Econometric Theory, **16**, 2000, 441–450. Printed in the United States of America.

BOOK REVIEW

FORECASTING ECONOMIC TIME SERIES

by Michael P. Clements and David F. Hendry

Cambridge University Press, 1998

REVIEWED BY
FRANK SCHORFHEIDE
University of Pennsylvania

1. INTRODUCTION AND OVERVIEW

The prediction of future events and developments is an exciting and perhaps mysterious task, often associated with the aura of prophets and seers instead of probabilistic models and computer screens. The reality of macroeconomic forecasting, however, is quite mundane. Predictions of macroeconomic aggregates play an important role in the decision making of private enterprises, central banks, and governments. In general, forecasts become less popular if they turn out to be inaccurate ex post, and the postwar history of macroeconomic forecasting has had its share of disappointments. For instance, in the early 1980's, economists tested inflation forecasts taken over the previous 20 years and found that the forecasts were poor, partly as a result of the oil price shocks in the 1970's. A recent study (Croushore, 1998) with data up to 1996 provides a more favorable assessment of the quality of inflation forecasts.

The practice of macroeconomic forecasting is an often not fully transparent mixture of number crunching and judgmental adjustments. The econometric theory, on the other hand, is very clean, almost sterile: postulate a probability distribution for future observations, take a loss function, compute the prediction that minimizes the expected forecast loss under the entertained probability model, maybe replace some unknown parameters by estimates or integrate out the parameters with respect to a posterior distribution. End of theory.

Of course, I am oversimplifying. Both practice and theory of forecasting are richer than my sketch. But there is a gap. Conditional on a probability model, econometricians have a well-equipped tool box that allows them to generate predictions. Unfortunately, there are many competing probability models in practice. It is often not at all obvious which one to use, how to use it, or whether to consider several models simultaneously. This leads to a variety of apparently

Address correspondence to: Frank Schorfheide, Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297; e-mail: schorf@ssc.upenn.edu, URL http://www.econ.upenn.edu/~schorf.

© 2000 Cambridge University Press 0266-4666/00 $9.50

https://doi.org/10.1017/S0266466600003066 Published online by Cambridge University Press

heuristic adjustments to the theoretical procedure: add-factors are tacked onto model predictions, forecasts from different models are combined, forecast loss functions are used to compute parameter estimates.

The recent book by Michael Clements and David Hendry, *Forecasting Economic Time Series*, is written to bridge the gap between practice and theory. Much of the analysis in the book is based on the assumption that the forecasting model is potentially misspecified. The objective of the monograph is "to provide a formal analysis of the models, procedures and measures of economic forecasting with a view to improving forecasting practice." The book shares its title with a monograph by Clive Granger and Paul Newbold, published in a second edition in 1986. It revisits many of the fundamental themes in theoretical and applied forecasting without reviewing the basic theory of time series. The perspective is of course a different one, taking many advances of the past decade into account.

The subject matter is complex and colorful, especially if it is tailored toward applied aspects. Consequently, it cannot be exhaustively covered in a 370-page monograph. As in any book, some hard choices with respect to selection and emphasis of the material have to be made. These choices are naturally subject to the personal taste of the authors, two prominent scholars in the field of macroeconomic forecasting. Although the overall coverage of the topic and the literature is impressive, the exposition is centered around the research agenda of the authors. The monograph draws extensively on papers previously published by Michael Clements, David Hendry, and their co-authors. It is assumed that the reader has an econometric theory background at the level of Hamilton (1994) or Hendry (1995). The book focuses on, but is not limited to, forecasting cointegrated linear vector processes, which are potentially subject to structural breaks. The conditional expectation is the workhorse predictor. The following topics are less emphasized: explicit derivation of optimal predictors for a variety of linear models, such as ARIMA and state space models; interval forecasts and density forecasts; forecasting under nonquadratic loss functions; Bayesian forecasting models for vector autoregressive processes; fractionally integrated models; prediction of conditional heterogeneity. However, the book provides many references related to these subjects.

The book is organized in thirteen chapters and a postscript. Each chapter is preceded by an abstract and followed by a summary. This structure makes it easy for the reader to obtain an overview of the main issues and to navigate back and forth through the monograph. The red thread that connects the different parts of the book is the notion that the forecast model might ex post not have provided an adequate probability distribution for the data, or in Clements and Hendry's terms: a potential discrepancy between the data generating process (DGP) and the forecast model. Let me illustrate this point by a simple example that by no means encompasses all the different cases analyzed in the book but highlights the important dimensions of the problem. At time T, the forecaster has to predict the observation y_{T+1} under quadratic forecast error loss,

https://doi.org/10.1017/S0266466600003066 Published online by Cambridge University Press

based on observations $y_1, \ldots, y_T$. Suppose that the forecast is computed according to $\hat{y}_{T+1|T} = \phi_0 + \phi y_T$. The predictor $\hat{y}_{T+1|T}$ corresponds to the conditional expectation of y_T under an AR(1) model with parameters ϕ_0 and ϕ_1 that might be replaced by some estimates. What are the properties of the forecast $\hat{y}_{T+1|T}$ if $\mathbb{E}_T[y_{T+1}] \neq \phi_0 + \phi_1 y_T$?

Suppose that between period $T-1$ and T, or between T and $T+1$, the coefficient ϕ_0 changes to ϕ_0^*. This can be interpreted as a structural break, an example of the inherent nonconstancy of the DGP in the language of the authors. Throughout the chapters of the book, the reader learns a lot about the calculation of prediction error losses for $y_{T+1|T}$ in the presence of a discrepancy between the forecast model and the DGP. The authors analyze in great detail how deviations from the "pseudo-optimal" predictor $\hat{y}_{T+1|T}$ can actually improve forecasts. Intercept corrections may help, or the combination of conditional and unconditional mean predictions can lead to superior forecasts. Simple forecasting rules, such as "no change" forecasts, are surprisingly robust in the presence of structural changes, in particular if the change occurred in the very recent past.

Unfortunately, the theory set forth in the book does not really address the question of how a forecaster can determine in period T whether to use the predictor $\hat{y}_{T+1|T}$ directly, or to calculate an intercept correction, or to switch to a "no change" forecast. If a forecaster is worried about a structural break during or immediately before the forecasting period the forecaster could, for instance, choose a predictor according to minimax considerations over a reasonable set of possible parameter changes, or the forecaster could place probability over the possible parameter changes and choose the predictor that minimizes the conditional expected forecast error loss. The specification and estimation of a constant coefficient model precludes the forecaster from learning about the probability of parameter nonconstancy through the data $y_1, \ldots, y_T$. Hence, the possibility of a parameter change can only be contemplated from an a priori perspective or based on an information set that is larger than the one used by the probability model. Alternatively, the forecaster could rewrite the probability model as a time varying parameter model and try to learn from the occurrence of past changes about the likelihood of future changes. The latter strategy is not explored in detail, except for a short excursion to self-exciting threshold autoregressive models and Markov-switching models.

The book gives the impression that the authors are convinced that regardless how hard one tries to model the past, there is always the possibility that the model breaks down between yesterday and tomorrow. While this proposition is correct in a trivial sense, the possibility is in practice small enough to turn the decision whether to use $\hat{y}_{T+1|T}$, an intercept corrected version of it, or simply the "no change" predictor, into an important and interesting problem. For some work along these lines see Schorfheide (1998). I enjoyed reading the analysis of the consequences on the performance of predictors under various forms of discrepancies between the forecasting model and what the authors call the data

https://doi.org/10.1017/S0266466600003066 Published online by Cambridge University Press

generating process. I was a bit disappointed that the authors did not attempt to systematically analyze the problem of how a forecaster at time T can make a good choice among these predictors based on the available information $y_1, \ldots, y_T$.

2. A CLOSER LOOK AT THE CONTENTS

The first chapter of the book begins with a brief review of the history of econometric forecasting. Clements and Hendry discuss a variety of forecasting methods such as guessing, extrapolation, leading indicators, surveys, time series models, and econometric systems. Chapter 2 lays the groundwork for the theory that is developed subsequently. The chapter starts with the fundamental result that the conditional expectation of y_{T+h} given time T information minimizes the expected forecast error loss, if the loss function is quadratic. The forecasting problem is not explicitly treated in a decision theoretic framework. The book focuses on expected quadratic prediction error losses because the authors believe that context-specific loss functions are rarely available to macroeconomic forecasters. Brief reviews of alternative loss functions and the corresponding optimal predictors appear in later chapters.

The authors continue with the definition of the concepts "predictability" and "forecastability." A process is unpredictable conditional on an information set, if its conditional distribution is the same as its unconditional distribution. The notion of forecastability is less clear cut. A weakly stationary process is regarded as unforecastable at horizon h if the (frequentist) expected loss of a forecast that is made without conditioning information is only ϵ smaller than the (posterior) expected loss of a forecast that is made based on the conditioning information. It seems difficult to make such statements precisely without adopting a decision theoretic approach and a careful distinction between posterior and frequentist risk. Clements and Hendry's concept of forecastability is based on two different notions of risk, which are only identical in a special case where the posterior risk does not depend on the conditioning information.

The authors discuss several implications of predictability and forecastability. For instance, intertemporal transforms of a process affect its predictability: a random walk is predictable, but its first differences are not. Therefore, no unique measure of predictability and forecasting accuracy exists. I did not find this interpretation very helpful. Clements and Hendry fix a quadratic loss function, $L(\hat{y}_{T+h}, y_{T+h}) = (\hat{y}_{T+h} - y_{T+h})^2$, and consider transformations of the data under the same loss function: $L(\Delta\hat{y}_{T+h}, \Delta y_{T+h})$. Why not regard the latter as state dependent loss function $L_T^*(\hat{y}_{T+h}, y_{T+h})$ and argue that rankings of forecasts and prediction procedures are, in general, sensitive to the choice of loss function? The subsequent section of Chapter 2 reviews forecasting with ARIMA models and provides a brief overview on forecasting in a multivariate framework. The chapter ends with an analysis of the role of causal information. The discussion provides a first flavor of the consequences of misspecification. In an AR(1) model with intercept the lagged dependent variable is "causal." Includ-

https://doi.org/10.1017/S0266466600003066 Published online by Cambridge University Press

ing it in a forecast model will reduce the forecast error loss. However, if the underlying process is subject to change, the unconditional mean forecast that ignores the causal information is potentially preferable.

Chapter 3 considers the evaluation of point forecasts. A natural starting point is the computation of average forecast error losses as an approximation to the frequentist risk. Under the adopted quadratic loss function this is simply the mean square forecast error (MSFE). Clements and Hendry discuss work on testing rationality of forecasts, fixed event forecasts, and various MSFE measures. The main argument set forth in Chapter 3 is a criticism of the lack of invariance of MSFE measures to affine data transformations. By switching from levels to differences, the ranking of predictors can change. Analytical calculations for univariate and multivariate models are used to illustrate this point. However, it is important to note that under suitable regularity conditions, the conditional expectation derived from a correctly specified model, parameters known or efficiently estimated, dominates other forecast models in terms of frequentist risk, regardless of the data transformation that is employed. Only the ranking between misspecified or inefficiently estimated forecasting models changes. To guarantee a ranking that is robust to affine data transformations, Clements and Hendry propose to evaluate predictions over horizons 1 to h according to the determinant of the covariance matrix of stacked forecast errors. This criterion does not have an interpretation as expected forecast error loss. It is useful for situations in which it is not clear whether the audience is interested in forecasting levels or differences.

The fourth chapter covers the prediction of univariate processes. The exposition does not focus on the derivation of conditional expectations of y_{T+h} for a variety of time series models. Such calculations can be found in many other time series books. Instead, Clements and Hendry derive one-step and multistep forecast error losses for various univariate time series models and predictors, such as the AR(1) model, the random walk, and a trend stationary model. Parameter uncertainty is taken into account, and large sample approximations to the prediction risks are derived. The chapter also reviews various methods of calculating predictors for nonlinear models, such as setting the disturbances equal to zero, deriving exact conditional expectations, and approximations based on Monte Carlo or bootstrapping techniques. A discussion of forecasting processes that exhibit conditional heteroskedasticity under asymmetric loss functions follows. Throughout the chapter, it is assumed that the forecast model is correctly specified. The results could be interpreted as bounds on how well a forecaster can do under ideal circumstances.

Chapter 5 reviews some basic Monte Carlo techniques that are relevant for the computation of expected prediction losses and determinants of forecast error covariance matrices. The chapter is short and probably does not contain much new material for the target audience of the book. At the end of the chapter, however, the reader can find some proofs of the unbiasedness of forecasts based on autoregressive models. Although the estimator of the autoregressive

https://doi.org/10.1017/S0266466600003066 Published online by Cambridge University Press

parameter in an AR(1) model is severely biased in small samples, it can be shown that the forecasts based on the estimated coefficient are not. The result dates back to the early 1970's and is initially a bit surprising. The proof is based on the idea of "antithetics." If the distribution of the stochastic disturbance in the AR(1) model is symmetric, then the sequences $\{u_t\}$ and $\{-u_t\}$ are "equally likely." It turns out that the absolute values of the forecast errors under $\{u_t\}$ and $\{-u_t\}$ are the same, but the signs differ. Because the forecast errors average out, they are unbiased. In the context of Monte Carlo simulation antithetic variates are often used to reduce the variance of the Monte Carlo estimate. I did not quite understand why the unbiasedness of forecasts is discussed in a chapter on Monte Carlo simulation, but it is nice to find this argument in a recent monograph.

Chapter 6 extends the analysis of Chapter 4, and the last section of Chapter 5, to cointegrated vector autoregressive (VAR) models. Analytical derivations are spelled out in detail. The interesting question asked in the chapter is the following: suppose the goal is to forecast a nonstationary vector process. Which of the following procedures is preferable: estimate an unrestricted VAR in levels, estimate a VAR in first differences, or pretest for cointegration, for instance via Johansen's or the Engle–Granger two-step procedure? The chapter provides both some theoretical results and Monte Carlo evidence. Of course, the precise ranking of the procedures depends on the assumptions for the underlying DGP. Linear combinations that are stationary under the DGP are poorly forecasted by a VAR in differences. However, the VAR in differences does reasonably well in forecasting level variables. The Monte Carlo results indicate that the determinant criterion proposed in Chapter 3 penalizes the inability of the VAR in differences to forecast stationary linear combinations. Unfortunately, the comparison in Chapter 6 excludes model selection based procedures to determine the cointegration rank and Bayesian vector autoregressions, as discussed in Phillips (1995, 1996).

Chapter 7 considers forecasting with large scale macroeconometric models and develops a taxonomy of forecast errors. The forecast errors fall in roughly five categories, such as residual variance, parameter uncertainty due to estimation, and mismeasurement of the initial condition or forecast origin y_T, possibly due to initially inaccurate data. In addition, Clements and Hendry emphasize model misspecification, for instance the imposition of invalid cointegration restrictions, and parameter nonconstancy. The taxonomy provides an accounting framework for possible causes of predictive failure. Ex post, however, it seems very difficult to decompose forecast error into these categories, and the section on forecast evaluation techniques does not attempt to make such a decomposition.

A theory of intercept corrections is developed in Chapter 8. Intercept corrections are additive adjustments made to a point forecast from a probabilistic model. Such adjustments can lead to improved forecasts for a variety of reasons related to the taxonomy of forecast errors set forth in the previous chapter. Suppose the time series is nonlinear of the form $y_{T+1} = f(y_T, \epsilon_{T+1})$. The easiest

https://doi.org/10.1017/S0266466600003066 Published online by Cambridge University Press

method to calculate a predictor is $\tilde{y}_{T+1} = f(y_T, 0)$. If f is a nonlinear function of ϵ_{T+1} the predictor differs from the conditional expectation. An intercept correction potentially narrows the discrepancy between $\tilde{y}_{T+1}$ and $\mathbb{E}_T[f(y_T, \epsilon_{T+1})]$. However, given the recent acceleration of computer speed, the conditional expectation can in many applications be approximated by numerical integration methods. A second motivation for intercept corrections is the possibility that the forecaster has information in addition to the sample $y_1, \ldots, y_T$ that makes it possible to anticipate future events not incorporated in the specification of the probability model.

A third justification for intercept corrections, emphasized in Chapter 8, is robustness. Two types of correction are formally analyzed: shrinking the conditional expectation toward the unconditional mean; and "setting the forecast back on track" by adding the last observed prediction error $\hat{y}_{T|T-1} - y_T$. For instance, it is shown that the latter correction can lead to improved forecasts if a structural break did occur between periods $T - 1$ and T. Several strategies for multistep forecasting are compared with respect to bias and forecast error variance: no intercept correction, hold intercept correction constant over forecast period, only adjust the one-step forecast, adjust the h-step forecast by the full amount of the period T error. However, intercept corrections will make forecasts worse if no structural break occurred immediately prior to time T. As pointed out in the beginning of this review, the authors do not really discuss how to translate their insights about the potential benefits of intercept corrections into feasible forecasting rules.

Chapter 9 examines the role of leading indicators in macroeconomic forecasting. While some authors, for example Granger and Newbold (1986), have argued that leading indicators are mainly tools to predict turning points of the business cycle, Clements and Hendry regard them as tools for point forecasting. The chapter starts with a brief review of how composite leading indicators are constructed in the United Kingdom. A theoretical analysis is conducted in the context of a first-order cointegrated vector process. Clements and Hendry examine under what conditions a composite leading indicator (CLI), that is, a linear combination of time $T - 1$ variables, is helpful to forecast an index of aggregate activity, represented by a linear combination of time T variables. After some transformations of the multivariate system it is shown that only in rare cases is the leading indicator an optimal predictor of economic activity. The potential advantage of a leading indicator is parsimony, which reduces forecast error variance at the expense of increasing the bias. The issue then becomes whether there are other parsimonious forecasting models for the index of aggregate activity that do not fall into the CLI category. This is mostly an empirical question, and Clements and Hendry provide illustrations with U.K. data. The authors regard CLI's at best as an adjunct to, but not a substitute for, econometric modeling.

Chapter 10 discusses the combination of forecasts. Suppose that the data are generated from some probability distribution with conditional mean $\mathbb{E}_T[y_{T+1}] =$

https://doi.org/10.1017/S0266466600003066 Published online by Cambridge University Press

$m(y_T)$ and forecasts $f_1(y_T)$ and $f_2(y_T)$ are calculated based on two models $\mathcal{M}_1$ and $\mathcal{M}_2$. A combination of forecasts can lead to a reduction of frequentist forecast error loss if there exists a $\lambda \in (0,1)$ that minimizes $\mathbb{E}[m(y_T) - \lambda f_1(y_T) - (1 - \lambda) f_2(y_T)]^2$. If the conditional mean of one of the forecast models coincides with the conditional mean of y_T under the DGP, then no gains are possible from the combination. In a Bayesian framework, the combination of forecasts arises if nonzero prior probabilities are placed on $\mathcal{M}_1$ and $\mathcal{M}_2$. Under a quadratic forecast error loss function, the predictions of the two models are weighted by the respective posterior model probabilities. Provided that the dimension of the parameter space under the two models is constant and that some additional regularity conditions hold, the posterior probability of one of the models will converge to one. Thus, the Bayesian combination occurs only in finite samples but not asymptotically.

The Bayesian approach to the combination of forecasts (see, e.g., Min and Zellner, 1993) is not discussed in the book. Instead, the authors review various approaches to find combination weights that minimize the expected distance between $m(Y_T)$ and $\lambda f_1(y_T) + (1 - \lambda) f_2(y_T)$, such as the so-called regression method, the Granger–Bates approach, and variants thereof. Clements and Hendry provide a simple analytical example in which there is no gain from the combination of forecasts. The remainder of the chapter discusses the combination of conditional and unconditional forecasts. The conditional forecasts have potentially a high variance due to parameter uncertainty, whereas the unconditional forecasts are biased conditional on time T information. The combination tries to balance this trade-off. If the forecaster knew the distribution from which the data are generated, it would be possible to compute the optimal weights. However, in practice this distribution is unknown, and the weights have to be calculated based on sample information, which is likely to reduce the gains from combination. Some references to the extensive empirical literature on this subject are provided.

Multistep estimation of forecasting models is examined in Chapter 11. The fundamental issue is whether the loss function that is used to evaluate forecasts should also be used for parameter estimation. The answers provided in the literature range from yes to no and depend on the assumed degree of misspecification of the forecasting model relative to the sample size. If the misspecification is small, then efficiency gains through pseudo maximum likelihood or single-step estimation outweigh the increased asymptotic bias relative to loss function estimation. Clements and Hendry discuss the problem in terms of the forecast error taxonomy developed in Chapter 7 and provide a simple analytic example in which data are generated from an MA(1) model and the forecasts are based on an AR(1) model. The authors also examine the impact of small sample biases on the choice between single-step and multistep estimators. Monte Carlo simulations for a correctly specified AR(1) model suggest that the ranking of the estimators is not reversed in finite samples. A larger Monte Carlo study compares single-step and multistep estimators under various forms of misspec-

https://doi.org/10.1017/S0266466600003066 Published online by Cambridge University Press

ification. Unlike in the study by Weiss (1991), Clements and Hendry also consider nonstationary specifications. If the underlying process has a unit root and the estimated model omits some MA components, then the multistep estimators seem to be preferable at short horizons because they attain a more precise estimate of the unit root. Some asymptotic forecast error loss calculations are provided that support the Monte Carlo evidence.

Chapter 12 discusses the role of parsimony in forecasting. Increasing the dimensionality of a model helps improve the in-sample fit but often has opposite effects on the forecasting performance. There are many approaches to penalize dimensionality. In a Bayesian framework, model selection is based on the marginal densities of the data conditional on different models. The ratio between marginal data density and maximized value of the likelihood function could be interpreted as penalty for dimensionality. This penalty is not only a function of the number of parameters but also of the concentration of the prior distribution. Alternative approaches that lead to parsimonious model specifications and have a forecasting interpretation are predictive least squares (Wei, 1992) and prequential analysis (Dawid, 1992).

Clements and Hendry center their discussion around the following type of analysis: suppose $y_t = x_t' \beta + u_t$. Omitting a regressor $x_{1,t}$ introduces bias in the forecast but on the other hand reduces the variance of parameter estimates and the predictor. As a function of β_1 we can calculate the expected forecast error loss for predictors that include or omit regression $x_{1,t}$. If $\beta_1^2 > c$, for some threshold c, then it is preferable to estimate the parameter instead of imposing it to be equal to zero. Clements and Hendry propose forecasting rules of the following form: retain regressor $x_{1,t}$ if the squared t-statistic for $H_0 : \beta_1 = 0$ is greater than 2. This idea is generalized to more complicated dynamic specifications. The problem with these model selection strategies is that they are inconsistent: even in large samples the squared t-statistic can be greater than 2 if $\beta_1^2 < c$ and smaller than 2 if $\beta_1^2 > c$. Unfortunately, the frequentist properties of such pretest procedures are difficult to analyze theoretically. An empirical example at the end of the chapter provides some illustration.

Chapter 13 reviews tests for predictive failure and comparative forecasting accuracy, and Chapter 14 provides a summary of the main issues discussed in the book and an outlook to a second volume.

3. CONCLUSION

Michael Clements and David Hendry's monograph provides a comprehensive theoretical treatment of macroeconomic forecasting with an emphasis on potential misspecification of the probabilistic model that is used to generate the predictions. Taking such misspecification into account is an important step toward bridging the gap between theory and practice and will remain a fruitful research area in the future. Hence, the monograph is recommended reading for econometricians with applied and theoretical interests in macroeconomic fore-

https://doi.org/10.1017/S0266466600003066 Published online by Cambridge University Press

casting. Even though the theory developed in the book does not provide easy-to-use recipes to cope with potential model misspecification, practitioners can learn a lot about the theoretical underpinnings of heuristic adjustment procedures for probability model predictions. Because the book does not contain a systematic introduction to time series econometrics, it is not a replacement for one of the standard time series texts in a graduate course. However, it is a valuable supplement if the course emphasizes forecasting and should be on top of the reading list for a topics course on macroeconomic forecasting.

REFERENCES

Croushore, D. (1998) Evaluating Inflation Forecasts. Working paper 98-14, Federal Reserve Bank of Philadelphia.

Dawid, A.P. (1992) Prequential analysis, stochastic complexity and Bayesian inference. In J.M. Bernado, J.O. Berger, A.P. Dawid, & A.F.M. Smith (eds.), *Bayesian Statistics 4*, 109–125.

Granger, C.W.J., & P. Newbold (1986) *Forecasting Economic Time Series*, 2nd ed., New York: Academic Press.

Hamilton, J.D. (1994) *Time Series Analysis.* Princeton, New Jersey: Princeton University Press.

Hendry, D.F. (1995) *Dynamic Econometrics.* New York: Oxford University Press.

Min Chung-ki & A. Zellner (1993) Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics* 56, 89–118.

Phillips, P.C.B. (1995) Bayesian model selection and prediction with empirical applications, with discussion. *Journal of Econometrics* 69, 289–365.

Phillips, P.C.B. (1996) Econometric model determination. *Econometrica* 64, 673–812.

Schorfheide, F. (1998) *Econometric Modeling of Macroeconomic Aggregates.* Ph.D. diss., Yale University.

Wei, C.Z. (1992) On predictive least squares principles. *Annals of Statistics* 20, 1–42.

Weiss, A. (1991) Multi-step estimation and forecasting in dynamic models. *Journal of Econometrics* 48, 135–149.

https://doi.org/10.1017/S0266466600003066 Published online by Cambridge University Press