Journal of Econometrics 40 (1989) 87–96. North-Holland

INTERVAL FORECASTING
An Analysis Based Upon ARCH-quantile Estimators

C.W.J. GRANGER, Halbert WHITE and Mark KAMSTRA*

University of California San Diego, La Jolla, CA 92093, USA

In this paper we explore techniques for obtaining interval forecasts based on estimated time-series models for processes which may exhibit autoregressive conditional heteroskedasticity (ARCH). To deal with the available variety of possible interval forecasts, we propose a method for combining these forecasts based on quantile regression techniques. Our approach is practical rather than theoretical, with attention focused directly on obtaining interval forecasts for two U.S. time series: a measure of unemployment and a Treasury bill rate. We evaluate the performance of our procedures using a variety of diagnostics. We find interval estimates which perform reasonably well, judged by both in-sample and out-of-sample criteria. Our experience suggests that a certain amount of care is required in order to obtain useful forecasts.

1. Introduction

All major economies produce many forecasts of the most important economic variables, such as GNP growth, prices, and unemployment rates; however, in virtually all cases only point forecasts are provided. When forecast errors are normally distributed with constant variance, confidence intervals, at any desired level, can be easily provided. However, assumptions of constant variances or normality are often at odds with the evidence of the data. In this paper we take a general, empirical approach to the question of estimating specified quantiles, possibly time-varying, to indicate forecast uncertainty. For concreteness and because of their ready interpretability, we focus on the 25% and 75% quantiles. We also discuss results for the 10% and 90% quantiles.

Our strategy is to form a univariate model to forecast conditional means or medians, and then to consider several different ways to estimate time-varying quantiles from this model's residuals. When several different quantile estimates are available, these are combined using methods of quantile regression. We evaluate the resulting time-varying quantiles using simple time-series methods.

*The authors are grateful for the comments and suggestions of Richard Carson, Robert Engle, Peter Schmidt, Ross Starr, and the referees. All errors are the authors' responsibility. Granger's participation was supported by NSF Grant SES 85-13858, White's by NSF Grant SES 85-10637, and Kamstra's by NSF Grant SES 85-10637 and SSHRCC Fellowship Grant 453-86-0717.

0304-4076/89/$3.50

The techniques are applied to two U.S. series, an unemployment rate and a Treasury bill rate. To allow for possible non-linearity in the behavior of these series, both levels and logarithms of the series are modelled. The ultimate aim, however, is taken to be finding good approximations to the quantiles of the levels. To this end, we examine the behavior of four time series, both 'in-sample' and later 'out-of-sample' for evaluation purposes. The series are: (a) the U.S. unemployment rate, monthly, February 1948 through February 1986, denoted *UE*; (b) the logarithm of *UE*, denoted *LUE*; (c) the yield to maturity for 12-month U.S. Treasury bills aged 11 months, monthly, September 1964 through December 1985, denoted *TB*; (d) the logarithm of *TB*, denoted *LTB*.

2. Modeling the conditional mean and variance processes

Consider a univariate time series y_t and let $f_t(y)$ be the conditional probability density function of y_t conditional on the information set I_{t-1}: y_{t-j}, $j \geq 1$. The conditional mean is $m_t \equiv \mathrm{E}[y_t | I_{t-1}]$ with deviation $e_t = y_t - m_t$. The conditional variance of y_t is $V_t \equiv \mathrm{E}[e_t^2 | I_{t-1}]$. We also define $\sigma_t \equiv V_t^{1/2}$. Modeling m_t using Box–Jenkins techniques is common, but only recently has attention been paid to modeling V_t. A class of models for V_t has been proposed by Engle (1982) called ARCH models, with specification

$$v_t(\alpha) = \sum_{j=1}^{q} \alpha_j e_{t-j}^2 + \alpha_0, \qquad q \in N. \tag{2.1}$$

Note that $v_t(\alpha)$ is positive if all coefficients α_j are positive. ARCH models are usually estimated by maximum likelihood, assuming that e_t/σ_t is i.i.d. standard normal or standard t. As we avoid such specific assumptions, the following initial analysis is appropriate.

(i) Univariate Box–Jenkins procedures using heteroskedasticity-consistent standard errors (a protection against ARCH) give us useful specifications for m_t. For simplicity, we consider only ARMA$(p,0)$ specifications in $y_t = x_t - x_{t-1}$, where x_t is the level or logarithm of *UE* or *TB*. Thus, we obtain estimated autoregressive models for the difference in levels and the difference in logarithms for our series.

(ii) The squared estimated residuals $\hat{e}_t^2$ were then put into the (heteroskedasticity-consistent) Box–Jenkins identification process. Appropriate models explaining e_t^2 in terms of e_{t-j}^2, $j \geq 1$, and possible other explanatory variables were considered. This leads to an estimated model for V_t for the differences in levels and logarithms for our series.

This modeling process yields AR(12) models for the change in unemployment and also for the change in log unemployment, with significant terms at lags 1,2,10,12 (and lag 3 for logs). For changes in levels, $R^2 = 0.21$; for changes in logarithms, $R^2 = 0.17$. Autoregressive models for the squared residuals found (one-sided) significant lagged terms only for changes in levels (lag 3) and, using lags up to 11, gave R^2 values less than 0.05. Treasury bill rates were found to be effectively random walks. The residuals squared showed some structure, particularly for changes in levels. An AR model, using lags 1,4,12, gave an R^2 of 0.18. However, for logarithms of rates we found an R^2 of only 0.07. All of the estimated models were overfit, in that all lags having t-statistics greater than one in absolute value were retained. Details of the models fitted are available from the authors.

These estimated models give us estimates of m_t, σ_t, and e_t. If e_t is conditionally normal, the search for (time-varying) quantiles is easily concluded. For 25% and 75% quantiles these are $m_t \pm 0.6745\sigma_t$ (for 10% and 90%, $m_t \pm 1.282\sigma_t$). However, although these provide candidate estimates for the required quantiles, they are based on strong assumptions. As we do not wish to rely too heavily on such assumptions, we now consider some appropriate alternatives.

3. Simple quantile estimators

The statistics of this section make direct use of our estimates of m_t, e_t, and σ_t. With normal errors in the change in levels equations, the 75% quantile is

$$Z_{2t}(0.75) \equiv m_t + 0.6745\sigma_t,$$

where m_t and σ_t are now estimates from steps (i) and (ii) above for the change in levels. With normal errors in the change in logarithms equations, the 75% quantile for the change in levels is

$$Z_{3t}(0.75) \equiv x_{t-1}\left[\exp(\bar{m}_t + 0.6745\bar{\sigma}_t) - 1\right],$$

where $\bar{m}_t$ and $\bar{\sigma}_t$ are estimates from steps (i) and (ii) for the change in logarithms, and x_{t-1} is the relevant lagged level.

These estimates rely heavily on normality. Analogous estimates which avoid this are

$$Z_{4t}(0.75) = m_t + \hat{Q}_{0.75}\sigma_t,$$

$$Z_{5t}(0.75) = x_{t-1}\left[\exp(\bar{m}_t + \bar{Q}_{0.75}\bar{\sigma}_t) - 1\right],$$

where $\hat{Q}_{0.75}$ is the empirical 75% quantile (the $[3n/4]$-order statistic) of the standardized residuals e_t/σ_t and $\bar{Q}_{0.75}$ is the empirical 75% quantile of the standardized residuals $\bar{e}_t/\bar{\sigma}_t$, where $\bar{e}_t$ is obtained from the change in loga-

Table 1

Summary statistics for in-sample and out-of-sample performance of raw interval estimates/forecasts.[a]

50% intervals

	Number in sample			Number out of sample			Low/In/High[b]		Autocorrelation[c]		Average interval width	
	Low	In	High	Low	In	High	In sample	Out of sample	In sample	Out of sample	In sample	Out of sample
							Unemployment					
Z_2	69	164	67	33	73	27	2.64	1.67	4.96	7.53	0.27	0.27
Z_3	72	156	72	24	91	18	0.48	17.4	5.97	2.29	0.27	0.38
Z_4	75	150	75	34	70	29	0.00	0.62	0.85	6.09	0.24	0.25
Z_5	75	150	75	24	90	19	0.00	14.25	3.91	2.92	0.27	0.37
Z_6	75	150	75	34	70	29	0.00	0.74	12.09	8.89	0.20	0.20
	Dependent variable:				In-sample mean = 0.002 Out-of sample mean = 0.0038				RMSE = 0.224 RMSE = 0.222			
							Treasury bill rate					
											$\times 10^{-3}$	$\times 10^{-3}$
Z_2	45	91	35	18	32	10	4.60	0.17	1.93	1.41	0.685	1.19
Z_3	35	96	40	17	33	10	2.30	2.30	3.10	3.31	0.643	1.13
Z_4	45	81	45	18	29	13	0.00	0.27	0.96	2.40	0.585	1.02
Z_5	42	86	43	18	32	10	0.00	0.17	6.07	2.11	0.606	1.06
Z_6	42	86	43	19	26	15	0.00	1.6	1.73	3.92	0.55	0.55
	Dependent variable:			In-sample mean = 0.496×10^{-4} Out-of-sample mean = 0.337×10^{-4}					RMSE = 0.708×10^{-3} RMSE = 0.142×10^{-2}			

rithms. As a benchmark, we also consider the time-invariant quantile

$$Z_{6t}(0.75) = \tilde{Q}_{0.75},$$

where $\tilde{Q}_{0.75}$ is the empirical 75% quantile (the $[3n/4]$-order statistic) for the differences in levels, $y_t = x_t - x_{t-1}$, in sample. Similar statistics are used to estimate 25%, 10% and 90% quantiles. E.g., in $Z_{2t}(0.25)$ and $Z_{3t}(0.25)$, -0.6745 replaces 0.6745, and $Q_{0.25}$ replaces $Q_{0.75}$ in $Z_{4t}(0.25)$ and $Z_{5t}(0.25)$. These quantile estimates and the derived interval estimates will be called 'raw' estimates.

Table 1 gives the summary statistics which reflect the performance of the raw interval estimates generated by the raw quantile estimators. These statistics measure: (1) 'low/in/high': the extent of the bias, i.e., the departure of the interval estimates from the theoretical performance level of, e.g., 25% of observed values lying below the estimated 0.25 quantile, 50% lying within the estimated 0.25 and 0.75 quantiles, and 25% lying above the estimated 0.75

Table 1 (continued)

Summary statistics for in-sample and out-of-sample performance of raw interval estimators/forecasts.[a]

80% intervals

	Number in sample			Number out of sample			Low/In/High[b]		Autocorrelation[c]		Average interval width	
	Low	In	High	Low	In	High	In sample	Out of sample	In sample	Out of sample	In sample	Out of sample
							Unemployment					
Z_2	31	242	26	11	110	12	0.58	0.32	2.41	3.58	0.51	0.52
Z_3	27	242	30	5	121	7	0.31	8.17	1.95	4.42	0.52	0.74
Z_4	30	239	30	11	109	13	0.00	0.15	2.64	2.74	0.50	0.52
Z_5	30	239	30	5	122	6	0.00	5.26	2.64	3.46	0.51	0.72
Z_6	30	239	30	16	102	15	0.00	0.95	24.04	4.54	0.60	0.60
							Treasury bill rate					
											$\times 10^{-3}$	$\times 10^{-3}$
Z_2	12	139	20	7	49	4	2.05	0.85	1.87	2.85	1.30	2.26
Z_3	15	140	16	7	49	4	0.40	0.85	3.33	7.65	1.23	2.15
Z_4	17	137	17	9	49	2	0.00	4.19	4.18	0.96	1.31	2.27
Z_5	17	137	17	8	47	5	0.00	0.85	2.03	10.44	1.19	2.08
Z_6	18	136	17	12	44	4	0.05	1.17	10.52	8.86	1.22	1.22

[a]Let $d_t = 1$ if $x_t > Z_t(0.75)$, $d_t = 0$ if $Z_t(0.25) \le x_t \le Z_t(0.75)$, $d_t = -1$ if $x_t < Z_t(0.25)$, $t = 1, \dots, n$.

[b]Low/In/High $= (D_1 - n/4)^2/(n/4) + (D_0 - n/2)^2/(n/2) + (D_{-1} - n/4)^2/(n/4)$, where $D_i = \#[t\colon\ d_t = i]$ for 50% intervals, with $n/10$ and $8n/10$ replacing $n/4$ and $n/2$ for 80% intervals. With no bias, it is reasonable to expect that this statistic is distributed asymptotically as χ^2_2 for the out-of-sample results. The distribution for the within-sample results remains to be verified.

[c]Autocorrelation $= \sum_{i=-1}^{1} \sum_{j=-1}^{1} (D_{ij} - D_i D_j/n)^2/(D_i D_j/n)$, where $D_{ij} = \#[t\colon\ d_t = i,\ d_{t-1} = j]$. With no autocorrelation, it is reasonable to expect that this statistic is distributed asymptotically as χ^2_4 for the out-of-sample results. The distribution for the in-sample results remains to be verified.

quantile; (2) 'autocorrelation': the degree of serial correlation in a variable measuring whether a given observation falls below, inside or above the central interval estimate, and (3) the average width of the interval estimate. These quantities are computed both within and outside the estimation sample. The estimation sample ends in January 1975 for *UE*, giving 300 observations. For *TB* it ends in December 1979, giving 171 observations. Out-of-sample observations number 133 for *UE* and 60 for *TB*. A 'good' interval estimate should have little evidence of bias or autocorrelation and should give a narrow interval on average, both in and out of sample.

For unemployment, all raw interval estimates perform well in sample according to the low/in/high and autocorrelation measures. Out of sample, however, only Z_2 and Z_4 yield adequate interval estimates according to the low/in/high and autocorrelation measures. The other two appear to give significantly biased intervals. Z_4 yields the smallest average time-varying intervals. Interestingly, the time-invariant interval is narrowest for the 50% interval.

For Treasury bills things are different. Here, all new statistics perform well both in and out of sample with respect to bias and autocorrelation (except Z_5 for the 80% interval). Unlike unemployment, the out-of-sample average interval widths are sometimes almost twice those in sample. Perhaps this reflects greater uncertainty in interest rate levels for (post-sample) years in which the Federal Reserve Bank changed its control rules. Of all the raw statistics, Z_4 again gives the best overall performance, both in and out of sample for 50% intervals, while Z_3 performs well overall for 80% intervals.

It is interesting that both series exhibit sufficient stability to allow relatively good out-of-sample performance for at least some raw statistics, even though they use parameter estimates from (i) and (ii) that are 5.5 (UE) and 2.5 (TB) years old on average.

4. Combining quantile estimators

When more than one point forecast is available, it may be possible to obtain a superior forecast by combining the individual forecasts [Bates and Granger (1969), Engle, Granger and Kraft (1985)]. We now propose a method for combining raw quantile estimators to obtain potentially superior interval forecasts.

We first introduce an additional quantile estimator appropriate when there are no ARCH effects. In this case, the conditional quantile at level p is $Q_{p,t} = a_p + Q_{0.5,t}$, where $Q_{0.5,t}$ is the conditional median given I_{t-1}, and a_p is a constant. Thus, we set

$$Z_{1t} \equiv \hat{Q}_{0.5,t},$$

where $\hat{Q}_{0.5,t}$ is the predicted value from a least absolute deviations regression on the difference in levels using the same lags as in the model identified in step (i). We also use $Z_{0t} \equiv 1$, a constant. Including Z_{0t} allows adjustment for biases in the raw quantile estimators.

The method proposed for combining the raw statistics is to conduct a quantile regression [similar to that of Koenker and Bassett (1978)] using appropriate subsets of the Z_t's as explanatory variables. We do not include all the Z_t's because of their high intercorrelations.

Quantile regression coefficients $\hat{\gamma}$ are obtained by solving

$$\min_{\gamma} n^{-1} \sum_{t=1}^{n} |y_t - z_t\gamma| \left(p\, 1\left[y_t \geq x_t\gamma \right] + (1-p)\, 1\left[y_t < z_t\gamma \right] \right),$$

where $y_t = z_t - x_{t-1}$ (the change in levels), z_t is the vector of explanatory variables, and p is the desired quantile. Here, $p = 0.10, 0.25, 0.75, 0.90$. The criterion used is equivalent to a cost function of the form $C(e) = a|e|$, $e \geq 0$, $C(e) = b|e|$, $e < 0$, so if $a \neq b$, the cost function is a 'trick function'. The values of a and b depend on the p-value of the quantile. If a median is to be estimated, then $p = 0.5$ and $a = b$, giving the usual 'mean absolute deviation' cost function.

The motivation for this proposal is that the estimated coefficients $\hat{\gamma}$ are generally consistent for parameters γ^* giving an approximation $z_t\gamma^*$ to $Q_{p,t}$ which is optimal in the sense of minimizing the expected value of the cost function. The combining procedure can thus be viewed as quantile estimation of a misspecified model for the conditional quantile. [Because the explanatory variables (the raw statistics) contain parameters estimated in a first stage, the procedure in fact amounts to two-stage quantile estimation of a misspecified model.] Because we are interested in forecasting, the misspecification only causes an approximation error and is thus not fatal. Indeed, the purpose of the combining is to reduce the approximation error.

Until recently, quantile estimation has been hampered by the absence of sufficiently precise estimation algorithms. Even if $\hat{\gamma}$ is consistent, rounding errors from pivoting operations in linear programming solutions to our minimization problem can easily lead to useless computed estimates $\hat{\gamma}$. Fortunately, there is now available software which can produce good computed estimates $\hat{\gamma}$ [Fulton, Subramanian and Carson (1985)].

Our procedure is first applied to quantile models of the form $\alpha + \beta Z_j$, for each $j = 1, \ldots, 5$. If the assumptions underlying the construction of Z_j are correct, one should expect $\alpha = 0$ and $\beta = 1$. For the unemployment series these expectations are largely met, but the resulting 'adjusted' estimates and forecasts have properties which differ in interesting ways from those of the raw statistics.

Most noticeably, all of the estimated in-sample intervals for unemployment appear seriously biased (results available on request). We believe that this is due to a few influential observations. However, two of the adjusted intervals perform well out of sample over a period of eleven years. Now, however, Z_3 and Z_5 appear to perform best for the 50% and 80% intervals, yielding unbiased intervals, not too much autocorrelation, and average interval widths smaller than the best widths for the raw statistics.

For the Treasury bill series, the results are somewhat different, as the estimated slope coefficients are rather farther from their anticipated values. As with the raw satistics, the Treasury bill intervals appear to be relatively unbiased both in and out of sample, with little evidence of autocorrelation. Interestingly, the smallest out-of-sample interval width for the 50% interval is given by Z_1, with out-of-sample performance comparable to that in sample (average width = 0.00054). Of the other raw statistics Z_4 appears to perform

Table 2

Summary statistics for in-sample and out-of-sample performance of combined interval estimates/forecasts.

		Coefficient (standared error)					Number in sample			Number out of sample			Low/In/High		Autocorrelation		Average interval width	
Quantile	Intercept	Z_1	Z_3	Z_4	Z_5	Objective function	Low	In	High	Low	In	High	In sample	Out of sample	In sample	Out of sample	In sample	Out of sample
						Unemployment												
0.25	0.047			0.87	0.07	0.06126												
	(0.024)			(0.35)	(0.35)													
0.75	−0.058			0.44	0.47	0.06555	107	85	109	39	55	39	58.1	3.98	2.40	8.18	0.136	0.208
	(0.027)			(0.46)	(0.41)													
0.10	0.11	0.14	0.90			0.040												
	(0.052)	(0.29)	(0.20)															
0.90	−0.07	0.27	0.68			0.047	83	135	81	20	93	20	181.6	8.44	1.4	6.93	0.241	0.404
	(0.067)	(0.40)	(0.23)															
						Treasury bill rate												
0.25	0.95×10^{-4}	−0.80	1.11			0.148×10^{-3}												
	(0.14×10^{-3})	(0.90)	(0.33)															
0.75	-0.94×10^{-4}	−1.61	1.67			0.168×10^{-3}	46	81	44	14	36	10	0.52	2.93	0.73	3.00	0.6×10^{-3}	1.18×10^{-5}
	(0.16×10^{-3})	(1.55)	(0.65)															
0.10	-0.159×10^{-3}	0.81	0.72			0.833×10^{-4}												
	(0.18×10^{-3})	(1.0)	(0.25)															
0.90	0.108×10^{-3}	0.98	0.72			0.102×10^{-2}	19	135	17	10	45	5	0.234	3.52	2.85	6.69	1.22×10^{-3}	1.91×10^{-3}
	(0.23×10^{-3})	(1.34)	(0.29)															

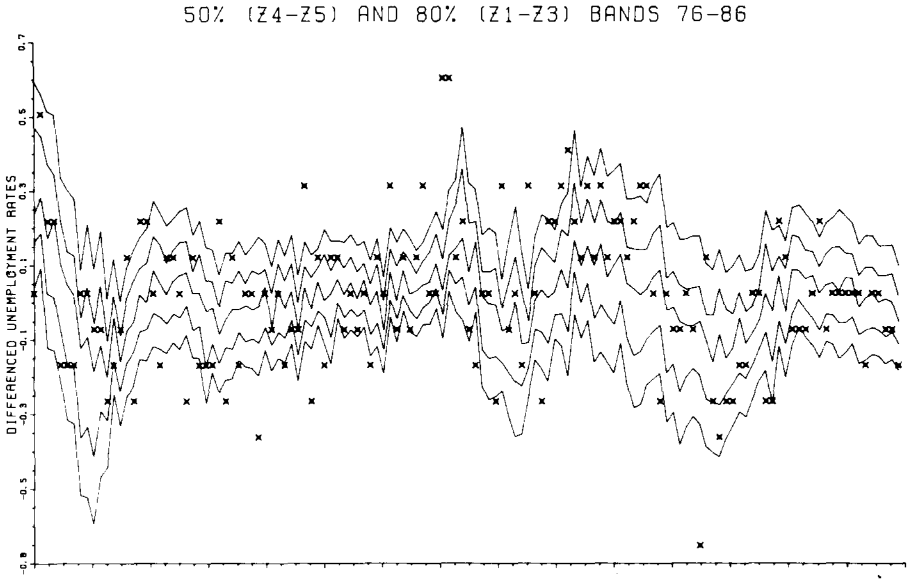

Fig. 1

best for the 50% interval. The results are different for the 80% interval. Again Z_1 gives the narrowest interval, but shows significant autocorrelation. Z_2 has no autocorrelation and a somewhat larger average interval.

Next, we consider the effects of combining several different quantile estimators. Because of the high collinearity between the raw statistics, we consider only cases in which two raw statistics are combined with a constant.

Table 2 presents some of the better results for this exercise. For 50% intervals for unemployment, we see apparent bias in sample, but little evidence of bias out of sample. The autocorrelation measure is lowest for the combination of Z_4 and Z_5, yielding 10% smaller intervals on average than any of the unbiased raw intervals. Thus there appears to be some gain from combining the raw measures. Similar results obtain for 80% intervals; here the combination of Z_1 and Z_3 performs best. Fig. 1 plots the resulting out-of-sample interval estimates for 50% and 80% intervals centered around the forecast value of the conditional median.

For Treasury bill rates, combining does not yield improvements over using an adjusted version of Z_1. We attribute this to finite-sample variation introduced by the estimation of an additional parameter. It is also possible that the relatively good out-of-sample performance of Z_1 is simply a fluke. In either event, this reveals the need for care in combining, and that more needs to be known about the procedures proposed here.

5. Summary and concluding remarks

In this paper we have started the study of methods for constructing quantile estimates for time series which provide potentially useful interval forecasts. Our procedure starts by modelling the conditional means and variances of the series and considering various 'raw' quantile estimates. Combinations are then investigated using quantile regression estimation. The resulting estimates are evaluated both in and out of sample. Our examples suggest that well-behaved interval forecasts can be constructed and that there may be gains from adjusting for bias and combining.

We emphasize that, because the methods that perform best in sample do not necessarily do best out of sample, further diagnostics need to be considered. Cross-validation methods have proven useful for cross-section data [Stone (1974), Efron (1983)] but are so far less developed for time series [but see Freedman (1986) and Carson (1987)]. At present, a sensible procedure is to hold back a number of observations from the estimation sample for out-of-sample evaluation and to use these results to choose the best procedure.

References

Bates, J.M. and C.W.J. Granger, 1969, The combination of forecasts, Operations Research Quarterly 20, 451–468.

Carson, R., 1987, On the use of cross validation and bootstrapping to determine the likely predictive power of macroeconomic models, Department of Economics discussion paper no. 87-7 (University of California at San Diego, CA).

Efron, B., 1983, Estimating the error rate of a prediction rule: Improvement on cross validation, Journal of the American Statistical Association 78, 316–331.

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

Engle, R.F., C.W.J. Granger, and D. Kraft, 1984, Combining competing forecasts of inflation using a bivariate ARCH model, Journal of Economic Dynamics and Control 8, 151–165.

Freedman, D.A., 1984, On bootstrapping two-stage least-squares estimates in stationary linear models, Annals of Statistics 12, 827–842.

Fulton, M., S. Subramanian, and R. Carson, 1987, Estimating fast regression quantiles using a modification of the Barrodale and Roberts L_1 algorithm, Department of Economics discussion paper no. 87-8 (University of California at San Diego, CA).

Koenker, R. and G. Bassett, 1978, Regression quantiles, Econometrica 46, 33–55.

Stone, M., 1974, Cross-validatory choice and assessment of statistical prediction, Journal of the Royal Statistical Association B 36, 111–147.

White, H., 1980, A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48, 817–838.