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

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.

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 technique (Hendry & Richard, 1982). To properly measure the forecast performance of an econometric model, first, forecasts should be for outside the period used for estimation and, second, all actual variables that would be unknown at the time the forecast is being made should be replaced by their forecasts.

In a review of forecasting, we have less interest in the last two activities of econometricians, inference and policy analysis. Structural models are often used to make forecasts as well as structural inferences. These are cause and effect relationships measured over time. Historically, short-term and long-term elasticities have been used. For demonstration they rely on showing the effect of a change in a causal variable or an impulse shock to the system that may or may not happen. Therefore, there is no check on the accuracy of the inference beyond tests of model adequacy (specification and misspecification tests). One can also use the approach to demonstrate the sensitivity of the model to particular shocks, which might indicate the areas where the model needs to be examined in more detail. But again, this is the *model* response, not reality. A structural model that embeds theory regarded as widely accepted and has been shown to closely approximate the historical record can provide forecasts that appeal to decision makers. Whether the forecasts are more accurate because of these facts is a different matter entirely.

Finally, a model for policy analysis requires that a proposed policy change be related to changes in the model's causal variables. A common problem is that at the level of aggregation commonly employed, parameters represent a mixture of technical and behavioral responses, making them subject to the Lucas critique. From a theoretical standpoint, analysis based on decision rules is fundamentally defective for producing conditional forecasts. The parameters are not constant because a change in policy induces a change in behavior and the model is not sufficiently "deep" to describe how the parameters change. Whether the problem is of any consequence in practice is less clear. Rudebusch (2005) argues that it is not: although the behavior of US monetary policymakers has changed during the past few decades, statistical analyses of reduced forms, such as VARs, often have not rejected the null of structural stability. The magnitude of policy shifts has been relatively small and the particular reduced forms relatively robust.

2. Early work

Econometric forecasting has a much longer history than this twenty-five year retrospective. (For a brief review emphasizing developments in theory prior to the 1980s, see Clements & Hendry, 1998.) One of its pioneers was Charles Sarle (Sarle, 1925), whose single equation model for forecasting the price of hogs was published in a special supplement to the *American Economic Review*. His work won the Babson prize, awarded for the best essay submitted by a student, as judged by a committee of eminent economists. Sarle was several decades ahead of his time. He used lagged explanatory variables, so that their values were known at the time of forecast, an early form of leading indicator; and he performed both within-sample and out-of-sample forecasts. Although his work was published in the leading economic journal, it was then largely ignored. Such is the fate of many a pioneer. Why this occurred is the subject of a fascinating reappraisal of Sarle's work by Gordon and Kerr (1997). With the advantage of modern techniques and computing power, Gordon and Kerr determined that Sarle's model was reasonably well specified. They surmise that it remained unknown for two reasons. First, Ezekiel, who did become well known, wrote a subsequent article in which he criticized Sarle's choice of variables. Second, econometric forecasting lost popularity shortly after the publication of Sarle's article. Arguments circulating at the time that the nature of economic data made forecasting them impossible were apparently persuasive enough to lead to the demise of econometric forecasting. It reappeared in the 1940s and 1950s, by which time articles published in the 1920s had been largely forgotten.

By the 1920s, the product-moment method of calculating correlations and regressions was known, thanks to the path-breaking paper by Karl Pearson (Pearson, 1896). Today, Sarle's model and method are regarded as primitive, handicapped as he was by both the length of variable series he could collect, and by the processing power of mechanical calculators then available.

3. Development of econometric models

The Cowles Commission was founded by Alfred Cowles in 1932. It gave its name to the approach that came to dominate macroeconometric practice (and much other econometric work) for the latter half of the 20th century. As Christ (1994, p. 31) notes: "The Cowles program was intended to combine economic theory, statistical methods, and observed data to construct and estimate a system of simultaneous equations that could describe the workings of the economy." Its followers expected economic theory to guide them in the choice of variables and functional form to be used in a particular situation. We have come to realize that economic theory is capable of much less than this, particularly in specifying dynamic relationships. (See Blaug, 1992, p. 87 for a highly entertaining view.) The emphasis on appropriate econometric methods and the almost complete dismissal of measurement error (in sharp contrast to engineers) continues to influence the discipline.

Work began in earnest in the 1940s. In two fundamental papers, Haavelmo (1943, 1944) first provided a demonstration of the role of an explicit probabilistic framework and then explained the concept of a simultaneous equations model. The first macroeconomic model using equation-by-equation estimation on 21 annual observations of the United States economy appeared in Klein's (1950) monograph. With the mechanical calculators available at the time, this represented the limit of a system that could be estimated. As computing power and data series both increased, Klein's simple models became the basis of more complex macroeconomic models that were used for forecasting.

By 1980, much dissatisfaction existed over the state of econometric forecasting. The state of the art consisted of some single-equation models and small to large systems used in macroeconomics and microeconomics (industry sectors, especially agriculture) for forecasting, structural inference and policy analysis. The oil crisis of 1972 led to a series of price shocks that were not picked up by structural models. Some attempts were made to fix the problems with the Keynesian-based structural models, for example, by the introduction of rational expectations (Fair, 1984). Fair's model is relatively small (30 stochastic equations, about 100 identities) compared with some of the macroeconomic models developed using the Cowles Commission approach, has a long forecasting history uncontaminated by judgmental fixes (the norm with these kinds of models), and is freely available for experimentation (at http://fairmodel.econ.yale.edu/main2.htm). These modifications have not solved problems with forecasting, as is illustrated by a story told by Zellner (2001, p. 93): "...in answer to my question as to whether his model caught the 1991 downturn in a U. of Chicago workshop a few years ago, Ray Fair answered, 'Damn it, Arnold, you had to ask that question. My model missed the downturn along with all of the others'."

4. Vector autoregression

The appearance of vector autoregression (VAR) signalled the beginning of a revolution in econometric forecasting. The philosophy switched from an emphasis on picking theoretically justified causal variables with little attention to dynamics, to emphasizing dynamics using a short list of causal variables. A VAR is just a multivariate generalization of a univariate autoregressive process usually associated with the work of Box and Jenkins (1970), though its roots can be found in Quenouille's (1957) book; however Sims' (1980) highly cited article makes no mention of either.

The impetus for VAR appears to be grounded in the limitations involved with the structural approach to simultaneous equations estimation. Beginning with Haavelmo's series of published papers in 1943–44 on simultaneity, literature on this subject dealt mainly with overidentified relationships. Liu (1960) used the Klein–Goldberger model of the US economy to trace the reasons for this apparent preoccupation. In so doing, he articulated the dilemma faced by structural model builders and suggested several reasonable courses of action. He noted, for example, that including a complete list of variables theoretically belonging in the model's investment function leads to such an enlarged specification that it became underidentified.

The proverbial pendulum then swung in the opposite direction, and the accepted solution had been to whittle down the "appropriate" set of variables, resulting in overidentification. These early models were characterized by fitting oversimplified relationships. Experiments by practitioners using different combinations of explanatory variables — and whose intention it was to obtain a set of significant and reasonable parameter estimates — had become the guiding principle for model specification. Liu (1960) emphasized that this approach was anything but an estimate of the structure. As he so concisely stated (p. 860), although it is the econometrician's goal to estimate economic structures, it should not be the econometrician's job to derive "structural" relationships by artificially overidentifying the structure.

In turn, Liu suggested that it would be more appropriate to direct efforts toward straightforward forecasting and to the development of techniques for dealing with underidentified relationships. His vehicle for exploration was the reduced form. In this context, he promoted the usefulness of lagged variables in the reduced form, citing the work of Nerlove and Addison (1958) in adaptive expectations and supply response. This was the thinking that eventually led to VAR.

Sims (1980) extended Liu's arguments and, in the process, promoted VAR modeling. (Although he fails to say so in his article, his approach is a direct descendant of the multivariate case described by Mann and Wald (1943). Epstein's (1987) history provides a more accessible description of their work and a critique of Sims' arguments.) Effectively, he brought Liu's work to a new level. One of Sims' enlightening points concerned normalization. If a parameterization from economic theory, e.g., a structural form, fails to be identified, we can transform or map all points in this original parameter space into the same points in a new parameter space. More specifically, and in the context of the original structural form, the resulting normalization is the reduced form. Furthermore, Sims indicates that most of the restrictions placed on underidentified structural models to make them identified are false in the first place. He asserts that the resulting restricted models are, then, only nominally overidentified. He refers to the restrictions that assist in achieving identification as *incredible* identification restrictions. Both Liu (1960) and Sims (1980) emphasize the importance of the reduced form being matched with the "correct" underidentified structure rather than with the "incorrect" overidentified structure. (Not all econometricians agree: one reviewer argues that Sims showed "a singular lack of understanding of simultaneous equations theory" in his characterization of identification restrictions.)

The system originally proposed by Sims was characterized as "profligate in parameters". This is only an issue in forecasting when measurement and estimation errors outweigh errors from misspecification. Zellner's (1992) advice to "keep it sophisticatedly simple" (the KISS principle) turns out to be powerfully applicable. Sargent and Sims (1977) proposed introducing the variables through two or three indexes; this approach did not catch on. Restricting the parameters on the longest lags of all the variables in an n-equation VAR to zero imposes n^2 restrictions, the fastest way of reducing parameter profligacy. Commonly, in following this strategy, a specification test determines how much the lag length can be shortened. Many studies have confirmed that this simplification strategy improves forecast accuracy. Further reduction in lag length, variable by variable, can be entertained, though the number of possibilities is large and the strategies are usually ad hoc. This has not been a popular route to simplification. Forecasters who follow it usually adopt Hsiao's (1979) method, in order to limit the number of restricted models entertained. Bayesians employed an alternative approach to dealing with parameter profligacy, imposing priors on each parameter, frequently of zero, based on the work of Shiller (1973) and Leamer (1973).

Sims also noted that the structural equations of the early large-scale macroeconometric models were usually estimated one equation at a time rather than as a system. This removed the possibility of looking at feedback between variables in the different equations. Not accounting for feedback when it actually exists is another form of an incredible restriction.

During the middle 1980s to the early 1990s, a large amount of attention was paid to specification issues in VAR models. Three types of VAR models had emerged. They were classified as reduced form, recursive, and structural. Reduced-form VARs express each dependent variable as a function of past values of everything in the model (this is the form of VAR addressed above). The error term (or shock) is assumed to be serially uncorrelated with itself, but it can be correlated with other error terms across equations.

In a recursive VAR, the first equation of the model, for example, expresses the dependent variable as a function of past values of itself and of other variables in the model. The second equation expresses its dependent variable as a function of past values of itself, of other variables in the model, and of the current value of the dependent variable appearing in the first equation. In the third equation, the pattern of the second equation repeats, but this time it includes the current value of the dependent variables appearing in the first and second equations. The process continues for the remaining equations. Equation ordering affects estimation results. The end result is a triangular simultaneous equations system with richer dynamics than is traditional. Krolzig (2003) proposed this as the starting point for a general-to-specific model reduction. We know of no forecasting applications.

An important source of information in a VAR is the estimated residual or shock in each equation. To appreciate the information contained in these shocks, we simply need to recall that an autoregression has a moving average representation and, equivalently, that a vector autoregression (VAR) can be written as a vector moving average (VMA). This means that the left-hand-side variables in a VAR model can be expressed in terms of current and past values of the shocks of the system. As presented in Sims (1980), this VMA representation allows one to trace out the time path of the various shocks on the variables contained in the VAR model. This led to the development of impulse response functions in a VAR setting.

When the standard reduced-form VAR was used for structural inference by imposing shocks to the system, macroeconomic modelers complained that the results could not be tied to economically meaningful variables. Identifying restrictions are needed, bringing us full circle to where a structural VAR resembles a structural simultaneous equations model. Generally, though not always, the resulting system is just identified. As is well known from classical simultaneous equations work, if each structural equation is just identified, the structural parameters can be recovered from the reduced form. Conversely, if sufficient restrictions are imposed on a reduced form system, one can just identify structural parameters tied to particular variables. As Dhrymes and Thomakos (1998) point out, this is a reparameterization of the reduced form which is untestable, not unique and sheds no additional light on the system over and above that illuminated by the reduced form. Other criteria based on economic theory or information about dynamics must be used. We seem to be back with incredible identifying restrictions, though in the context of a model with richer dynamics. A large body of literature concerned with macroeconomic structural inference and policy analysis followed the original concepts of structural VARs, though their use for forecasting appears to be minimal and evidence on their forecast performance is almost non-existent.

As time passed, other important issues surfaced and received attention. For example, when it came to one important type of non-stationarity, estimation by way of time-varying parameters in a VAR was suggested as a viable procedure for capturing the effects of drift (Sims, 1993). Likewise, the perennial problem with VARs is the wholesale loss of degrees of freedom with the inclusion of additional lag structures. In this context, Bayesian approaches received increased attention as a way to impose a common structure on regression coefficients (Litterman, 1986).

There are challenges in the future relating to VAR modeling. If the shocks that are present in each VAR equation reflect omitted variables, and if these are correlated with included explanatory variables, the result is bias. When these effects end up in the error term, the shock that is used to estimate the impulse response is incorrect. This, indeed, is the all-too-familiar specification problem. Beyond challenges that address specification and information which can be used to impose identifying restrictions, others remain of a more technical nature. Work for the future remains in richer nonlinear structures extended to VARs.

As a forecasting method, has VAR been successful? The answer is a largely unqualified "yes". As Table 1 shows, neither single nor simultaneous

Table 1
Econometric versus univariate forecasts, recorded as (better, no difference, or worse) according to the specified accuracy criterion, by *series*, with the number of studies in parentheses

	Classical single equation	VAR	Structural sector model	All econometric
Pre-1985	71,6,46 (52)	4,0,4 (2)	114,10,79 (23)	189,16,129 (77)
Better as percent of total	58	50	*56*	*57*
1985 on	56,12,56 (39)	189,10,44 (44)	77,6,46 (9)	321,28,146 (91)
Better as percent of total	45	78	59	65

Most forecasts are one-step ahead and RMSE is the usual accuracy criterion.
Source: Chapter 11 of Armstrong (2001).

equation methods show impressive forecast accuracy compared with naïve benchmarks. Developments in econometric theory since about 1950 do not seem to have led to improvements, and given the discussion above, we can conclude that the simplification methods used are failing to deliver models that match the structure of the system they are meant to capture. Conversely, VAR models, which only appeared on the scene around 1985, adopt an approach that is noticeably more successful.

Since a VAR is a self-contained system, there is a tendency to believe that it represents the entire economy, and that the biases of the undoubtedly large number of omitted variables are taken care of by including a large enough number of lags of the variables actually included. Two specialized uses of the VAR approach are described later. When VAR models began to be used in dynamic stochastic general equilibrium, they were indeed intended to represent the entire economy. Familiar macroeconomic variables such as consumption, investment, and interest rate appeared. The theoretical underpinning of general equilibrium models implied a number of parameter restrictions, which if imposed created a structural model. Second, in forecasting using leading indicators, it is possible to construct a VAR out of the composite indexes (coincident and leading) or out of any combination of the individual components of the indexes. The atheoretical nature of VARs and of leading indicators gives them a natural connection.

5. Dynamic stochastic general equilibrium (DSGE) models

While forecasting requires only reduced-form models, policy analysis needs structural ones. Dissatisfaction with the early systems-of-equations models centered on the way decision rules was used to specify the structural relations. Lucas in particular, in a series of writings best summarized by his critique (Lucas, 1976) cogently argued that parameters dependent on agents' expectations would change when policies changed and were therefore inadequate for either forecasting or policy analysis. In addition, business cycles returned to the US economy in the 1970s and were not picked up by the forecasting models then in use. This paved the way for a return to the microeconomic roots of macroeconomics through general equilibrium theory.

The great strength of a DSGE model is that it is structural and based on economic theory: aggregate output, consumption and possibly other macroeconomic variables are linked to parameters describing agents' tastes and technologies. Its great disadvantage is that to maintain the theory-induced parameter restrictions all equations must be estimated simultaneously. The system is both small and necessarily stylized. Some reduction in size of system is possible by making use of identities. For example, making use of the well-known identity "output equals consumption plus investment", only two stochastic equations are needed for two variables.

Although all microeconomists accept the fundamental structure and optimizing behavior of general equilibrium theory, many simplifying assumptions are needed to make the theory operational, leading to many different models. A single representative agent is assumed, who consumes, invests and provides labor. The agent's production function is frequently assumed to be of Cobb–Douglas form. The first-order conditions arising from utility and profit maximization generate the observation equations of the system. In the simplest framework, dynamics are introduced with a single equation of motion describing technology shocks. Under more complex assumptions, innovations are introduced at other points, for example, changes in tastes.

Kydland and Prescott (1982) introduced multiperiod investment, inter-temporal leisure–work transfers and real technology shocks (hence "real business cycle" analysis) into a model whose parameters were calibrated rather than estimated. The real business cycle school has attracted its critics: calibration methods will not be accurate enough for forecasting purposes, and the stylized representative-agent models may not in fact address the Lucas critique. That is, although model construction is based on taste and technology structures which are invariant under different policy regimes, at the level of aggregation needed to make the approach operational, the parameter aggregates are not constant. Diebold (1998) provides a more detailed discussion and suggests several needed developments, of which the first is more heterogeneity, to more fully address the Lucas critique ("deeper" parameters).

Early real business cycle models were calibrated to match the historical behavior of the economy they represented. DSGE modelers estimated coefficients that fit historical data. They paid limited attention to the properties of coefficients, rather more attention to dynamic behavior through impulse response functions and no attention to out-of-sample forecast performance, particularly in comparison with other models. This is changing and is likely to become the focus of future research.

The resulting model contains the same kinds of variables as are found in systems of equations describing an economy: output, consumption, employment, etc. The same variables can be used in a VAR model and the question asked is: how can we define each model in a way that permits a realistic comparison of their performances? Ireland (2004) proposed one approach: augment the structural model with a serially correlated error term, so that the result is a DSGE model with its residuals described by a VAR. He compared out-of-sample forecasts one to four steps ahead from the hybrid DSGE model with those from an unrestricted first-order VAR in the same three variables (output, consumption and hours worked). (Investment was also forecast, using the identity described earlier.) Forecast accuracy was better for the hybrid model for almost all horizons and variables, and was frequently significantly better, indicating that the restrictions imposed by the DSGE model contained valuable information for the forecaster.

Del Negro, Schorfheide, Smets, and Wouters (2004) adopted a different strategy. Working with seven variables they first constructed a DSGE model then derived an approximate VAR representation with four lags. Their approach is Bayesian in spirit. The VAR has many implied cross-equation parameter restrictions. An ordinary unrestricted VAR can be viewed as having diffuse prior parameter restrictions, and the VAR that most closely approximates the DSGE model as having tight prior restrictions. This range is indexed by a hyperparameter that ranges from zero in the first case to infinity in the last one. By allowing the hyperparameter to take on intermediate values the method causes the DSGE restrictions to carry more or less weight in the estimation of coefficients.

Perhaps because the DSGE model is expressed as rates of change (in output, consumption, etc), the corresponding VAR is of all variables in first differences, which contains several unit root restrictions. Working backwards to a less-restrictive specification, the authors add three cointegrating vectors based on the DSGE model restrictions and produce a vector error correction model (VECM). Within sample, the DSGE-VAR models have higher likelihoods than DSGE-VECM models, suggesting that the differenced VAR is more robust to breaks than the VECM. In both cases the optimal hyperparameter value which maximizes the likelihood implies that the DSGE restrictions should carry little weight. There are benefits, though. For most of the variables and for most of the horizons (up to 12 steps ahead), out-of-sample forecasts are most accurate for the DSGE-VECM (using the optimal hyperparameter estimate), followed by the pure DSGE model, followed by the VECM.

Diebold (1998) predicted that DSGE models would grow in size from 3–4 variables to 8–10 because of the importance of parsimony and because parameters must be jointly estimated. Some models, like that described above (Del Negro et al., 2004) are almost at that size. It is hard to see how the call for more heterogeneity can be met with models of this size, so we might anticipate that with more computing power, these models will grow even larger, if they are to become successful forecasting tools. Diebold (1998) also predicted that the future will bring better stochastic dynamics for technology shocks, etc. — regime switching models, multiple sources of uncertainty, diagnosis of misspecifications via formal procedures, and greater use of techniques to "shrink" coefficients in particular directions, e.g., via Bayesian methods.

6. Leading indicators

Leading indicators have a long history in macroeconomic forecasting. Work on describing business cycles goes back at least into the 1920s with the work of Wesley Mitchell, followed by a search for leading, coincident and lagging indicators, leading to two landmark publications, one on statistical indicators (Mitchell & Burns, 1938) and the encyclopedic "Measuring Business Cycles" (Burns & Mitchell, 1946). Interest in leading indicator analysis rose during the early 1990s from almost nothing to a level that has been sustained ever since (based on citation counts in the database Econlit®). Diebold and Rudebusch (1996) and, most recently, Marcellino (2005) review the literature.

An economic variable should possess several properties if it is to be a useful leading indicator. It should consistently anticipate turning points of the target variable. It should also change in the same way as the target variable at other points of the business cycle. Requirements for consistent co-movement between leading and target variables led to the use of composite indexes. Leading indicators should provide accurate measures of the variable of interest, i.e., be subject to little measurement error, and also be subject to little later revision. Causes of business cycles remain not well understood. Concern that the connections between indicator and target variables had weak theoretical underpinnings led to the objection that leading indicator analysis was "measurement without theory," the title of a critical paper by Koopmans (1947).

In 1981, Geoffrey Moore, already an authority on leading indicator analysis, and later a fellow of the International Institute of Forecasters, delivered a keynote address at the first International Symposium of Forecasting. He compared the work of those wanting to monitor the business cycle with those wanting to forecast it, who were viewed then as rather separate groups. (Moore's co-author provides a later re-examination of the original topics, see Cullity, 1993.)

A working paper by Klein and Moore (1982), which seems to have never appeared as a journal article, summarized the non-model-based approach to the construction of composite indexes. Even then model-based approaches were being developed which would supplant the earlier method. The retrospective part of Klein and Moore's paper overlooks them entirely. Non-model-based composite indexes are essentially the simple average of standardized percentage changes of the variables in the index added to the previous value of the index. Standardization gives each transformed variable the same standard deviation. Such indexes are simple to construct and explain. They have received several criticisms: the indicator variables have no explicit connection with the target variable, the index uses fixed weights, no lagged values of target or indicator variables are used, even though they may contain useful information, and derivation of standard errors for the forecast of the target variable is difficult. If that sounds like an opportunity for vector autoregression models, such is indeed the case, yet a problem arises there too. Burns and Mitchell (1946) analyzed hundreds of economic series to produce composite indexes; a standard VAR model would only be able to handle a few of them.

Dynamic factor models, originally developed for cross-sectional analysis and adapted to time series by Geweke (1977) and Sargent and Sims (1977), provided one answer. They represent one line of research that addresses the early concern of measuring the consistent pattern of co-movement among leading and target variables. In a one-factor model (e.g., Stock & Watson, 1991) there is a single unobserved common factor. It can be thought of as the representation of many shocks affecting the system, which result in proportional effects on the variables. Stock and Watson (1991) made identifying assumptions and estimated the resulting state space model.

A second concern of early business cycle analysts was that leading indicators anticipated target variables differently in the expansion and contraction phases of the business cycle. In econometric terms, the parameters in the different phases of the cycle are not constant. Early regime-switching models relied on past values of the variables in the system to switch from one state to another. Hamilton (1989) proposed a Markov switching (MS) scheme that has since proved popular. In the simplest case of a binary switch between expansions and recessions, the binary state variable is not observable but depends on the value of the target variable, from which the probability of being in either state is calculated. In leading indicator work, it was first used by Lahiri and Wang (1994) to predict cyclical turning points. Filardo (1994) extended their work to permit time varying transition probabilities, which appeared to improve forecasts marginally. Diebold and Rudebusch (1996) provided another early analysis.

How well do leading-indicator models perform? Since forecasting is a major purpose of the leading indicator approach, one might think that this question had been investigated in detail, however, such does not appear to be the case. While there are many studies that compare the performance of different leading indicators in forecasting the target variable, or different combinations of leading variables, or different ways of weighting composite indicators, or the difference between using original variable values (so-called "real-time analysis) and later revisions, we found relatively few studies that compared forecasts across methods. The 22 papers collected by Lahiri and Moore (1991) illustrate the range of efforts, and only one (Webb, 1991) makes such a comparison. Using a fairly short time interval, from 1971 to 1983, Webb concluded that for a four-quarters-ahead turning-point prediction a VAR had a slight advantage, while for one quarter ahead, neither a VAR nor a composite leading indicator came out ahead of a naïve-no-change prediction. Some could argue that a method designed to predict turning points of business cycles, and which is inherently nonlinear, should not be compared with linear models such as VARs. However, linear models have been used for this purpose, and ideal leading indicators lead the target variable at all stages of the cycle.

Megna and Xu (2003) used six leading indicator variables to forecast the composite coincident indicator that they constructed for the New York state economy. They compared out-of-sample forecasts up to 12 months ahead from a standard unrestricted VAR with 6 lags, three variants of Bayesian VARs and a restricted VAR. As might be expected, the more parsimonious restricted VAR generally performed the best, followed by the less tightly restricted BVAR. Marcellino (2005) compared forecasts one and six months ahead from six different models plus both a simple average and mean-square-error weighted combined forecasts. Once again, combining was a successful strategy. The best single method was a five-equation VAR(2) with the four individual components of the coincident index and the composite leading indicator. A VAR(1) with the composite leading indicator decomposed into its 10 components gave poor forecasts, while decomposing both indexes (into a VAR(1) with 14 equations) was more successful for some reason. A VECM(2) with the two indexes plus an error correction term performed poorly, worse than a VAR(2) in levels, which is difficult to explain since it suggests that the latter was more robust to equilibrium shifts than the former.

Asymmetric responses to expansions and contractions have been demonstrated for the Markov-switching approach and seem to matter for forecast accuracy, at least some of the time. Clements and Krolzig (1998) presented graphical comparisons of one- through 16-steps-ahead forecasts of US GNP using univariate models only, over two different time periods (1985:1–1990:4 and 1991:1–1996:2). In the first time period, the MS model was slightly more accurate at shorter horizons (about 5% better than either a linear AR(4) or a self-exciting threshold autoregression (SETAR(2,2,2)) model). In the second time period, except for the poor performance of the SETAR model at shorter horizons, there was little to choose between them. Constant transition probability univariate MS models do not appear to offer much promise. Marcellino (2005) found the VAR(2) above had worse one-step-ahead turning-point predictions than the equivalent VAR-MS constant transition probability model. Layton and Katsuura (2001) found that MS regressions had much better one-step-ahead forecasts of an expansion or contraction of the economy than binomial logit, multinomial logit or binomial probit models. The MS model where the transition probabilities were allowed to vary depending on the value of a leading indicator was even better.

There are probably three lines of research that will be pursued in the future. There will be efforts to increase the number of common factors of the leading indicator variables, a form of principal components, to accommodate a larger number of variables with more flexibility, following the work of Stock and Watson (1991). A greater variety of nonlinear models than the Markov-switching model are likely to be tried. And following arguments by Clements and Hendry (1998, pp. 213–214), greater use of economic theory is likely, particularly if cointegration and co-breaking relationships can be inserted into the composite leading indicator. If series can be found that break together, because of a causal relationship, then this increases the prospect for consistent behavior, and if one variable consistently breaks before a target variable, then the ability to predict turning points of the target variable will be enhanced. Kajal Lahiri (personal communication) while not disagreeing with the above expectations, expresses regret at the impending demise of the old-style leading indicators, whose simplicity gives them popularity and continuing market value, noting that forecasts from more complex methods are little better.

7. Error correction and cointegration

In the context of a single equation, the term "error correction" appears to have been first applied by Hendry and von Ungern-Sternberg (1981), though the concept has a much longer history. Phillips (1954), whose statement on the relationship between wages and employment became known as the Phillips curve, noted (pp. 293–4) "Stabilisation policy consists of detecting any error and taking corrective action." In a later paper, referring to an electric-circuit type of diagram he states (Phillips, 1957, p. 267) "The relations shown at the top of the diagram represent an error correction type of stabilisation policy." But it was Sargan, in probably his best-known paper (Sargan, 1964), who proposed the equation in its currently recognizable form. He showed the decomposition of the equation into a dynamic adjustment component in response to error and an equilibrium condition. He noted, almost in passing, that his specification represented a correction to a 'Phillips' type equation. Hendry referred to the approach in a conference organized by Sims (Hendry, 1977), and it was used in the widely cited paper by Davidson, Hendry, Srba, and Yeo (1978) before getting the label that has stuck for many years. Only recently has Hendry (Clements & Hendry, 1996) begun referring to the phenomenon as "equilibrium correction". We can expect a few years of confusion before the term "error correction" stops being used for situations where "equilibrium correction" is in fact what is meant.

Error correction refers to a model's ability to eliminate forecast errors. A shock to an economic system shifts its equilibrium and an error-correction model, such as its most extreme form, the random walk, adjusts quickly, while the equilibrium-correction model (EqCM) attempts to return to the old equilibrium. Early specifications of equilibrium-correction models imposed parameter restrictions that could be expressed as unit root restrictions. Of course, when one starts with a restricted model it can be difficult to see what the more general case might be, so it was some time before it was realized that any VAR in levels can be rewritten into general equilibrium-correction form (best described in chapter 2 of Banerjee, Dolado, Galbraith, & Hendry, 1993).

According to Clive Granger (Granger, 2004), awareness of the relationship between error correction and cointegration (the term was coined by Granger) came about when David Hendry stated that the difference between a pair of integrated series could be stationary and Granger argued that it could not. In setting out to prove Hendry wrong, Granger came to the conclusion that he was in fact correct, and the rest, as they say, is history (Granger, 2004, p. 423). Granger's (1981) paper is the seminal article.

It was probably no accident that the original unit root test was published about this time (Dickey & Fuller, 1979). Cointegration first requires the variables to possess unit roots, then requires parameter restrictions — unit root restrictions needed for the linear combination of variables to be stationary. Initially Engle and Granger (1987) introduced the restrictions via a second equation, the cointegrating equation (only one with a two-variable system). Johansen (1988) introduced the generalization of the unit root test to a system of autoregression equations.

Where tests show the unit root restrictions to be valid, we would expect EqCMs to give better forecasts than VARs in levels. In practice this has not clearly been the case, as Table 2 illustrates.

Focusing discussion on variables that are integrated of order one, I(1), when there are no cointegrating vectors, unit root restrictions can be imposed on all variables and the VAR in first differences is the correct specification. The first panel of Table 2 shows that this specification (in bold) is more accurate than an EqCM, as we would expect. When there is a cointegrating vector, imposing that restriction does not generally lead to more accurate forecasts, though imposing more inappropriate unit root restrictions and estimating a differenced VAR is not a good strategy. There are too few comparisons when multiple cointegrating vectors are detected to make much of the comparison.

These results are a little surprising. When there is a shift in equilibrium, an EqCM tends to return to the old equilibrium while a VAR with the variables in first differences adjusts quickly and is in that sense robust to change, even though it is misspecified. The VAR which is less parsimonious in levels should behave as the EqCM but seems to do no worse or better. And the robustness of the differenced VAR in response to structural changes does not seem to confer any advantage.

As one stopping point on the progression from general-to-specific models, equilibrium-correction

Table 2
Pairwise comparison of estimating different vector autoregression models, by out-of-sample forecast RMSE, lead times not specified

Methods	First method best	Second method best	Total series
No cointegrating vectors detected			
EqCM vs VAR levels	5	12	17
EqCM vs VAR differenced	5	**8**	13
One cointegrating vector detected			
EqCM vs VAR levels	**11**	12	25**
EqCM vs VAR differenced	**20**	4	24
Two cointegrating vectors detected			
EqCM vs VAR levels	**0**	1	1
EqCM vs VAR differenced	**1**	0	1
Three cointegrating vectors detected			
EqCM vs VAR levels	**3**	3	7*
EqCM vs VAR differenced	3	2	7**
Four cointegrating vectors detected			
EqCM vs VAR levels	**4**	0	4

Number of series for which the first method is more accurate than the second method.
The number of asterisks (*) shows the number of series tied, with equal RMSE for each method.
Source: Chapter 11 of Forecasting Principles. For details of sources see Appendix Table A1 of that chapter, found on the http://forecastingprinciples.com website, or directly at http://fourps.wharton.upenn.edu/forecast/paperpdf/principles%20appendix.pdf.

models are likely to stay. Alternatively, if theory assumes a more dominant role in model specification than at present, such models may be derived from theory-based cointegrating restrictions. Their success will depend on establishing the conditions under which the parameter restrictions lead to better forecasts. Although econometric theory confidently asserts that imposing true parameter restrictions improves efficiency and should lead to more accurate forecasts, analyses of real series do not confirm this expectation. Either increased efficiency does not lead to better forecasts, or, more likely, tests supporting parameter restrictions are not sufficiently powerful.

8. Autoregressive conditional heteroscedasticity

No one source provides a more understandable description of autoregressive conditional heteroscedasticity (ARCH) than Engle's recent *American Economic Review* article (Engle, 2004). It is refreshing in that it provides historical perspective on the issues that led to its development, on its metamorphosis into more generalized models over time, and especially on the players involved in its development and their contributions. Its basic premise is that many economic time series, particularly financial series, exhibit periods of high volatility followed by periods of relative calm. Consequently, the assumption of constant variance is violated.

As partial background, Engle (2004) points out that Granger had an interest in bilinear time-series models. That is, he observed that squared residuals from a time-series model were frequently autocorrelated although the residuals themselves were not. Out of such bilinear models a test using squared residuals was developed. Discussions at the London School of Economics in 1979 with Sargan, Durbin, and Hendry (who is credited with inventing the name) led Engle to the original specification. Frank Srba wrote the first ARCH program, and Engle's (1982) article in *Econometrica* is the first published reference.

Tim Bollerslev, Engle's student, generalized the model to GARCH in a 1986 *Journal of Econometrics* article (Bollerslev, 1986). It is the second most-cited paper in that journal (see http://www1.elsevier.com/homepage/sae/econworld/econbase/econom/econom_starpapers.htm). The review paper of Bollerslev, Chou, and Kroner (1992) on ARCH is the third most cited.

The GARCH(1,1) model has become "the workhorse of financial applications" (Engle, 2004, p. 408) when describing volatility dynamics. A whole alphabet soup of generalizations of the original ARCH model has been proposed, and this work dominates the financial forecasting literature. ARCH has made much less headway in other applications of econometric forecasting.

Covariance and correlation forecasting is an important aspect of forecasting when dealing with high-frequency data. The PIER working paper by Andersen, Bollerslev, Christoffersen and Diebold (2005) is a compendium of available techniques. It is thorough in its treatment of theoretical and applied issues. The authors discuss volatility modeling and forecasting from the perspective of GARCH volatility, stochastic volatility, realized volatility, and multivariate volatility paradigms. The realized volatility approach has been both useful and successful when it comes to covariance and correlation forecasting, as is illustrated by the studies reviewed by Poon and Granger (2003). On page 506 they make forecasting-performance comparisons among the various volatility approaches. For example, in 39 studies comparing the forecast performance between historical volatility models (HISVOL) and GARCH, 22 of these studies (or 56 % of the total) found HISVOL to be better at forecasting than GARCH, and 17 of the studies (or 44 % of the total) found GARCH to be better at forecasting than HISVOL.

In the future, work can be expected to continue in two areas: understanding the determinants of volatility persistence, and extracting more information about volatility to improve forecast accuracy. The same problems that arise in forecasting means also arise in forecasting variances. Poon and Granger (2003) cite the work of Lamoureux and Lastrapes (1990) and Gallant, Rossi, and Tauchen (1993) in exploring various ARCH processes as providing a promising framework for analyzing volatility persistence. Understanding both the historical and implied causes of volatility should help improve time-series methods. Future research could also be directed to issues of forecast evaluation and combining forecasts.

9. The future

If we start with the technological and institutional future, two things are easy to forecast. First, computational ability and the software to go with it will both become more powerful. Techniques and analyses that we can only dream of today will become feasible, and once feasible will be used. Second, the amount of data, its length and level of disaggregation, will also increase. Taken together, these two components of the future indicate that data mining based on large data sets will become more common. In reference to the use of VAR in macroeconomic forecasting, Diebold makes the following observation (1998, p. 183): "The future of nonstructural economic forecasting will be more of the same — steady progress — fueled by cheap, fast computing, massive storage, and increased sophistication of numerical and simulation techniques."

Granger (2001), in writing about macroeconomics, observed that its practitioners operated at one extreme or the other: employing either rather small vector models using vector autoregressions or error-correction specifications at one extreme, or very large models involving hundreds or thousands of equations at the other. The real future, he believes, should lie somewhere between these extremes. Diebold (1998) agrees; he expects models to grow from three or four variables to eight or ten, but for the expansion to stop there, first because bigger is not necessarily better, and second because models that require system estimation methods are still limited by computing power.

Some things will not go away, however. Model simplification (Zellner's KISS principle referred to earlier) holds a powerful appeal and is likely to remain a key element. Clements and Hendry (2005, p. 717) refer to the benefits of parsimony in reducing estimation uncertainty and susceptibility to breaks as "a well-known claim in the folklore of economic forecasting," and would certainly not elevate it to a principle. They note later in the same page that this claim confounds simplicity with robustness. While a simple model offers the benefit that the forecast will be easier to interpret, a model that is robust to structural breaks is likely to provide the better forecast. A variable reduction strategy will not identify economic structure without a good deal of help from other sources. One question is whether economic theory will be developed in a way that can provide that help. Three decades ago, theory dominated. Commentators at that time noted that economic theory was being pushed too hard by econometricians: it could answer qualitative, directional questions but gave little insight into either the quantitative size of the change or the short-term dynamics. How much theory (and which theory) should be used? Should the theory provide a background to the model that is to be embedded in it, or constrain it?

An alternative approach to using model shrinkage is to use Bayesian methods. Although Bayesian econometrics has been around for some fifty years, it has become the preferred philosophy of only a small minority of econometricians. It was used by Sims to estimate his original VAR models, yet has failed to make inroads on approaches used by classical econometricians. There is no sign that it will do so in the future.

In sharp contrast, work on testing is likely to explode in the next decade. Some might argue that it has already started. The history of forecasting is marked by pragmatism: use the method or methods that work. This was the incentive for the large-scale forecasting competitions organized by Makridakis et al. (1982). Out of those competitions emerged the standard practice of comparing the accuracy of forecasts for the period outside the sample used to fit the model. This practice carried over into the multivariate arena, though Hendry and Clements (2003) have recently pointed out that it is not the most useful criterion for model selection, especially if the model is to be used in policy analysis.

In the general-to-specific modeling strategy advocated by Hendry (summarized by Pagan (1987), who regards the evaluation and testing of the resulting model as differentiating Hendry from his predecessors), the first layer of tests should be misspecification or diagnostic tests on residuals. A failed misspecification test is a rather negative thing. It says only that the model as estimated is an inadequate summary of the data. Since all model specifications are simplifications of reality, there is a temptation to press on in spite of the warning signals.

One problem is that the number of tests available for a particular form of misspecification, for example, departure from normality, is large and growing. By applying all available tests there is a good chance that the model will fail one of them at the chosen level of significance. Applying additional tests for heteroscedasticity, autocorrelation and parameter constancy almost guarantee it. Several test procedures make assumptions about structure, for example, some parameter constancy tests assume homoscedasticity, which may be undermined by other tests. We need answers to questions: In forecasting, which misspecifications matter? Which tests are sufficiently robust to other misspecifications to be relied on? How “independent” are the individual tests from each other? What sequence and how many tests should be applied? And what level of significance should be employed? Out of these issues, some standard practice is likely to emerge in the next decade.

There is a general sense that, given a constant-parameter model, structural breaks, regime changes, and outliers are the most serious causes of misspecification. How do we best detect them? What do we do once they are detected? The answer to the second question is, of course, to make the model more complicated. Possibilities for complication are: add more explanatory variables, move to functional forms that are nonlinear in parameters, or recognize that the structure of the model is changing over time. Once the more complicated model is found, we move to the second issue, specification testing, where the result of the test does point to a particular model. Although new specification tests are being reported in the literature, this area of testing is much more settled than misspecification testing.

About nonlinear-in-parameters functional forms there is little to be said, except that we can expect these to become more common. A generic test for nonlinearity would be useful. A much wider set of models becomes available and estimation of them becomes practical with increased computing power. Even with the greater flexibility afforded compared with linear models, the problem of structural breaks still exists, with the same consequences for forecast performance. There is a danger of asking more of the data than they are capable of delivering, and the need to enforce the KISS principle is strong.

In the context of linear-in-parameters models, are time-varying parameters the solution? Here also a much larger set of models becomes available, and includes systematically varying parameters, stationary and non-stationary stochastic parameters. Some of these have a long history and are hardly regarded as varying parameters, for example, the use of dummy variables. Along similar lines, parameters are made functions of variables, whether already in the model or not. Sometimes these are qualitative and determine a known or unknown join point of two fixed-parameter regressions, and sometimes the join point is determined stochastically, as with Markov switching. The stationary random parameter models developed by Hildreth and Houck (1968) and by Swamy (1970) have a long history. Later, Harvey and Phillips (1982) generalized the Hildreth and Houck specification in a way that led naturally to a state-space formulation. In contrast, the approach by Cooley and Prescott (1973) treated parameters as random walks.

Stochastically varying parameter approaches have not come into general use and the forces against them are strong, represented by econometricians dedicated to improving fixed-parameter approaches. Swamy argues that their dedication (exposited by econometrics textbooks) to improve fixed-parameter approaches has failed miserably (a recent statement of his views is in Swamy, Yaghi, Mehta, & Chang, in press). The biasing effects of omitted variables, measurement errors, and incorrect functional forms are a pervasive problem in applied statistics. Swamy argues from a "true" model with time-varying coefficients (TVC), which is "true" if all economic variables are included and are correctly measured, and if coefficients can take on different values over time and across individuals. Even this model still seems open to the charge of misspecified functional form unless it is viewed as a Taylor's series. What is clear is that an operational TVC model derived from the "true" model gives more accurate out-of-sample forecasts than a fixed-parameter random-effects model of the same variables. This result may well always hold in a large-sample sense. What matters in practice are the conditions under which varying parameter approaches will dominate. As long as mainstream econometricians continue to put effort into improving the specifications of fixed-parameter models, they will make that dominance a greater challenge.

Recently, developments of automatic model selection have become feasible with increases in computing power. Workers in this area include Hendry and Krolzig (2001), Hoover and Perez (1999), Krolzig and Hendry (2001) and Phillips (1992, 1995). Their algorithms rely on automated significance tests in conjunction with model selection rules for dealing with rival specifications that are unresolved by significance testing, and have an impressive ability to sort through thousands of members of a model class to discover in seconds a parsimonious model that previously took experienced econometricians weeks of effort. Most work to date has been with single equation models, though Krolzig (2003) provides a multivariate extension. Phillips (2003) proposes a scheme to offer automatic econometric modeling methods to a wider community of users by means of the Internet, providing unsophisticated users with access to best-practice econometric techniques and econometric software. Just as the driver of a modern automobile needs to know little about the machinery that propels it, so might a community of forecasters know little about the machinery that generates their results. Will this approach become the mainstream of econometric forecasting or just an interesting and useful tool in the forecaster's toolkit?

Finally, we can expect density forecasting to take on a more substantial role than it has at present. The point forecast conveys no information about uncertainty, the standard construction of a prediction interval conveys some, and the density forecast, or estimate of the complete probability distribution of the variable of interest, conveys the most. Granger (2001) notes that as more data become available, forecasts will not be made just for conditional means and variances, but for the whole predictive distribution. Any user can then find any forecast required using any given cost function. This line of work will gain momentum when the role of the forecaster in decision making becomes more clearly defined. To date, almost all density forecasts have been for macroeconomic and finance variables. If relative page counts in the Tay and Wallis (2000) survey are a guide, financial forecasters are currently ahead, perhaps because users' desires to make a profit and to control risk provide a clear loss function, in contrast with government policymakers. Investors' need for risk assessment (measured, for example, by Value-at-Risk, with its unfortunately confusing contraction of VaR) is another motivation. Tay and Wallis regard the ARCH modeling work as the beginning of density forecasting and point out that higher moments of the distribution, which matter in density forecasting, are quite inflexible in the array of models developed from the original ARCH specification.

Future developments of density forecasting are likely in areas of evaluation or calibration (Tay & Wallis, 2000). Questions of how closely the forecast distribution from one method matches the true probability distribution of the variable compared with the forecast distribution of another method raise complex questions. Issues of combining distributions and evaluating multi-step-ahead density forecasts raise more. More computing power is definitely needed. Paying more attention to the presentation of the forecast and to the relationship of the forecast to the user (the user's loss function) are desirable. More power! More care! Perhaps these are the future for all parts of econometric forecasting.

References

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2005). Volatility forecasting. *Penn Institute for Economic Research (PIER) Working Paper 05-011*. Department of Economics, University of Pennsylvania.

Armstrong, J. S. (Ed.). (2001). *Principles of forecasting: A handbook for researchers and practitioners*. Norwell MA: Kluwer Academic Publishers.

Banerjee, A., Dolado, J. J., Galbraith, J. W., & Hendry, D. F. (1993). *Co-integration, error-correction, and the econometric analysis of non-stationary data*. Oxford: Oxford University Press.

Blaug, M. (1992). *The methodology of economics, or how economists explain*. Cambridge: Cambridge University Press.

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

Bollerslev, T., Chou, R., & Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, *52*, 5–59.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis. Forecasting and control*. San Francisco: Holden-Day.

Burns, A. F., & Mitchell, W. C. (1946). Measuring business cycles. *Studies in business cycles*, *vol. 2*. . National Bureau of Economic Research.

Christ, C. F. (1994). The Cowles Commission's contributions to econometrics at Chicago, 1939–1955. *Journal of Economic Literature*, *32*, 30–59.

Clements, M. P., & Hendry, D. F. (1996). Intercept corrections and structural change. *Journal of Applied Econometrics*, *11*, 475–494.

Clements, M. P., & Hendry, D. F. (1998). *Forecasting economic time series*. Cambridge: Cambridge University Press.

Clements, M. P., & Hendry, D. F. (2005). 'Guest editors' introduction: Information in economic forecasting. *Oxford Bulletin of Economics and Statistics*, *67*, 713–753.

Clements, M. P., & Krolzig, H. M. (1998). A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP. *Econometrics Journal*, *1*, C47–C75.

Cooley, T., & Prescott, E. (1973). Varying parameter regression, a theory and some applications. *Annals of Social and Economic Measurement*, *2*, 463–474.

Cullity, J. P. (1993). Monitoring business conditions at the CIBCR [Center for International Business Cycle Research at Columbia University]. *International Journal of Forecasting*, *9*, 49–60.

Davidson, J. E. H., Hendry, D. F., Srba, F., & Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, *88*, 661–692.

Del Negro, M., Schorfheide, F., Smets, F., & Wouters, R. (2004). On the fit and forecasting performance of new Keynesian models. *Federal Reserve Bank of Atlanta Working Paper 2004-37*.

Dhrymes, P. J., & Thomakos, D. D. (1998). Structural VAR, MARMA and open economy models. *International Journal of Forecasting*, *14*, 187–198.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*, 427–431.

Diebold, F. X. (1998). The past, present, and future of macroeconomic forecasting. *Journal of Economic Perspectives*, *12*(2), 175–192.

Diebold, F. X. (2001). Econometrics: Retrospect and prospect. *Journal of Econometrics*, *100*, 73–75.

Diebold, F. X., & Rudebusch, G. D. (1996). Measuring business cycles: A modern perspective. *Review of Economics and Statistics*, *78*, 67–77.

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

Engle, R. F. (2004). Risk and volatility: Econometric models and financial practice. *The American Economic Review*, *94*, 405–420.

Engle, R. F., & Granger, C. W. J. (1987). Cointegration and error-correction: Representation, estimation, and testing. *Econometrica*, *55*, 251–276.

Epstein, R. J. (1987). A history of econometrics. *Contributions to economic analysis series*, *vol. 165*. Amsterdam: North-Holland.

Fair, R. C. (1984). *Specification, estimation, and analysis of macroeconometric models*. Cambridge, MA: Harvard University Press.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics*, *12*, 299–308.

Gallant, A. R., Rossi, P. E., & Tauchen, G. (1993). Nonlinear dynamic structures. *Econometrica*, *61*, 871–907.

Geweke, J. (1977). The dynamic factor analysis of time series. In D. J. Aigner, & A. S. Goldberger (Eds.), *Latent variables in socioeconomic models, chapter 19*. Amsterdam: North Holland.

Gordon, D. V., & Kerr, W. A. (1997). Was the Babson prize deserved? An enquiry into an early forecasting model. *Economic Modelling*, *14*, 417–433.

Granger, C. W. J. (1981). Some properties of time series data in econometric model specification. *Journal of Econometrics*, *16*, 121–130.

Granger, C. W. J. (2001). Macroeconometrics—past and future. *Journal of Econometrics*, *100*, 17–19.

Granger, C. W. J. (2004). Time series analysis, cointegration, and applications. *American Economic Review*, *94*, 421–425.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, *11*, 1–12.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, *12*, 1–115.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*, 357–384.

Harvey, A., & Phillips, G. (1982). Estimation of regression models with time varying parameters. *Proceedings of a Symposium in Honour of Oskar Morgenstern, Vienna, May 1980*. Wurzburg: Physica-Verlag.

Hendry, D. F. (1977). Comments on Granger–Newbold's 'Time series approach to econometric model building' and Sargent–

Sims' 'Business cycle modeling without pretending to have too much a priori economic theory'. In C. A. Sims (Ed.), *New methods in business cycle research: Proceedings from a conference* (pp. 183–202). Minneapolis: Federal Reserve Bank of Minneapolis.

Hendry, D. F., & Clements, M. P. (2003). Economic forecasting: Some lessons from recent research. *Economic Modelling*, *20*, 301–328.

Hendry, D. F., & Krolzig, H.-M. (2001). *Automatic econometric model selection*. London: Timberlake Consultants Press.

Hendry, D. F., & Richard, J.-F. (1982). On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics*, *20*, 3–33.

Hendry, D. F., & von Ungern-Sternberg, T. (1981). Liquidity and inflation effects on consumers' expenditure. In A. S. Deaton (Ed.), *Essays in the theory and measurement of consumer behaviour: In honour of Sir Richard Stone* (pp. 237–260). Cambridge: Cambridge University Press.

Hildreth, C., & Houck, J. P. (1968). Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association*, *63*, 584–595.

Hoover, K. D., & Perez, S. J. (1999). Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, *2*, 167–191.

Hsiao, C. (1979). Autoregressive modeling of Canadian money and income data. *Journal of the American Statistical Association*, *74*, 553–560.

Ireland, P. N. (2004). A method for taking models to the data. *Journal of Economic Dynamics & Control*, *28*, 1205–1226.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics & Control*, *12*, 231–254.

Klein, L. R. (1950). Economic fluctuations in the United States 1921–1941. *Cowles commission monograph*, *vol. 11*. New York: Wiley.

Klein, P. A., & Moore, G. H. (1982). The leading indicator approach to economic forecasting: Retrospect and prospect. *NBER Working Paper No. 0941*. National Bureau of Economic Research, Inc.

Koopmans, T. C. (1947). Measurement without theory. *Review of Economics and Statistics*, *29*, 161–179.

Krolzig, H.-M. (2003). General-to-specific model selection procedures for structural vector autoregressions. *Oxford Bulletin of Economics and Statistics, Supplement*, *65*, 769–801.

Krolzig, H.-M., & Hendry, D. F. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics & Control*, *25*, 831–866.

Kydland, F. E., & Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica*, *50*, 1345–1370.

Lahiri, K., & Moore, G. H. (Eds.). (1991). *Leading economic indicators: New approaches and forecasting records*. Cambridge (UK): Cambridge University Press.

Lahiri, K., & Wang, J. G. (1994). Predicting cyclical turning points with a leading index in a Markov switching model. *Journal of Forecasting*, *13*, 245–264.

Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence in variance, structural change and the GARCH model. *Journal of Business & Economic Statistics*, *8*, 225–235.

Layton, A. P., & Katsuura, M. (2001). Comparison of regime switching, probit and logit models in dating and forecasting US business cycles. *International Journal of Forecasting*, *17*, 403–417.

Leamer, E. F. (1973). Multicollinearity: A Bayesian interpretation. *Review of Economics and Statistics*, *55*, 271–280.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions: Five years of experience. *Journal of Business & Economic Statistics*, *4*, 25–38.

Liu, T. C. (1960). Underidentification, structural estimation, and forecasting. *Econometrica*, *28*, 855–865.

Lucas, R. E. Jr. (1976). Econometric policy evaluation: A critique. In K. Brunner, & A. H. Meltzer (Eds.), *The Phillips curve and labor markets*. *Carnegie-Rochester conference series on public policy* (pp. 19–46). Amsterdam: North Holland.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., & Lewandowski, R., et al. (1982). The accuracy of extrapolation (time-series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*, 111–153.

Mann, H. B., & Wald, A. (1943). On the statistical treatment of linear difference equations. *Econometrica*, *11*, 173–220.

Marcellino, M. (2005). Leading indicators: What have we learned? *CEPR Discussion Paper*, *vol. 4977*.

Megna, R., & Xu, Q. (2003). Forecasting the New York State economy: The coincident and leading indicators approach. *International Journal of Forecasting*, *19*, 701–713.

Mitchell, W. C., & Burns, A. F. (1938). Statistical indicators of cyclical revivals. *National Bureau of Economic Research, Bulletin*, *vol. 69* (pp. 69).

Nerlove, M., & Addison, W. (1958). Statistical estimation of long-run elasticities of supply and demand. *Journal of Farm Economics*, 861–880.

Pagan, A. (1987). Three econometric methodologies: A critical appraisal. *Journal of Economic Surveys*, *1*, 3–24.

Pearson, K. (1896). Mathematical contributions to the theory of evolution: III. Regression, Heredity, and Panmixia. *Philosophical Transactions of the Royal Society of London*, *187*, 253–318.

Phillips, A. W. (1954). Stabilisation policy in a closed economy. *Economic Journal*, *64*(254), 290–323.

Phillips, A. W. (1957). Stabilisation policy and the time-forms of lagged responses. *Economic Journal*, *67*(266), 265–277.

Phillips, P. C. B. (1992). Bayes methods for trending multiple time series with an empirical application to the US economy. *Cowles Foundation Discussion Paper*, *vol. 1025*.

Phillips, P. C. B. (1995). Bayesian model selection and prediction with empirical applications. *Journal of Econometrics*, *69*, 289–331.

Phillips, P. C. B. (2003). Laws and limits of econometrics. *Economic Journal*, *113*, C26–C52.

Poon, S., & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, *41*, 478–539.

Quenouille, M. H. (1957). *The analysis of multiple time series*. London: Griffin.

Rudebusch, Glenn D. (2005). Assessing the Lucas critique in monetary policy models. *Journal of Money, Credit and Banking*, *37*, 245–272.

Sargan, J.D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology. In P.E. Hart, G. Mills, J.K. Whittaker (Eds.), *Econometric analysis for National Economic Planning*. Butterworths, London. Reprinted. In D.F. Hendry, & K.F. Wallis (eds.), *Econometrics and Quantitative Econometrics*. Blackwell, Oxford, 1984.

Sargent, T. J., & Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. In C. A. Sims (Ed.), *New methods in business cycle research: Proceedings from a conference*. Minneapolis: Federal Reserve Bank of Minneapolis.

Sarle, C. F. (1925). The forecasting of the price of hogs. *American Economic Review*, *15, Number 3*(Supplement Number 2), 1–22.

Shiller, R. L. (1973). A distributed lag estimator derived from smoothness priors. *Econometrica*, *41*, 775–788.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*(1), 1–48.

Sims, C. A. (1993). A nine-variable probabilistic macroeconomic forecasting model. In J. H. Stock, & M. W. Watson (Eds.), *NBER studies in business: Business cycles, indicators, and forecasting*, *vol. 28*. (pp. 179–214). Chicago: University of Chicago Press.

Stock, J. H. (2001). Macro-econometrics. *Journal of Econometrics*, *100*, 29–32.

Stock, J. H., & Watson, M. W. (1991). A probability model of the coincident indicators. In K. Lahiri, & G. H. Moore (Eds.), *Leading economic indicators: New approaches and forecasting records* (pp. 63–90). Cambridge (UK): Cambridge University Press.

Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, *15*(4), 101–115.

Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, *38*, 311–323.

Swamy, P. A. V. B., Yaghi, W., Mehta, J. S., & Chang, I. -L. (in press). Empirical best linear unbiased prediction in misspecified and improved panel data models, with an application to gasoline demand. *Computational Statistics and Data Analysis*.

Tay, A. S., & Wallis, K. F. (2000). Density forecasting: A survey. *Journal of Forecasting*, *19*, 235–254.

Webb, R. H. (1991). On predicting the stage of the business cycle. In K. Lahiri, & G. H. Moore (Eds.), *Leading economic indicators: New approaches and forecasting records* (pp. 109–127). Cambridge (UK): Cambridge University Press.

Zellner, A. (1992). Statistics, science and public policy. *Journal of the American Statistical Association*, *87*, 1–6.

Zellner, A. (2001). Comments on papers by Engle, Geweke and Granger. *Journal of Econometrics*, *100*, 93–94.

Further reading

McCloskey, D. (2002). *The secret sins of economics*. Chicago: Prickly-Paradigm Press.

Pagan, A. R., & Robertson, J. C. (1998). Structural models of the liquidity effect. *Review of Economics and Statistics*, *80*, 202–217.