeling strategy is often structural for the conditional mean, but non-structural for the conditional variance. The reason is that while economic theory, usually based on some intertemporal optimization problem, often provides a functional form for the former, very rarely can it suggest such for the latter, because, as Pagan and Hong (1991) emphasize, there is no optimizing theory that could yield such form. The present paper considers the case when structural modeling of the mean together with a nature of the problem impose certain restrictions on how one is allowed to model the conditional variance. We consider multiperiod conditional moment restrictions, that is, the situation when prediction errors are overlapping (which sometimes is informally called *overlapping data*). A variety of intertemporal macroeconomic and financial models give rise to multiperiod conditional moment restrictions, where the moment function is serially correlated of known order: models with overlapping prediction horizons (Hodrick, 1987; Mishkin, 1990, 1992; Rich, Raymond and Butler, 1992), with temporal aggregation (Hall, 1988; Hansen and Singleton, 1996), or with complex decision rules (Eichenbaum, Hansen and Singleton, 1988). Some of the issues to be discussed are mentioned in Rich, Raymond and Butler (1992), but the authors decided not to stay in the maximum likelihood framework typical for ARCH models and instead turned to distribution-free GMM.

Suppose one has a problem of two-step-ahead prediction and wishes to model heteroskedasticity within the ARCH framework. A likely model may be the following:

$$y_{t+2} = x_t'\beta + e_{t+2}, \tag{1}$$

$$e_{t+2} = \varepsilon_{t+2} + \theta\varepsilon_{t+1}, \tag{2}$$

where the variables are indexed with the time when they get realized, and the conditional mean is linear for simplicity. Under optimal prediction, the entire error term e_{t+2} has mean zero conditional on time t information $\Im_t = \{x_t, y_t, x_{t-1}, y_{t-1}, \cdots\}$, which automatically includes the history of the disturbance. The variance equation is typically specified for the innovations ε_{t+1}. For example, it may take the GARCH(1,1) form

$$\varepsilon_{t+1}|\Im_t \sim \mathcal{N}(0, h_t), \quad h_t = \omega + \alpha\varepsilon_t^2 + \phi h_{t-1}, \tag{3}$$

that is, ε_{t+1} is conditionally Gaussian with conditional mean zero and conditional variance h_t. This type of modeling the evolution of the conditional variance has proved to often provide a good approximation to its actual evolution in the data. See, for example, Pagan and Schwert (1990), Baillie and DeGennaro (1990), McCurdy and Morgan (1991).

The error e_{t+2} is made at time t and is interpreted as the error of the optimal forecast of y_{t+2}. However, in (3) the part ε_{t+2} of the error is as though formed at time $t+1$, the fact that may seem strange. In addition, as Hayashi and Sims (1983) argued, the Wold innovation of the prediction error in a rational expectations model is in general *not* a martingale difference sequence relative to the history of innovations. Therefore, making the forecast error have the above structure is restrictive and arbitrary.[1] One more doubtful consequence of (1)–(3) is the following. If ε_{t+1} conditional on $\Im_t$ is Gaussian, the entire error term e_{t+2} is *not* Gaussian conditional on $\Im_t$. This can be seen by noting that while the component $\theta\varepsilon_{t+1}$ is conditionally normal, the

[1] Even for problems with serially uncorrelated errors, Francq and Zakoïan (2000) relax the assumption that the innovation forms a martingale difference.

other component, ε_{t+2}, is conditionally (on $\Im_t$) leptokurtic, the latter fact following by the virtue of the conditional Jensen inequality. Thus the idea to impose normality on the nonpredictable part of the predicted variable is not really carried out by (3).

We adapt a more natural set of assumptions about the error term. The assumptions will concern only the conditional behavior of the whole error, and not its innovations. We let the conditional distribution of the error to be Gaussian. However, to completely specify the likelihood function of the process, we have to impose distributional assumptions on the error conditional on *future* information sets. We set these distributions to be Gaussian too. Such strategy leads to a hybrid of structural and non-structural modeling of the mean and non-structural (ARCH-type) modeling of the variance.

The paper is organized as follows. Section 2 outlines ideas and sets forth the modeling strategy for a 2-step-ahead prediction problem, while a generalization is relegated to the Appendix. Sections 3 and 4 are devoted to applications: section 3 revisits estimation of Mishkin's (1990) relationship between the term structure and future inflation at short horizons, and section 4 attempts to resolve the dispute between Bomberger (1996, 1999) and Rich and Butler (1998) on whether disagreement is an appropriate measure for forecast uncertainty. Finally, section 5 concludes.

2 Modeling strategy

Consider the framework of the foregoing 2-step-ahead prediction problem. Let us denote $\mathcal{X}_t = \sigma(x_t, x_{t-1}, \cdots)$, $\mathcal{E}_t = \sigma(e_t, e_{t-1}, \cdots)$, and $\Im_t = \mathcal{X}_t \vee \mathcal{E}_t$. Following the original ARCH tradition, we would like to model e_{t+2} as conditional Gaussian[2]: $e_{t+2}|\Im_t \sim \mathcal{N}(0, h_t)$. This formulation is slack, however, since the likelihood function is not uniquely defined. Rich, Raymond and Butler (1992) used a similar approach, but without the normality assumption, and modeled h_t as a linear function of squared past prediction errors. Since they were dealing with prediction data from surveys, they called their model *Survey Data – ARCH*. Incomplete specification allows only GMM estimation, which leads to the problem of instrument selection and inefficiency.

We take an approach of full distributional specification. Assuming that T is even, the conditional part of the likelihood function can be partitioned as

$$f_{(e_1,e_2)|\Im_0}(e_1, e_2) \cdot \prod_{\substack{t=2 \\ t \text{ even}}}^{T-2} f_{(e_{t+1},e_{t+2})|\Im_t}(e_{t+1}, e_{t+2}).$$

If T is odd, the first factor in the latter partitioning is $f_{e_1|\Im_{-1}}(e_1)$, but this boundary effect does not matter asymptotically. Since we already have a restriction on $e_{t+2}|\Im_t$, it is natural to specify the joint distribution of $(e_{t+1}, e_{t+2})'$ conditional on $\Im_t$ to take advantage of the above partitioning and achieve unique identification of the likelihood. We set that distribution to bivariate normal:

$$\begin{pmatrix} e_{t+1} \\ e_{t+2} \end{pmatrix} |\Im_t \sim \mathcal{N}\left(\begin{bmatrix} \mu_t \\ 0 \end{bmatrix}, \begin{bmatrix} \omega_{t-1} & \lambda_t \\ \lambda_t & \omega_t + E[\mu_{t+1}^2|\Im_t] \end{bmatrix} \right), \tag{4}$$

where in addition $E[\mu_{t+1}|\Im_t] = 0$ and $\lambda_t^2 = \omega_{t-1}E[\mu_{t+1}^2|\Im_t]$. The newly introduced time varying parameters ω_t and λ_t are $\Im_t$-measurable, while the parameter μ_t is $(\mathcal{X}_{t-1} \vee \mathcal{E}_t)$-measurable. All these restrictions are imposed on the entries of the mean and variance

[2]Of course, the gaussianity assumption can be changed to another distributional specification. Whether the subsequent strategy still goes through is a topic of future research.

to make the distributional specification (4) intrinsically compatible, because both elements of the vector are essentially the same thing with a lag difference. Such specification is easily interpretable: at the moment when the prediction is made, the prediction error is conditionally Gaussian, and so is uncertainty left of the previous prediction period error when the forecaster wants to revise that prediction.

It is easy to check that under the specification (4) the conditional density of two successive disturbances partitions as

$$f_{(e_{t+1},e_{t+2})|\Im_t}\left(e_{t+1},e_{t+2}\right)=f_{e_{t+1}|\mathcal{X}_{t-1},\mathcal{E}_t}\left(e_{t+1}\right)f_{e_{t+2}|\mathcal{X}_t,\mathcal{E}_{t+1}}\left(e_{t+2}\right),$$

so that the distributional assumption (4) together with the compatibility restrictions is equivalent to

$$e_{t+1}|\mathcal{X}_{t-1},\mathcal{E}_t\sim\mathcal{N}\left(\mu_t,\omega_{t-1}\right) \tag{5}$$

subject to $E[\mu_{t+1}|\Im_t]=0$. Note that the conditional variance ω_{t-1} is $\Im_{t-1}$-measurable (cf. Rich, Raymond and Butler, 1992). This property makes intuitive sense, since we regard the error e_{t+1} as made at time $t-1$, so the variance of the residual part after time t is realized was determined at time $t-1$ too. The following gives a rough idea of the structure of the error e_{t+2}. It consists of two components, $\varepsilon_{1,t+1}$ and $\varepsilon_{2,t+2}$, that are made at time t and will be realized in the two subsequent periods. Both have $\Im_t$-conditional mean zero and $\Im_t$-measurable conditional variance. At time $t+1$ the first error component, $\varepsilon_{1,t+1}$, is realized, and conditional on this additional information, the forecaster would be willing to reconsider her forecast, if she could. That is, the second error component, $\varepsilon_{2,t+2}$, is not the optimal prediction error based on the expanded information set, but instead has $(\mathcal{X}_t \vee \mathcal{E}_{t+1})$-conditional mean μ_{t+1}. From the perspective of time t, this μ_{t+1} is unpredictable and hence has $\Im_t$-conditional mean zero.

It is interesting to contrast specification (5) with the conventional one for the Wold innovation: $\varepsilon_{t+1}|\Im_t \sim N(0, h_t)$. While (5) has a more flexible specification for the mean, that for the variance is more constrained. Therefore, neither model nests the other. Following Pagan and Hong's (1991) logic, we note that since our presumption is that the economic theory imposes only the restriction $E[e_{t+2}|\Im_t] = 0$ and says nothing about the structure of μ_t, the latter has to be modeled in a non-structural way, similarly to how ω_t is usually modeled. Thus our next step is to specify the parametric forms of μ_t and ω_t.

For ω_t, a natural strategy is to set it to a linear function of past squared prediction errors, similarly to Rich, Raymond and Butler's (1992) choice:

$$\omega_t = \omega + \sum_{i=1}^{q} \alpha_i e_{t-i+1}^2, \tag{6}$$

or to a modification of another member of the ARCH family. Of course, such specifications do not nest the conventional one. The conventional analog of (6) is a linear function of the past squared *innovations*, having which in the specification is unnatural in the present context. Besides, while working with stock returns indices, Pagan and Schwert (1990) indicated that in the conventional specification (1)–(3) with conditional variance from the ARCH family, no important discrepancies resulted from use of lags of e_t^2 instead of lags of ε_t^2 in the specification for conditional variance h_t in terms of fit.

To specify the form of μ_t, observe that due to the conditional bivariate normality[3],

$$\mu_{t+1} = \delta_t \left(e_{t+1} - \mu_t \right), \tag{7}$$

where δ_t is $\Im_t$-measurable, and $E[\log|\delta_t|] < 0$ in order for μ_t to be stationary (see Brandt, 1986). This has a natural interpretation: the forecaster wants to revise her initial prediction of e_{t+2} equalling zero, after the next period's variables are realized, by taking the fraction δ_t of the realized prediction error $e_{t+1} - \mu_t$ from the previous round, in the spirit of adaptive way of expectation formation. Since $\mu_t = E[e_{t+1}|\Im_t]$, from (7) it follows that e_{t+1} has a nonlinear moving-average representation

$$e_{t+2} = \nu_{t+2} + \delta_t \nu_{t+1}, \tag{8}$$

where $\nu_{t+1} = e_{t+1} - \mu_t$ is an error of the "revised" prediction, and the MA coefficient is $\Im_t$-measurable. The necessary and sufficient condition of invertibility of (8) is again $E[\log|\delta_t|] < 0$. In the special case when the adaptation parameter δ_t is time invariant, i.e. $\delta_t = \delta$ for all t, we see that (8) is the usual Wold representation for the disturbance. That is, the constancy of δ_t makes the prediction problem specified in the conventional MA-ARCH spirit[4].

[3]Indeed, since under normality the conditional expectations coincide with linear projections, projecting e_{t+2} on $(\mathcal{X}_t \vee \mathcal{E}_{t+1})$ using (4) yields $\mu_{t+1} = \omega_{t-1}^{-1} \lambda_t \left(e_{t+1} - \mu_t \right)$.

[4]Alternatively, one can see this by noting that when δ_t is time invariant, μ_t is a linear combination of all past errors. This means that $E[e_{t+1}|\Im_t]$ is the same as the linear predictor $\hat{E}[e_{t+1}|e_t, e_{t-1}, ...]$, which in turn implies that the Wold innovation in e_{t+1} is a martingale difference.

It is left to specify the form of δ_t. Note that although we know that $\delta_t = \omega_{t-1}^{-1}\lambda_t$, the only constraint put on δ_t is the stability condition $E[\log|\delta_t|] < 0$, as the form of λ_t is not constrained by (4)[5]. In general, δ_t may be specified in a variety of ways as some function of past prediction errors. It is desirable to specify it so that it is convenient to test for its constancy in time, which amounts to testing for the martingale difference hypothesis of the innovation of e_t. The form that we explore in applications is

$$\delta_t = \delta + \sum_{i=1}^{r} \gamma_i e_{t-i+1}, \tag{9}$$

which appears to be a natural choice: as one often models the conditional second moment by a quadratic form of past errors, it is natural to model the conditional first moment by a linear form of them. With the specification (9), the martingale difference hypothesis for the innovation of e_t can be tested via the null $H_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_r = 0$.

In the Appendix, we generalize this modeling framework for the case of a $(J+1)$-step-ahead prediction problem, where $J \geq 1$. In the following two sections we apply this framework for two prediction problems with $J = 2$ and $J = 1$.

3 Application 1: Term Structure and Future Inflation

Frederic Mishkin (1990) estimates a relationship between the term structure of nominal interest rates, on the one hand, and future inflation and real interest rate, on

[5] Although λ_t must satisfy $\lambda_t^2 = \omega_{t-1}E[\mu_{t+1}^2|\Im_t]$, this equality carries no information about its form, even after imposing (6) and (7).

the other, at horizons from 1 to 12 months. The main question that he asks is: how much information about changes in future inflation and changes in ex post real interest rate is embedded in the slope of the term structure given that expectations are rational? To answer this and related questions, Mishkin specifies the following econometric model:

$$\pi_t^m - \pi_t^n = \alpha_{m,n} + \beta_{m,n} \left(i_t^m - i_t^n \right) + \eta_t^{m,n}, \quad E_t \left[\eta_t^{m,n} \right] = 0, \tag{10}$$

where π_t^k is inflation rate between t and $t+k$, i_t^k is the current nominal interest rate on k-month Treasury bills, and $\eta_t^{m,n}$ is the prediction error. Rejection of the null hypothesis $H_0 : \beta_{m,n} = 0$ corresponds to predictability of inflation, while rejection of $H_0 : \beta_{m,n} = 1$ – to predictability of real interest rates. The combinations (m, n) are $(3, 1)$, $(6, 3)$, $(9, 6)$, $(12, 6)$, and $(12, 9)$, although Mishkin presents the results for all combinations of m and n not exceeding 6 in an appendix to his working paper version (Mishkin, 1988).

Since the data are overlapping, there is conditional serial correlation in $\eta_t^{m,n}$ of order $\max(m, n) - 1$. Mishkin utilizes OLS estimation with HAC robust calculation of standard errors to perform the tests. He also acknowledges poor performance of asymptotic approximation and simulates more relevant for the given sample sizes critical values. Needless to say, the OLS estimator is inefficient in the class of instrumental variables due to the presence of serial correlation and possibly conditional heteroskedasticity. We apply the ideas developed in the above sections to suggest an alternative way to estimate (10). What follows does not pretend to be a rigorous investigation of the applied problem at hand, but rather a complementary illustrative material.

The assumptions we make entail conditional normality, which is quite plausible for the problem under consideration. The estimator is less robust than the class of instrumental variables estimators, but if the normality does hold, the resulting tests are expected to have more power. Given a short time span of the data, we consider the shortest horizons that Mishkin (1990) uses: $m = 3$, $n = 1$. The framework outlined in the Appendix applies, with $J = 2$, $y_{t+3} = \pi_t^3 - \pi_t^1$, $x_t' = (1, i_t^3 - i_t^1)$, $\beta' = \left(\alpha_{3,1}, \beta_{3,1}\right)$, $g(x_t, \beta) = x_t'\beta$, $e_{t+3} = \eta_t^{3,1}$. The expectation adjustment processes $\delta_{1,t}$ and $\delta_{2,t}$ are specified as linear functions of the last realized prediction error:

$$\delta_{j,t} = \delta_j + \gamma_j e_t, \quad j = 1, 2,$$

with δ_j and γ_j satisfying the stability condition. We parametrize the conditional variance ω_t as a linear function of the squared last realized error, in the ARCH(1) spirit:

$$\omega_t = \omega + \alpha e_t^2,$$

where $\omega > 0$ and $\alpha \geq 0$. The reason for such parsimonious specifications lies in restricted sample sizes that the original author exploits.

The sample for the period from 1964.02 to 1986.12 (panel A in Table 1, 275 observations) is split into three subperiods: from 1964.02 to 1979.10 (panel B, 189 observations), from 1979.11 to 1982.10 (panel C, 36 observations) and from 1982.11 to 1986.12 (panel D, 50 observations). For reasons, see Mishkin (1990) and references therein. Mishkin (1992) finds that the series of nominal interest rates and inflation may contain unit roots. Indeed, conventional unit root tests fail to reject unit roots in the interest rate series. However, after taking differences between interest rates at different maturities, these unit root tests strongly reject the presence of unit roots in the resulting series.

Table 1 displays the results. With the more efficient estimation method, it is clearer that splitting the sample is well justified: the results heavily differ across the subperiods. In all but period C, numerical values for the structural parameters $\alpha_{3,1}$ and $\beta_{3,1}$ agree with Mishkin's (1990), but the inferences are different in that more efficient estimates happen to often be more significant, sometimes reverting Mishkin's (1990) conclusion of the absence of information about the path of future inflation in short term nominal interest rates.

The other tests yield the following results. Testing $\gamma_1 = \gamma_2 = 0$, i.e. that the Wold innovation of e_t is a martingale difference, sometimes yields rejection, thus invalidating the conventional specification for the error term. Testing $\alpha = 0$, i.e. that there is no conditional heteroskedasticity given time-varying expectation adjustment parameters, always yields acceptance. This goes at variance with the presence of ARCH effects in the error term of an ARMA representation of the inflation series documented in Mishkin (1990, 1992). The discrepancy may be due to the documented difficulty to disentangle the ARCH effect and bilinearity (Weiss, 1986).

4 Application 2: Disagreement and Uncertainty

As mentioned earlier, Rich, Raymond and Butler (1992) noted inappropriateness of conventional modelling of the variance when the data are overlapping. To deal with multiperiod prediction data from surveys properly, they configured a conditional moment restriction $E\left[e_{t+J+1}^2|\Im_t\right] = \omega_t$ and estimated this *Survey Data – ARCH model* by GMM. They also applied the idea to the estimation of the relationship between forecast dispersion and forecast uncertainty from data on inflation expectations from several surveys, the Livingston survey series among them. In 1996, William Bomberger estimated a similar relationship on the Livingston data with the use of the maximum likelihood method that violates the multiperiod prediction nature of the problem. Rich and Butler (1998) reacted with a critique which pointed out misspecification of the ML procedure and suggested again the GMM framework. They found that the structural specification of the variance equation is dominated by an ARCH-type one. In his reply, Bomberger (1999) compared the results of GMM and misspecified ML procedures and argued that despite misspecification, the results are reconcilable, especially if one takes into account the time span differences. In this section, we would like to reexamine the relationship within this paper's theoretical framework and possibly find solid arguments for the Rich-Raymond-Butler – Bomberger dispute.

The series under investigation are[6]: (1) the observed mean error $e_{t+2} = \pi_{t+2} - p_t$, where π_{t+2} is a two-step-ahead[7] inflation, $p_t = \frac{1}{n_t}\sum_{i=1}^{n_t} p_{it}$ is its average forecast made

[6]We change time indexing from the original to be consistent with our notation.

[7]The prediction horizon is one year, while the sampling interval is semi-annual.

at time t, p_{it} is the forecast of forecaster i made at time t, and n_t is a number of forecasters at time t; and (2) the measure of disagreement among forecasters $s_t^2 = \frac{1}{n_t}\sum_{i=1}^{n_t}(p_{it} - p_t)^2$. The semi-annual data from the Livingston survey on p_t, s_t^2, e_{t+2} and n_t from 1946:1 to 1994:2 are given in Table 1 of Bomberger (1996). The object of interest is uncertainty $\omega_t = E\left[e_{t+2}^2|\Im_t\right]$ as function of time t information $\Im_t = \sigma\left(s_t, e_t, s_{t-1}, e_{t-1}, \ldots\right)$.

Because of the aberrant behavior of all variables at the beginning of the sample (which may be caused by a possibly transitional nature of the survey at initial stages of its existence), it may not be innocuous to include the initial 7 observations into the sample. The ARCH literature (e.g., van Dijk, Franses and Lucas, 1999) suggests that the presence of outliers (to which the initial observations may be attributed) may lead to spurious ARCH effects. Bomberger (1999) reports high degree of sensitivity of GMM estimates to where the sample begins, and we find the same tendency with our ML estimates. Since in these circumstances estimation from the full sample can hardly rely on asymptotic theory, we report the results for the subsamples beginning at 1949:2 (to rule out the transitional period) and at 1952:1 (to use one of Bomberger's subsamples), although we do that for the full sample too.

Rich and Butler (1998) found that in the model[8]

$$\omega_t = \omega + \sigma s_t^2 + \alpha_4 e_{t-4}^2,$$

σ is insignificant when the term $\alpha_4 e_{t-4}^2$ is included in the specification, and conclude

[8]Rich, Raymond and Butler (1992) included a square of the realized error's *fourth* lag on the basis of the sample autocorrelation function of the squared regression residuals.

that disagreement s_t^2 is not a useful proxy for uncertainty. We would like to verify this claim within the framework of the present paper keeping in mind that the ML procedure may be more powerful than the GMM test. To this end, we adopt the distributional assumption (5) with the specification (7) and

$$\delta_t = \delta + \kappa s_t + \gamma_4 e_{t-4}.$$

That is, we let the adaptation parameter δ_t to linearly depend on the same variables that enter the variance equation.

Table 2 shows the results for several subsamples. The lines that represent GMM estimates are taken from earlier referenced sources. Firstly, we can see that if one takes the standard errors at face value, the proposed ML procedure is indeed more efficient and yields tighter confidence intervals. Secondly, in the function for the adjustment parameter, κ is never significant, while γ_4 tends to be significant, especially in the stationary region. That is, the Wold innovation of e_t is *not* a martingale difference sequence, and the prediction should indeed be adjusted on the basis of the past prediction error. Thirdly, in the skedastic function, σ is always significant, which is in favor of Bomberger's use of the disagreement measure s_t^2 as a proxy for uncertainty. However, this is *not* its perfect measure, and the fit does improve when one adds the ARCH term, in contrast to Bomberger (1999). That is, we confirm Rich and Butler's (1998) reversal of what Bomberger (1999) calls "the result B". In addition, the results are at variance with Bomberger's pure proportionality specification $\omega_t = \sigma s_t^2$, as the constant ω turns out to be highly significant.

5 Conclusion

When the disturbance term is a multiperiod optimal prediction error and one wants to model its conditional variance within the ARCH framework, it is conventional to impose a linear moving average structure on the error making the innovation a martingale difference sequence following ARCH. In this paper we argued that this is overly restrictive in several respects and suggested Gaussian maximum likelihood framework that does not entail this assumption. We applied this framework to two empirical problems. One is estimation of Mishkin's (1990) relationship between the term structure of nominal interest rates and future inflation at short horizons. The other is the dispute between Bomberger (1996, 1999) and Rich and Butler (1998) on whether disagreement is an appropriate measure for forecast uncertainty.

Appendix: Generalization for many-step-ahead prediction

For the $(J+1)$-step-ahead prediction problem the generalization is straightforward. The mean equation is

$$y_{t+J+1} = g\left(x_t, \beta\right) + e_{t+J+1},$$

where e_{t+1} is modeled as Gaussian when conditioned on $\mathcal{X}_{t-1} \vee \mathcal{E}_t$, with conditional mean μ_t and conditional variance ω_{t-J}, the latter being measurable relative to $\Im_{t-J}$ to yield normality of e_{t+1} conditional on each of $\Im_t, \Im_{t-1}, \cdots, \Im_{t-J}$. That is,

$$e_{t+1}|\mathcal{X}_{t-1}, \mathcal{E}_t \sim \mathcal{N}\left(\mu_t, \omega_{t-J}\right)$$

subject to $E\left[\mu_{t+J}|\Im_t\right] = 0$. Driven by conditional normality, the specificetion for μ_t

is

$$\mu_{t+J} = \delta_{1,t+J-1}\left(e_{t+J}-\mu_{t+J-1}\right)+\delta_{2,t+J-2}\left(e_{t+J-1}-\mu_{t+J-2}\right) + \cdots + \delta_{J,t}\left(e_{t+1}-\mu_{t}\right),$$

where $\Im_t$-measurable processes $\delta_{j,t}$, $j = 1, \cdots, J$ satisfy the following condition for the stability and stationarity of μ_t:

$$E\left[\max(\log|\delta_{j,t}|, 0)\right] < \infty \quad \forall\, j = 1, \cdots, J, \quad E\left[\log\|M_t\|\right] < 0,$$

where

$$\underset{J\times J}{M_t} = \begin{pmatrix} \delta_{1,t+J-1} & \delta_{2,t+J-2} & \cdots & \delta_{J,t} \\ & \mathrm{I}_{J-1} & & \mathbf{0}_{J-1} \end{pmatrix}$$

(see Bougerol and Picard, 1992), and $\|\cdot\|$ is any matrix norm[9].

Natural specifications for processes $\delta_{j,t}$ are (possibly nonlinear) functions of the last r_j realized errors:

$$\delta_{j,t} = \Delta_j(e_t, e_{t-1}, \cdots, e_{t-r_j+1}; \delta_j) \quad j = 1, \cdots, J,$$

where δ_j's are "mean parameters". The conditional variance ω_t may be parametrized as a function of its previous p values and the last q realized errors:

$$\omega_t = \Omega(\omega_{t-1}, \omega_{t-2}, \cdots, \omega_{t-p}, e_t, e_{t-1}, \cdots, e_{t-q+1}; \omega),$$

where ω contains "variance parameters".

[9]It is preferable to use the spectral norm $\|A\| = \sqrt{\varrho\left(A'A\right)}$, where $\varrho(\cdot)$ is the spectral radius, because it is smaller than, for example, the Euclidean norm $\|A\| = \sqrt{tr\left(A'A\right)}$.

Let $\delta \equiv (\delta_1' \ \cdots \ \delta_J')'$ and $\theta \equiv (\beta' \ \delta' \ \omega')'$. Apart from a constant, the loglikelihood for one observation is:

$$l_{t+1}(\theta) = -\frac{1}{2}\left(\log \omega_{t-J} + \frac{(e_{t+1} - \mu_t)^2}{\omega_{t-J}}\right)$$

Maximization of $\sum_t l_{t+1}(\theta)$ yields the maximum likelihood estimator $\hat{\theta}$, which under suitable conditions is consistent for θ. Differentiating $l_{t+1}(\theta)$ yields

$$\begin{aligned}
\frac{\partial l_{t+1}}{\partial \beta} &= \frac{1}{2}\left(\epsilon_{t+1}^2 - \frac{1}{\omega_{t-J}}\right)\frac{\partial \omega_{t-J}}{\partial \beta} + \epsilon_{t+1}\left(\frac{\partial g\left(x_{t-J}, \beta\right)}{\partial \beta} + \frac{\partial \mu_t}{\partial \beta}\right) \\
\frac{\partial l_{t+1}}{\partial \delta} &= \epsilon_{t+1}\frac{\partial \mu_t}{\partial \delta} \\
\frac{\partial l_{t+1}}{\partial \omega} &= \frac{1}{2}\left(\epsilon_{t+1}^2 - \frac{1}{\omega_{t-J}}\right)\frac{\partial \omega_{t-J}}{\partial \omega}
\end{aligned}$$

where $\epsilon_{t+1} = \omega_{t-J}^{-1}\left(e_{t+1} - \mu_t\right)$, and

$$\begin{aligned}
\frac{\partial \mu_{t+J}}{\partial \beta} = & \frac{\partial \delta_{1,t+J-1}}{\partial \beta}\left(e_{t+J} - \mu_{t+J-1}\right)\cdots + \frac{\partial \delta_{J,t}}{\partial \beta}\left(e_{t+1} - \mu_t\right) \\
& -\delta_{1,t+J-1}\left(\frac{\partial g\left(x_{t-1}, \beta\right)}{\partial \beta} + \frac{\partial \mu_{t+J-1}}{\partial \beta}\right)\cdots - \delta_{J,t}\left(\frac{\partial g\left(x_{t-J}, \beta\right)}{\partial \beta} + \frac{\partial \mu_t}{\partial \beta}\right),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mu_{t+J}}{\partial \delta} = & \frac{\partial \delta_{1,t+J-1}}{\partial \delta}\left(e_{t+J} - \mu_{t+J-1}\right)\cdots + \frac{\partial \delta_{J,t}}{\partial \delta}\left(e_{t+1} - \mu_t\right) \\
& -\delta_{1,t+J-1}\frac{\partial \mu_{t+J-1}}{\partial \delta}\cdots - \delta_{J,t}\frac{\partial \mu_t}{\partial \delta},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \delta_{j,t+J-j}}{\partial \beta} = & -\frac{\partial \Delta_j(e_{t+J-j}, \cdots, e_{t+J-j-r_j+1}; \delta_j)}{\partial e_{t+J-j}}\frac{\partial g\left(x_{t-j-1}, \beta\right)}{\partial \beta} \\
& \cdots - \frac{\partial \Delta_j(e_{t+J-j}, \cdots, e_{t+J-j-r_j+1}; \delta_j)}{\partial e_{t+J-j-r_j+1}}\frac{\partial g\left(x_{t-j-r_j}, \beta\right)}{\partial \beta},
\end{aligned}$$

$$\frac{\partial \delta_{j,t+J-j}}{\partial \delta} = \frac{\partial \Delta_j(e_{t+J-j}, \cdots, e_{t+J-j-r_j+1}; \delta_j)}{\partial \delta},$$

$$
\begin{aligned}
\frac{\partial \omega_{t-J}}{\partial \omega} \quad = \quad & \frac{\partial \Omega(\omega_{t-J-1}, \cdots, \omega_{t-J-p}, e_t, \cdots, e_{t-q+1}; \omega)}{\partial \omega_{t-J-1}} \frac{\partial \omega_{t-J-1}}{\partial \omega} \\
& \cdots + \frac{\partial \Omega(\omega_{t-J-1}, \cdots, \omega_{t-J-p}, e_t, \cdots, e_{t-q+1}; \omega)}{\partial \omega_{t-J-p}} \frac{\partial \omega_{t-J-p}}{\partial \omega} \\
& + \frac{\partial \Omega(\omega_{t-1}, \cdots, \omega_{t-p}, e_t, \cdots, e_{t-q+1}; \omega)}{\partial \omega}.
\end{aligned}
$$

Under correct specification and suitable conditions, a consistent estimate of the asymptotic variance matrix of $\hat{\theta}$ is

$$
\hat{V}_{\hat{\theta}} = T \left(\sum_t \frac{\partial l_{t+1}}{\partial \theta} \frac{\partial l_{t+1}}{\partial \theta'} \right)^{-1}.
$$

References

Baillie, R.T. and R.P. DeGennaro, "Stock returns and volatility," *Journal of Financial and Quantitative Analysis* 25 (1990), 203–214.

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen, "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics* 74 (1996), 3–30.

Bera, A.K. and M.L. Higgins, "ARCH models: properties, estimations and testing," *Journal of Economic Surveys* 7 (1993), 305–366.

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

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

Bollerslev, T., R.F. Engle and D.B. Nelson, "ARCH models," in R.F. Engle and D.L. McFadden, eds., *Handbook of econometrics*, Vol. 4 (Amsterdam: North-Holland, 1994), 2959–3038.

Bomberger, W.A., "Disagreement as a measure of uncertainty," *Journal of Money, Credit and Banking* 28 (1996), 381–392.

Bomberger, W.A., "Disagreement and uncertainty: a reply to Rich and Butler," *Journal of Money, Credit and Banking* 31 (1999), 273–276.

Bougerol, P. and N. Picard, "Strict stationarity of generalized autoregressive processes," *Annals of Probability* 20 (1992), 1714–1730.

Brandt, A., "The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients," *Advances in Applied Probability* 18 (1986), 211–220.

Eichenbaum, M.S., L.P. Hansen and K.J. Singleton, "A time series analysis of representative agent models of consumption and leisure choice under uncertainty," *Quarterly Journal of Economics* 103 (1988), 51–78.

Engle, R.F., "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation," *Econometrica* 50 (1982), 987–1008.

Engle, R.F. and T. Bollerslev, "Modelling the persistence of conditional variances," *Econometric Reviews* 5 (1986), 1–87.

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

Francq, C. and J.-P. Zakoïan, "Estimating weak GARCH representations," *Econometric Theory* 16 (2000), 692–728.

Hall, R.E., "Intertemporal substitution in consumption," *Journal of Political Economy* 96 (1988), 339–357.

Hansen, L.P. and K.J. Singleton, "Efficient estimation of linear asset pricing models with moving-average errors," *Journal of Business and Economic Statistics* 14 (1996), 53–68.

Hayashi, F. and C. Sims, "Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments," *Econometrica* 51 (1983), 783–798.

Higgins, M.L. and A.K. Bera, "A class of nonlinear ARCH models," *International Economic Review* 33 (1992), 137–158.

Hodrick, R.J., *The empirical evidence on the efficiency of forward and futures foreign exchange markets* (New York: Harwood Academic Publishers, 1987).

McCurdy, T.H. and I.G. Morgan, "Tests for a systematic risk component in deviations from uncovered interest rate parity," *Review of Economic Studies* 58 (1991), 587–602.

Mishkin, F., "What does the term structure tell us about future inflation?" NBER working paper No. 2626, 1988.

Mishkin, F., "What does the term structure tell us about future inflation?" *Journal of Monetary Economics* 25 (1990), 77–95.

Mishkin, F., "Is the Fisher effect for real? A reexamination of the relationship between inflation and interest rates," *Journal of Monetary Economics* 30 (1992), 195–215.

Nelson, D.B., "Conditional heteroskedasticity in asset returns: a new approach," *Econometrica* 59 (1991), 347–370.

Pagan, A.R. and Y.S. Hong, "Nonparametric estimation and the risk premium," in W.A. Barnett, J.E. Powell and G.E. Tauchen, eds., *Semiparametric and nonparametric methods in statistics and econometrics* (Cambridge: Cambridge University Press, 1991).

Pagan, A.R. and G.W. Schwert, "Alternative models for conditional stock volatility," *Journal of Econometrics* 45 (1990), 267–290.

Rich, R.W. and J.S. Butler, "Disagreement as a measure of uncertainty: a comment on Bomberger," *Journal of Money, Credit and Banking* 30 (1998), 411–419.

Rich, R.W., J.E. Raymond and J.S. Butler, "The relationship between forecast dispersion and forecast uncertainty: evidence from a Survey Data – ARCH model," *Journal of Applied Econometrics* 7 (1992), 131–148.

van Dijk, D., P.H. Franses and A. Lucas, "Testing for ARCH in the presence of additive outliers," *Journal of Applied Econometrics* 14 (1999), 539–562.

Wang, K.-L., C. Fawson, C.B. Barrett and J.B. McDonald, "A flexible parametric GARCH model with an application to exchange rates," *Journal of Applied Econometrics* 16 (2001), 521–536.

Weiss, A.A., "ARCH and bilinear time series models: comparison and combination," *Journal of Business and Economic Statistics* 4 (1986), 59–70.

Zakoïan, J.-P., "Threshold heteroskedastic models," *Journal Of Economic Dynamics and Control* 18 (1994), 931–955.

Table 1. The estimation results for the interest rates–inflation problem.

Method	$\alpha_{3,1}$	$\beta_{3,1}$	δ_1	γ_1	δ_2	γ_2	ω	α
Panel A: 1964.02 – 1986.12								
ML	$.157^{\dagger}$	$-.276^{\dagger}$	$.194^{\dagger}$	$.071^{\dagger}$	$.524^{\dagger}$	$-.039^{\diamond}$	$2.81^{\dagger}$	.031
	(.059)	(.105)	(.052)	(.021)	(.050)	(.021)	(.25)	(.067)
OLS	.169	$-.318$	–	–	–	–	–	–
	(.142)	(.260)						
Panel B: 1964.02 – 1979.10								
ML	.189	$-.318$	$.155^{\diamond}$	$.081^{*}$	$.299^{\dagger}$	$-.004$	$2.47^{\dagger}$	.002
	(.157)	(.418)	(.082)	(.031)	(.086)	(.028)	(.24)	(.046)
OLS	.142	$-.313$	–	–	–	–	–	–
	(.185)	(.450)						
Panel C: 1979.11 – 1982.10								
ML	$.562^{\dagger}$	$-.817^{\dagger}$	$-.655^{\dagger}$	$-.074$	$1.169^{\dagger}$	$.259^{\diamond}$	1.44^{*}	.002
	(.172)	(.067)	(.236)	(.132)	(.273)	(.140)	(.75)	(.098)
OLS	.009	$-.177$	–	–	–	–	–	–
	(.564)	(.373)						
Panel D: 1982.11 – 1986.12								
ML	.571	$-.958$	$-.133$	.023	$.736^{\dagger}$	$-.041$	$2.36^{\dagger}$	.026
	(.604)	(1.02)	(.178)	(.085)	(.133)	(.074)	(.83)	(.176)
OLS	.646	$-.954$	–	–	–	–	–	–
	(.568)	(.913)						

Notes: The econometric model is $\pi_t^3 - \pi_t^1 = \alpha_{3,1} + \beta_{3,1}(i_t^3 - i_t^1) + e_{t+3}$, where π_t^k is inflation rate from t to $t+k$, i_t^k is current nominal interest rate at t on k-month Treasury bills, $k = 1, 3$, and e_{t+3} is prediction error. The assumption on e_{t+3} is $e_{t+1}|\mathcal{X}_{t-1}, \mathcal{E}_t \sim \mathcal{N}(\mu_t, \omega_{t-2})$, where $\mu_{t+2} = \delta_{1,t+1}(e_{t+2} - \mu_{t+1}) + \delta_{2,t}(e_{t+1} - \mu_t)$, $\delta_{1,t} = \delta_1 + \gamma_1 e_t$, $\delta_{2,t} = \delta_2 + \gamma_2 e_t$, $\omega_t = \omega + \alpha e_t^2$. Asymptotic standard errors are in parentheses. Lines denoted as "OLS" are reproduced from Mishkin (1990). Tests of individual significance of α is one-tailed; $^{\diamond}$ denotes significance at 10% level, * – at 5% level, † – at 1% level.

Table 2. The estimation results for the disagreement–uncertainty problem.

Method	δ	κ	γ_4	ω	σ	α_4
Full sample 46:1–94:2						
ML	$-.474^*$	$-.020$	.022	$1.28^\dagger$	$.142^*$	$.097^\diamond$
	(.186)	(.650)	(.065)	(.35)	(.070)	(.068)
GMM	–	–	–	$2.88^\diamond$	.131	$.159^*$
				(1.73)	(1.09)	(.089)
Subsample 47:2–94:2						
GMM	–	–	–	.38	$1.97^\dagger$	$-.02$
				(.89)	(.77)	(.05)
Subsample 49:2–94:2						
ML	$-.551^\dagger$	.048	$-.089^*$	$.696^*$	$.340^\dagger$	$.178^*$
	(.171)	(.071)	(.042)	(.348)	(.144)	(.089)
Subsample 52:1–94:2						
ML	$-.597^\dagger$	.000	$-.104^\dagger$	$.525^*$	$.322^*$	$.297^\dagger$
	(.140)	(.056)	(.032)	(.255)	(.145)	(.121)

Notes: The econometric model is $e_{t+1}|\mathcal{X}_{t-1}, \mathcal{E}_t \sim \mathcal{N}(\mu_t, \omega_{t-1})$, where e_{t+2} is observed average prediction error, $\mu_t = \delta_{t-1}(e_t - \mu_{t-1})$, $\delta_t = \delta + \kappa s_t + \gamma_4 e_{t-4}$, $\omega_t = \omega + \sigma s_t^2 + \alpha_4 e_{t-4}^2$, and s_t^2 is disagreement. Asymptotic standard errors are in parentheses. The first GMM line is reproduced from Rich and Butler (1998), the second – from Bomberger (1999). Tests of individual significance of σ and α_4 are one-tailed; $^\diamond$ denotes significance at 10% level, * – at 5% level, † – at 1% level.