Forecasting with Serially Corrected Regression Models

Yue Fang
Lundquist College of Business
University of Oregon, Eugene, OR 97403
yfang@oregon.yoregon.edu

Sergio G. Koreisha
Lundquist College of Business
University of Oregon, Eugene, OR 97403
sergiok@oregon.uoregon.edu

Since Cochrane and Orcutt (1949) and Durbin and Watson (1950) developed an approximate transformation to deal with and test for autoregressive disturbances of order 1 we have witnessed a plethora of studies dealing with the problems of serial correlation in regression models. Choudbhury *et al.* (1999), among others, have documented the evolution of approximate/exact transformations to deal with more complex serial correlation structures, as well as, the development of new, more powerful estimation methods that have occurred in the last fifty years. When disturbances exhibit serial correlation, least squares will yield unbiased, but inefficient estimators of parameters of the model, thus invalidating all tests of significance. In addition these estimated regression coefficients will have larger sampling variances than other estimators such as generalized least squares (GLS) that deal explicitly with the autocorrelation in the residuals. Furthermore, forecasts generated from such models can be seriously inefficient, not only because of the inefficiency of the parameter estimators, but also because the error between the fitted and actual value in the last observation is apt to persist into the actual forecast interval.

With few exceptions most of the studies that have been conducted on this topic assume that the error covariance matrix Ω from the regression model, $Y = X\beta + \epsilon$, where Y is a $(T \times 1)$ vector of observations on a dependent variable, X is a $(T \times k)$ design matrix and ϵ is a random vector, which follows an ARMA process, with $\mathrm{E}(\epsilon) = 0$ and $\mathrm{E}(\epsilon\epsilon') = \sigma^2\Omega$, is either known or could be estimated consistently from data. Until recently, of the very few studies that considered the properties of estimators, or of the forecasts when the structure of Ω was incorrectly identified, or when its parameters were inefficiently estimated, most dealt with or depended on asymptotic results (Goldberger, 1962; Amemiya, 1973; Engle, 1974). Koreisha and Fang (2000), on the other hand, compared the finite-sample efficiencies of ordinary least squares (OLS) and GLS vis-à-vis incorrect GLS (IGLS) estimators, i.e., estimators based on incorrectly identified Ω, and established theoretical efficiency bounds for IGLS relative to GLS and OLS. They found that GLS estimation based on autoregressive representations of autoregressive-moving average (ARMA) disturbances yielded more efficient estimates than OLS particularly when the order of the autoregression was set near $[\sqrt{T}/2]$, and that the differences in estimation efficiency between estimated IGLS and GLS were small.

In this article we augment the work of Koreisha and Fang (2000) by investigating the impact that estimated IGLS corrections may have on the forecasting performance of regression models with serial correlation. This work differs from others dealing with regression forecasts with autocorrelated disturbances such as Armstrong (1978) and Dielman (1985) in that it does not assume that the form of the autocorrelation is known (AR(1) in those cases) a priori. The main issue here is not determining which estimation procedure yields the best forecasts when Ω is known. Our goal is to show, both theoretically and with simulated data, that there exists a certain class of estimators based on incorrectly identified residual autocorrelation structures, namely, AR($\tilde{p}$) corrections with $\tilde{p} \approx [\sqrt{T}/2]$, that for finite samples can yield as good, and more often, better forecasts than those generated from OLS or from GLS using the correct form of the residual autocorrelation structure.

We have examined the relative forecast efficiency of GLS and IGLS predictors for regression models with serial correlation vis-à-vis each other and OLS predictors. We have derived new theorems associated with these predictors and established the form of the predictive mean squared errors as well a their magnitude relative to each other. From a large simulation study we have also found that for finite samples, estimated GLS corrections including those based on incorrectly identified disturbance structures yield more efficient short and medium term forecasts than OLS. Furthermore, when the order of autoregressive corrections is set at $[\sqrt{T}/2]$ the differences in forecast efficiency between EAR($\tilde{p}$) and EGLS is very small. This suggests that there is not much to be gained in trying to identify the correct order of OLS residuals when generating short to medium term forecasts. On the other hand, for longer horizons, it appears that OLS yields forecasts that are just as efficient as EGLS.

References

[1] Amemiya, T. (1973) Generalized least squares with an estimated autocovariance matrix. *Econometrica,* **41**, 723-732.

[2] Armstrong, J. (1978) *Long-Range Forecasting: from Crystal Ball to Computer.* New York: Wiley.

[3] Choudhury, A., Hubata R. and Louis R. (1999) Understanding time-series regression estimators. *The American Statistician*, **53**, 342-348.

[4] Cochrane, D. and Orcutt G. (1949) Application of least squares regression to relationships containing autocorrelated error terms. *Journal of American Statistical Association,* **44**, 32-61.

[5] Dielman, T. (1985) Regression forecasts when disturbances are autocorrelated. *Journal of Forecasting,* **4**, 263-271.

[6] Durbin, J. and Watson G. (1950) Testing for serial correction in least squares regression. *Biometrika,* **37**, 409-428.

[7] Engle, R. (1974) Specification of the disturbance for efficient estimation. *Econometrica,* **42**, 135-146.

[8] Goldberger, A. S. (1962) Best linear unbiased prediction in the generalized linear regression model. *Journal of American Statistical Association,* **57**, 369-375.

[9] Koreisha, S. and Fang Y. (2000) Generalized least squares with misspecified serial correlation structures. *Journal of the Royal Statistical Society, Series B,* forthcoming.