Forecasting With Structural Change

A. Thavaneswaran and M. Ghahramani

Department of Statistics

The University of Manitoba

Winnipeg, Manitoba, Canada R3T − 2N2

Abstract

Many analyses of time series forecasting have been based on the assumptions of a constant, time-invariant, data generating process, that is stationary (no structural change), and coincides with the time series model used. A simple example is when the process has undergone a regime shift (structural change), forecasts based on past information need not be unbiased despite being the previous conditional expectation. In this paper we study improved estimates and forecasts for time series models with structural change in the mean as well as volatility.

Keywords: Structural Change, GARCH, Improved Estimate/Forecast.

1 Introduction

In time series applications, minimum mean square error (MMSE) forecasts are generally presented as point estimates. MMSE forecasts are linear only in the Gaussian case. However, in most decision making problems where the underlying distribution is asymmetric and or the constraints are nonlinear, linear forecasts will not be sufficient and nonlinear forecasts will be be needed.

In order to generate non-linear forecasts for nonlinear models, a result of Godambe (1998) on optimal combination of orthogonal estimating functions for discrete time stochastic processes had been extended to non-orthogonal estimating functions and applied to nonnormal forecasting/filtering problems, in which the posterior mode of the signal distribution is efficient (see Thavaneswaran and Thompson(2002), Thavaneswaran and Heyde(1999)). The extensions so obtained may be applicable in a wider context than the standard notions based upon the conditional mean. For example, more efficient least absolute deviation median forecast/filtering equations for models having heavy tailed distributions have been studied by the aforementioned authors.

This paper addresses some interesting problems facing the theory of time series forecasting based on empirical time series models, especially with structural change, (see Clements and Hendry(1998)). The methods developed for forecasting are based on the conditional mean and variance are not adequate to accommodate the structural change in the mean or in the variance. Properties of forecasting models are often derived assuming that the model is *correctly specified* , in that it coincides with the mechanism which generated the observed data. At the forecasting step for linear models, we assume that the model parameters are known and the process is stationary, whereas for nonlinear models, we assume that the increments are martingale differences. A simple example is when the process has undergone a regime shift (structural change). Forecasts based on past information need not be unbiased despite being the previous conditional expectation (see Clements and Hendry (1998)).

Consider a simple example of an AR(1) model with structural change at time $T+1$, $y_t = \alpha + \rho y_{t-1} + a_t$. Then, $y_{T+1} = \alpha^* + \rho y_T + a_{T+1}$ and the forecast error is given by $e_{T+1} = y_{T+1} - \hat{y}_{T+1} = (\alpha^* - \alpha) + a_{T+1}$. Hence the associated mean squared error is $MSE_1 = (\alpha^* - \alpha)^2 + \sigma^2 \geq \sigma^2$. Since the forecast is biased, the usual forecasting method does not apply when there is structural change.

For ARCH or even for more general GARCH models the main assumption is that the differences form martingale differences. When the process undergoes a regime change or shift in the mean, we cannot use the usual MMSE forecast. In this paper, we propose an optimal, but biased estimate/forecast for the deterministic shift in the mean (or variance) of linear time series models and some nonlinear time series models such as GARCH models. This paper is organized as follows. In section 2, some basic results on best linear MMSE estimators for any stationary process with changes in the mean as well as in the volatility are given. In section 3, a regression model where there has been a single structural change is outlined and a method for obtaining improved estimates is proposed. Finally, we end the paper with a Conclusions section.

2 Improved Estimation/Forecasting

Recently Shalabh (2001) has studied an improved biased estimate of the mean of a normal population. In this section we use a similar approach to improve the estimator. Suppose that we have a set of independent observations $\{y_t\}, t = 1, 2, 3...T$ with the underlying model

$$y_t = \mu + a_t,$$

where a_t are i.i.d. normal with mean 0 and variance σ^2. Then a forecast of y_{T+1} is μ. Since the best linear unbiased estimate of μ, based on the the available observations is $\bar{y}$, we take

$$\hat{y}_{T+1} = \bar{y},$$

and the associated mean square error (MSE) of the forecast, regarding y_{T+1} as a variable is

$$E(y_{T+1} - \bar{y})^2 = \sigma^2(1 + \frac{1}{T}).$$

This is the Minimum MSE (MMSE) if we restrict the estimator of μ to be unbiased. We can relax this restriction and look for a biased estimator which has smaller MSE. Suppose such an estimator is $c\bar{y}$. The c that minimizes the MSE $E[y_{T+1} - c\bar{y}]^2$ with respect to c is given by $c = \frac{\mu^2}{E\bar{y}^2}$. Then the Minimum Forecast Square Error (MFSE) is given by

$$E(y_{T+1} - c\bar{y})^2 = \sigma^2(1 + \frac{1}{T}[1 + \frac{v^2}{T}]^{-1}) \leq \sigma^2(1 + \frac{1}{T}),$$

where $v = \frac{\sigma}{\mu}$ is the the coefficient of variation. If v is small then the improvement on the MSE is relatively small. However, if v is large then the improvement can be significant. In practice, we always assume that the model parameters are known at the forecasting step. In order to use $c\bar{y}$ as a forecast we have to replace v by $\hat{v}$, the sample coefficient of variation and hence the new forecast is given by

$$c\bar{y} = (1 + \frac{\hat{v}^2}{T})^{-1}\bar{y}.$$

That is the MFSE and bias decrease as the sample size T increases. That is, for large values of T, the estimate approaches the sample mean as well. Suppose there is some structural change(or a possible intervention) at time $T + 1$ and that the mean has shifted the mean from μ to μ_1 . The MMSE estimate of y_{T+1} based on $\{y_t\}, t = 1, 2, \ldots, T$ is obtained by minimizing $E[y_{T+1} - c\bar{y}]^2$ with respect to c and the optimal $c = \frac{\mu\mu_1}{E\bar{y}^2}$. The $MMSE = \sigma^2 + \mu_1^2[1 - \frac{\mu^2}{E\bar{y}^2}]$ and this can be simplified to $MMSE = \sigma^2 + \mu_1^2[\frac{\sigma^2}{\sigma^2 + T\mu^2}]$.

It is of interest to note that for the proposed new estimator the MMSE and bias decrease as T increases. In practice, if we have m observations after the structural change we could forecast y_{T+m+1} by replacing μ_1 by the average based on these m observations. A more refined approach based on a recursive algorithm is given in the next section. However, if we use the usual forecast, the sample mean $\bar{y}$ to forecast the future value at time $T+1$, the bias $\mu - \mu_1$, remains constant.

Now we consider the regression model of the form $y_{t+1} = \beta x_t + a_{t+1}$ where a_{t+1} are i.i.d. mean 0 and variance σ^2. Suppose there is a shift in the mean at time T and our interest lies in forecasting y_{T+1} given the observations $(x_t, y_t), t = 1, 2, 3, \ldots, T$. The usual estimator/forecast is $\hat{y}_{T+1} = \hat{\beta}_T x_T$, where $\hat{\beta}_T = \left(\sum_{t=2}^{T} x_{t-1}^2\right)^{-1} (\sum_{t=2}^{T} x_{t-1} y_t)$, and the conditional Mean square forecast error (MSFE) becomes $E[\hat{e}_{T+1}^2 | x_T] = \sigma^2 + Var(\hat{\beta}_T) x_T^2$. The corresponding improved MMSE forecast/predictor is $\tilde{y}_{T+1} = c\hat{\beta}_T x_T$ where $c = \frac{\beta^2}{E\hat{\beta}_T^2}$. The corresponding conditional MSFE becomes $E[e_{T+1}^2 | x_T] = \sigma^2 + \frac{\beta^2 Var(\hat{\beta}_T) x_T^2}{Var(\hat{\beta}_T) + \beta^2}$ where $Var(\hat{\beta}_T) = \left(\sum_{t=2}^{T} x_{t-1}^2\right)^{-1} \sigma^2$. This clearly shows that biased optimal predictors have smaller forecast error.

2.1 Time Series Models

Now consider a set of observations $\{y_t\}, t = 1, 2, 3 \ldots T$ from a stationary process $\{y_t\}$ having mean μ . Let $\{\rho_k\}$ be the autocorrelation function and let the $\{\gamma_k\}$ be the corresponding autocovariance function. Then it can be easily shown that the MMSE forecast of y_{T+1} is $c\bar{y}$ where $c = \frac{\mu^2}{E\bar{y}^2}$, and

$$E\bar{y}^2 = \mu^2 + \frac{\gamma_0}{T^2}(T + 2(T-1)\rho_1 + 2(T-2)\rho_2 + \ldots + 2\rho_{T-1}) = \mu^2 + \Sigma_{\bar{y}}^2 \tag{2.1}$$

with $\Sigma_{\bar{y}}^2 = Var(\bar{y})$ given by

$$\Sigma_{\bar{y}}^2 = \frac{\gamma_0}{T^2}(T + 2(T-1)\rho_1 + 2(T-2)\rho_2 + ... + 2\rho_{T-1}) \tag{2.2}$$

and the MSE is $MMSE_T = \sigma^2 + \mu^2(1 - \frac{\mu^2}{E\bar{y}^2}) = \sigma^2 + \frac{\mu^2\Sigma_{\bar{y}}^2}{\mu^2 + \Sigma_{\bar{y}}^2} \leq \sigma^2 + \Sigma_{\bar{y}}^2$.

Now suppose that the mean of the stationary series shifts from μ to μ_1 at time $T+1$ and also assume that the autocorrelation structure remains the same after the shift. It can be easily shown that the MMSE estimate of μ_1 is $c\bar{y}$ where $c = \frac{\mu\mu_1}{E\bar{y}^2}$. $E\bar{y}^2$ is as in (2.1) and $\Sigma_{\bar{y}}^2$ is as in (2.2). The MSE is $MMSE_T = \sigma^2 + \mu_1^2(1 - \frac{\mu^2}{E\bar{y}^2}) = \sigma^2 + \frac{\mu_1^2\Sigma_{\bar{y}}^2}{\mu^2 + \Sigma_{\bar{y}}^2}$. The minimum mean square estimate of μ_1, $c\bar{y}$ has smaller MSE than the sample mean as c was chosen optimally.

2.2 GARCH Models

Recently there has been growing interest in using nonlinear time series models in Finance and Economics. A nonlinear model was proposed in Abraham and Thavaneswaran (1991) and using nonlinear state space formulation, filtering and smoothing has been studied (see Granger (1998) for more details). Many financial series, such as returns on stocks and foreign exchange rates, exhibit leptokurtosis and time-varying volatility. These two features have been the subject of extensive studies ever since Nicholls and Quinn (1982) and Engle (1982) reported them. The Random Coefficient Autoregressive (RCA) models, the autoregressive conditional heteroscedastic (ARCH) models (see Engle (1982)), and the GARCH models of Bollerslev (1986) provide a convenient framework to study time-varying volatility in financial markets. In Thavaneswaran et al. (2004) the correlation properties for RCA GARCH models have been studied in detail.

Consider the general class of GARCH(p, q) models for the time series y_t where

$$y_t = \sqrt{h_t}Z_t, \tag{2.3}$$

$$h_t = \sigma_0^2 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_i h_{t-j}, \tag{2.4}$$

where Z_t is a sequence of independent, identically distributed random variables with zero mean, unit variance. Let $u_t = y_t^2 - h_t$ be the martingale difference and let σ_u^2 be the variance of u_t. Then (2.3) and (2.4) could be written as:

$$\begin{aligned} y_t^2 - u_t &= \sigma_0^2 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_i h_{t-j}, \\ \phi(B) y_t^2 &= \sigma_0^2 + \beta(B) u_t. \end{aligned} \tag{2.5}$$

where, $\phi(B) = 1 - \sum_{i=1}^{r} \phi_i B^i$, $\phi_i = (\alpha_i + \beta_i)$, $\beta(B) = 1 - \sum_{i=1}^{q} \beta_i B^i$ and $r = \max(p, q)$. We shall make the following stationarity assumptions for y_t^2 which has an ARMA(r, q) representation.

(A.1) all the zeroes of the polynomial $\phi(B)$ lies outside of the unit circle.

(A.2) $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ where the $\psi_i's$ are obtained from the relation $\psi(B)\phi(B) = \beta(B)$ with $\psi(B) = 1 + \sum_{i=1}^{\infty} \psi_i B^i$.

The assumptions ensure that the $u_t's$ are uncorrelated with zero mean and finite variance and that the y_t^2 process is weakly stationary. In order to calculate the variance and kurtosis for a GARCH process in terms of the ψ weights, we have the following theorem.

Theorem 2.1 For the GARCH(p, q) process volatility forecasts specified by (2.5), under the assumptions of stationarity and finite fourth moment, the kurtosis $K^{(y)}$ of the process is given by:

(a) $K^{(y)} = \dfrac{E(Z_t^4)}{E(Z_t^4) - [E(Z_t^4) - 1]\sum_{j=0}^{\infty} \psi_j^2}$,

(b) The variance of the y_t^2 process is $\gamma_0^{y^2} = \sum_{j=0}^{\infty} \psi_j^2 \sigma_u^2$, where $\sigma_u^2 = \dfrac{\mu^2(K^{(y)} - 1)}{\sum_{j=0}^{\infty} \psi_j^2}$ and $\sigma_y^2 = E(y_t^2) = \dfrac{\sigma_0^2}{1 - \phi_1 - \phi_2 - \cdots - \phi_r}$,

(c) $Var(y_t^2) = (K^{(y)} - 1)[Var(y_t)]^2$. That is, the square of the coefficient of variation for the squared process y_t^2 is $K^{(y)} - 1$.

Proof: Part (a) of Theorem 2.1 somewhat parallels the proof of Theorem of 2.1 in Thavaneswaran et al.(2004) for kurtosis. Part (b) follows from (2.5),

$$\sigma_y^2 = E(y_t^2) = \frac{\sigma_0^2}{1 - \phi_1 - \phi_2 - \ldots - \phi_r},\ K^{(y)} = \frac{E(y_t^4)}{(E(y_t^2))^2} = \frac{E(y_t^4)}{\sigma_y^2}, \text{ and } \sigma_u^2 = \frac{\sigma_y^2(K^{(y)} - 1)}{\sum_{j=0}^{\infty} \psi_j^2}.$$

Part (c) follows from the fact that the square of the coefficient of variation of y_t^2 equals $\dfrac{Var(y_t^2)}{(Var(y_t))^2}$.

Suppose we have observations from a GARCH(p,q) process and the unconditional mean of y_t^2, σ_0^2 shifts from σ_0^2 to σ_1^2 at time $t = T+1$. The following theorem extends Theorem 2.1 for the GARCH processes.

Theorem 2.2: For the GARCH processes described in (2.1) and (2.2), under suitable stationarity conditions,

(a) The improved estimate of $\sigma_1^2 = E(y_{T+1}^2)$ is $c\bar{y^2}$ where $\bar{y^2} = \frac{1}{T}\sum_{k=1}^{T} y_k^2$ and $c = \left[1 + \frac{(K^{(y)}-1)}{T}\right]^{-1}$.

(b) The MSE of the estimate of σ_1^2, $c\bar{y^2}$, is less than the MSE of the estimate of $\bar{y}^2$, of σ_1^2.

Proof: The proof of Part (a) follows from the fact that $Var(y_t^2) = (K^{(y)}-1)[Var(y_t)]^2$, and by the definition of the coefficient of variation. The proof of Part (b) is similar to the proof for linear time series models.

3 A Regression Model for a Single Structural Break

Consider the simple linear regression model studied in Pesaran and Timmerman (2004) of the form

$$y_{t+1} = \beta_1 x_t I_{(t \le T_1)} + I_{(t > T_1)} \beta_2 x_t + a_{t+1} \tag{3.1}$$

possibly with a shift in its variance from σ_1^2 to σ_2^2 . We know that β has changed at T_1, our interest lies in forecasting y_{T+1} given the observations $(x_t, y_t), t = 1, 2, \ldots, T$.

The problem of the number of observations needed to estimate a model that, when used to generate forecasts, will minimize the expected mean squared forecast error has been studied in Pesaran and Timmermann (2004).

Let m denote the starting point of the most recent observations to be used in estimation for the purpose of forecasting y_{T+1}. Then the least squares estimator (LSE) can be written as

$$\hat{\beta}_T(m) = \left(\sum_{t=m}^{T} x_{t-1}^2\right)^{-1} \left(\sum_{t=m}^{T} x_{t-1} y_t\right). \tag{3.2}$$

The forecast of y_{T+1} is given by $\hat{y}_{T+1} = \hat{\beta}_T(m)x_T$, with

$$\hat{\beta}_T(m) = \theta_m\beta_1 + (1-\theta_m)\beta_2 + \nu_T(m) \tag{3.3}$$

where $\nu_T(m) = \left(\sum_{t=m}^{T} x_{t-1}^2\right)^{-1} (\sum_{t=m}^{T} x_{t-1}a_t)$, and $\theta_m = \left(\sum_{t=m}^{T} x_{t-1}^2\right)^{-1} (\sum_{t=m}^{T_1} x_{t-1}^2)$. The forecast error in the prediction of y_{T+1} will be a function of the data sample used to estimate β and is given by the following: $e_{T+1}(m) = a_{T+1} + \theta_m(\beta_2 - \beta_1)x_T - \nu_T(m)x_T$. And, $E[\nu_T(m)^2|X_T] = \left(\sum_{t=m}^{T} x_{t-1}^2\right)^{-1} (\sigma_1^2\theta_m + (1-\theta_m)\sigma_2^2)$. The corresponding conditional MSFE becomes $E[e_{T+1}^2(m)|x_T] = \sigma_2^2 + \theta_m^2(\beta_2-\beta_1)^2x_T^2 + [\left(\sum_{t=m}^{T} x_{t-1}^2\right)^{-1} (\sigma_1^2\theta_m + (1-\theta_m)\sigma_2^2)]x_T^2$. For the proposed the MMSE predictor as $y_{T+1} = c\hat{\beta}_T(m)x_T$ where $c = \dfrac{\beta_2 E\hat{\beta}_T(m)}{E\hat{\beta}_T^2(m)}$, $e_{T+1}(m) = y_{T+1} - c\hat{\beta}_T(m)x_T = (\beta_2 - c\hat{\beta}_T(m))x_T + a_{T+1}$. The corresponding conditional MSFE becomes $E[e_{T+1}^2(m)|x_T] = \sigma_2^2 + [\beta_2^2(1 - \dfrac{(E\hat{\beta}_T(m))^2}{E\hat{\beta}_T^2(m)}]x_T^2$ which is smaller than the conditional MSFE of Persaran and Timmerman (2004).

3.1 Recursive Forecasts with Structural Change

For the model considered in Section 3, the estimate of β_2 with the starting point m of the most recent observations can be written as

$$\hat{\beta}_T(m) = \left(\sum_{t=m}^{T} x_{t-1}^2\right)^{-1} (\sum_{t=m}^{T} x_{t-1}y_t) = k_m(\sum_{t=m}^{T} x_{t-1}y_t), \tag{3.4}$$

where $k_m^{-1} = \left(\sum_{t=m}^{T} x_{t-1}^2\right)$, $k_m^{-1} - k_{m+1}^{-1} = x_{m-1}^2$.

In order to determine m,

Step 1: Recursively calculate $\hat{\beta}_T(m)$ and k_m^{-1} for each m using $\hat{\beta}_T(m) = \hat{\beta}_T(m+1) + k_m x_{m-1}\left(y_m - x_{m-1}\hat{\beta}_T(m+1)\right)$.

Step 2: Calculate the improved forecast and the corresponding conditional minimum mean square error for each m.

Step 3: Select the value of m which minimizes the minimum value given in Step 2 as the optimal choice of the window.

4 Conclusions

Recent evidence suggests that many economic time series are subject to structural breaks, yet little is known about the properties of alternative estimation/forecasting methods for such data. In this paper, we have proposed an improved biased estimator/forecast of the unconditional mean for linear as well as some nonlinear models (such as GARCH).

REFERENCES

[1] Abraham, B. and Thavaneswaran, A. (1991). A nonlinear time series model and estimation of missing observations. *Annals of the Institute of Statistical Mathematics*, **43**, no.3, 493-504

[2] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, **31**, 307-327.

[3] Clements, M.P. and Hendry, D.F. (1998). Forecasting Economic Time Series. *Cambridge University Press*, Cambridge.

[4] Engle R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50**, 987-1008.

[5] Granger, C.W.J. (1998). Overview of Non-Linear time series specification in Economics, *Berkeley NSF-Symposia*.

[6] Godambe, V.P. (1999). Linear Bayes and optimal estimation. *Annals of the Institute of Statistical Mathematics*, **51**, 201-215.

[7] Hendry, D.F. and Clements, M.P. (2003). Economic forecasting: some lessons from recent research. *Economic Modelling*, **20**, 301–329.

[8] Nicholls, D.F., Quinn, B.G. (1982). Random Coefficient Autoregressive Models: An Introduction. *Lecture Notes in Statistics*, **11**, Springer, New York.

[9] Pesaran, M.H. and Timmerman, A. (2004). Selection of estimation window with strictly exogenous regression. Working Paper.

[10] Shalabh (2001). Estimation of bias and standard error of an improved estimator of normal mean, *Metrika*, **54**, 43–51.

[11] Thavaneswaran, A., Appadoo, S.S., and Samanta, M. (2005). Random Coefficient GARCH models. *Mathematical and Computer Modeling*(To appear).

[12] Thavaneswaran, A. and Heyde, C.C. (1999). Prediction via estimating functions. *Journal of Statistical Planning and Inference*, **77**, 89-101.

[13] Thavaneswaran, A. and Thompson, M.E. (2002). Non-normal filtering via estimating functions. *Advances on theoretical and methodology aspects of probability and statistics* (Hamilton, ON. 1998). 173-183, Taylor and Francis: London.