Available online at www.sciencedirect.com

SCIENCE DIRECT®

International Journal of Forecasting 22 (2006) 475–492

international journal of forecasting

www.elsevier.com/locate/ijforecast

Twenty-five years of progress, problems, and conflicting evidence in econometric forecasting. What about the next 25 years?

P. Geoffrey Allen*, Bernard J. Morzuch[1]

Department of Resource Economics, Stockbridge Hall, University of Massachusetts, 80 Campus Center Way, Amherst, MA 01003-9246, USA

Abstract

In the early 1940s, the Cowles Commission for Research (later, the Cowles Foundation) fostered the development of statistical methodology for application in economics and paved the way for large-scale econometric models to be used for both structural estimation and forecasting. This approach stood for decades. Vector autoregression (VAR), appearing in the 1980s, was a clear improvement over early Cowles Foundation models, primarily because it paid attention to dynamic structure. As a way of imposing long-run equilibrium restrictions on sets of variables, cointegration and error-correction modeling (ECM) gained popularity in the 1980s and 1990s, though ECMs have so far failed to deliver on their early promise. ARCH and GARCH modeling have been used with great success in specialized financial areas to model dynamic heteroscedasticity, though in mainstream econometrics, evidence of their value is limited and conflicting. Concerning misspecification tests, any model will inevitably fail some of them for the simple reason that there are many possible tests. Which failures matter? The root of the difficulty regarding all issues related to modeling is that we can never know the true data generating process. In the next 25 years, what new avenues will open up? With ever greater computational capacity, more complex models with larger data sets seem the way to the future. Will they require the automatic model selection methods that have recently been introduced? Preliminary evidence suggests that these methods can do well. The quality of aggregate data is no better than it was. Will greater use of more disaggregated data be sufficient to provide better forecasts? That remains an open question.

Keywords: VAR; Cointegration; Error correction; Dynamic stochastic general equilibrium; Leading indicators; ARCH; GARCH; Automatic model selection

1. Introduction

The ending of a millennium and the quarter-century of a journal's existence provide an opportunity for retrospection and prospection. It is unsurprising, then that in January 2001, the 100th volume of the *Journal of Econometrics* carried a special issue doing just that. A number of prominent econome-

* Corresponding author. Tel.: +1 413 545 5715; fax: +1 413 545 5853.

E-mail address: allen@resecon.umass.edu (P.G. Allen), morzuch@resecon.umass.edu (B.J. Morzuch)

[1] Tel.: +1 413 545 5715; fax: +1 413 545 5853.

0169-2070/$ - see front matter © 2006 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2006.03.003

tricians were invited to write brief essays from their perspective on important past developments and likely future directions, while others took the roles of discussants and commentators. Diebold (2001) commented that the twelve essays illustrated the central themes of econometrics: the sweeping effect of improvements in information technology, that empirical finance and time-series econometrics are natural partners, and the continued importance and rapid development of ideas and methods related to forecasting. He notes that almost all the essays are concerned one way or another with forecasting. That statement captures the relationship between econometrics in general and econometric forecasting in particular.

Although forecasts of macroeconomic variables from a substantial proportion of the output of econometric forecasters, noticeable effort has gone into forecasts of prices and quantities of industry and sectoral outputs, particularly in agricultural sectors where government data collection has a long history. For more disaggregated variables, fewer forecasts can be found, in part because individual companies have incentives to keep their sales forecasts to themselves.

We take Diebold's first point as our main theme in this paper: developments in applied econometrics have always operated under the constraints imposed by a limited number of data points and limited computing power. We expect the future to be no different, and developments in econometric forecasting are likely to take advantage of increases in both of these. An easy prediction to make is that improvements in processing speed as well as data storage capabilities, and the electronic capturing of ever more high frequency data will continue to grow. Other forces expected to influence the direction are the impact of paying greater attention to the role of the forecaster (and the forecast) in decision making, and the need to provide more information in the form of forecast error distribution statistics.

Twenty-five years ago, 1980 marked a watershed moment in econometric forecasting. Two of its most useful tools began their rise in popularity at this time: vector autoregression and error correction models. So also did autoregressive conditional heteroscedasticity and the generalized method of moments, which became the main tools of financial econometrics. Twenty-five years ago, financial econometrics did not exist. Now it is a major area deserving its own review. None of the developments we describe started 25 years ago. Like any scientific endeavor, their roots can be traced back further, sometimes much further. Nevertheless, much of what is happening today in econometric forecasting received a dramatic boost in the early 1980s.

The structure described by Stock and Watson (2001), though in the context of macroeconomics, is a useful way of characterizing what econometricians do: describe and summarize data, make forecasts, make structural inferences, and analyze policies. These represent and require successively deeper levels of understanding about economic systems. Data description and summarization, which is what a statistic is, really falls under the job description of statisticians, since it requires no theory about what generated the observations. It captures correlation, not causation.

For forecasting, values of structural parameters are unimportant. Consequently, forecasting is frequently done with reduced-form models, following Klein's (1950) suggestion that forecasting can use different econometric practices from explanation. That is, attention is paid to the set of causal variables, not to the structural relationships. The definition of causal variables includes the particular lag or lags, if any, specified for each variable. This can be justified theoretically by realizing that forecasts from a structural model are just a function of current and past data. If the function can be estimated consistently then the resulting forecasts will have the same forecast error variance, asymptotically, as if the function were known (Stock, 2001).

In an attempt to illustrate the success of their endeavor, econometric forecasters often perform static simulations, in which the historical record is simulated by the model, using actual values of exogenous and lagged endogenous variables. Recognizing that the actual values of at least some of these variables will be unknown for the forecast horizon in question, it is claimed that the exercise merely demonstrates the performance of the model system, which is neither necessary nor sufficient to make useful forecasts. Arguably better is dynamic simulation, where actual exogenous variables and model-generated lagged endogenous variables are used as inputs, though this too has been shown to be an invalid model evaluation