Working Papers

Research Department

MEASURING PREDICTABILITY:
THEORY AND MACROECONOMIC APPLICATIONS

Francis X. Diebold
Department of Economics
University of Pennsylvania
Visiting Scholar, Federal Reserve Bank of Philadelphia

Lutz Kilian
Department of Economics
University of Michigan

WORKING PAPER NO. 97-23
MEASURING PREDICTABILITY:
THEORY AND MACROECONOMIC APPLICATIONS

Francis X. Diebold
Department of Economics
University of Pennsylvania
Visiting Scholar, Federal Reserve Bank of Philadelphia

Lutz Kilian
Department of Economics
University of Michigan

August 1996
This Print: December 20, 1997

ABSTRACT

We propose a measure of predictability based on the ratio of the expected loss of a short-run forecast to the expected loss of a long-run forecast. This predictability measure can be tailored to the forecast horizons of interest, and it allows for general loss functions, univariate or multivariate information sets, and stationary or nonstationary data. We propose a simple estimator, and we suggest resampling methods for inference. We then provide several macroeconomic applications. First, based on fitted parametric models, we assess the predictability of a variety of macroeconomic series. Second, we analyze the internal propagation mechanism of a standard dynamic macroeconomic model by comparing predictability of model inputs and model outputs. Third, we use predictability as a metric for assessing the similarity of data simulated from the model and actual data. Finally, we sketch several promising directions for future research.

Correspondence to:
F.X. Diebold
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104
fdiebold@mail.sas.upenn.edu

1. Introduction

It is natural and informative to judge forecasts by their accuracy. However, actual and forecasted values will differ, even for very good forecasts. To take an extreme example, consider a zero-mean white noise process. The optimal linear forecast under quadratic loss is simply zero, so the paths of forecasts and realizations will look different. These differences illustrate the inherent limits to predictability, even when using optimal forecasts. The extent of a series' predictability depends on how much information its past conveys regarding its future; as a result, some processes are inherently easy to forecast, and others are difficult.

In addition to being of interest to forecasters, predictability measures are potentially useful in empirical macroeconomics. Predictability provides a succinct measure of a key aspect of time series dynamics and is therefore useful for summarizing and comparing the behavior of economic series, as well as for assessing agreement between economic models and data.[1]

Remarkably little attention has been paid to methods for measuring predictability. Existing methods include those based on canonical correlations between past and future and those based on comparing the innovation variance and unconditional variance of stationary series.[2] Those methods, however, are inadequate in light of recent work stressing nonstationarities of various sorts, rich and high-dimensional information sets, nonquadratic and

---

[1] We do not advocate comparing models to data *purely* on the basis of predictability. Rather, predictability simply provides an easily digested summary distillation of certain important aspects of dynamics. More complete frameworks for assessing agreement between models and data are developed in King and Watson (1996) and Diebold, Ohanian and Berkowitz (1995).

[2] See Jewell and Bloomfield (1983), Jewell *et al*. (1983), Hannan and Poskitt (1988), and Granger and Newbold (1986).

possibly even asymmetric loss functions, and variations in forecast accuracy across horizons.[3]

The lack of methodological development coincides, not surprisingly, with a lack of substantive exploration. Even for the major macroeconomic aggregates, very little is known about comparative predictability. At first glance, the assertion that we know little about predictability seems exaggerated. We know, for example, that consumption is less volatile than output, and that investment is more volatile than output.[4] Such statements, however, concern *unconditional* variances, whereas predictability concerns variances *conditional* on varying information sets. The two concepts are very different, as illustrated, for example, in studies of inflation, in which the unconditional variance simply measures inflation variability, whereas the conditional forecast error variance measures inflation uncertainty.[5]

In this paper, we contribute to the theory of predictability measurement and apply our results in several macroeconomic contexts. In section 2, we discuss some of the difficulties involved in predictability measurement and propose a simple measure of relative predictability based on the ratio of the expected loss of a short-run forecast to the expected loss of a long-run forecast. Our measure allows for stationary or nonstationary series, univariate or multivariate information sets, general loss functions, and different forecast horizons of interest. In section 3, we propose methods for estimating the predictability of observed series, and we suggest using bootstrap methods for inference. In section 4, we assess the predictability of a variety of U.S. macroeconomic series. In section 5, we illustrate the use of predictability measures in assessing

---

[3] See, among others, Stock (1995), Forni and Reichlin (1995), Diebold and Mariano (1995), and Christoffersen and Diebold (1996, 1997).

[4] See, for example, Cooley and Prescott (1995).

[5] See, for example, Ball and Cecchetti (1990).

the propagation mechanisms of economic models and in assessing agreement between economic

models and data; the results are sharp and surprising.

The exercises we undertake in section 5 are very similar in spirit to those of Rotemberg

and Woodford's (1996) important paper, which focuses on predictions and predictability in the

analysis of a real business cycle model.[6] We share Rotemberg and Woodford's interest in

assessing models via the properties of their predictions; moreover, a comparison of their paper

and ours reveals that the two are highly complementary. Rotemberg and Woodford's primary

concern is with a certain class of business cycle models, and their analysis of predictability is

very tightly and appropriately linked to the assumptions of the particular models they study. Our

primary focus, in contrast, is on general macroeconometric methods of assessing predictability

that can be applied in a variety of situations, under minimal assumptions. Our methods are

designed both to provide model-free measures of the predictability of data and to compare

formally the predictability of actual and simulated data.

2. Population Predictability Measures

The expected loss of an optimal forecast will, in general, exceed zero, which illustrates

the inherent limits to predictability, even when using optimal forecasts. Put differently, poor

forecast accuracy does not necessarily imply that the forecaster failed. The extent of a series'

predictability in population depends on how much information its past conveys regarding its

future; given an information set, some processes are inherently easy to forecast, and others are

---

[6] We were not aware of Rotemberg and Woodford's work when writing the first draft of this paper. We thank Garey Ramey for bringing it to our attention.

difficult. This point is not purely academic. For example, a policymaker may choose to target

nominal income rather than inflation, if the former turns out to be much more predictable

(Cecchetti, 1995).

In measuring predictability it is important to keep three points in mind. First, the

question of whether a series is predictable or not should be replaced by one of *how* predictable it

is. Predictability always is a matter of degree.[7] Second, the question of how predictable a series

is cannot be answered in general. We have to be clear about the relevant forecast horizon and

loss function. For example, a series may be quite predictable at short horizons, but not very

predictable at long horizons. Third, to compare the predictability of several series we need a

common numeraire. It may be tempting to simply compare the expected losses of forecasts for

two series to assess their relative predictability, but that ignores the possibility that the two series

may be very different in magnitude or may be measured on different scales.

Granger and Newbold (1986, p. 310) therefore propose a natural measure of the

forecastability of covariance stationary series under squared-error loss, patterned after the

familiar R^2 of linear regression,

$$G = \frac{\text{var}\left(\hat{y}_{t+j,t}\right)}{\text{var}\left(y_{t+j}\right)} = 1 - \frac{\text{var}\left(e_{t+j,t}\right)}{\text{var}\left(y_{t+j}\right)},$$

where $\hat{y}_{t+j,t}$ is the optimal (i.e., conditional mean) forecast and $e_{t+j,t} = y_{t+j} - \hat{y}_{t+j,t}$. The Granger-

---

[7] In particular, questions such as "Are exchange rates predictable?" really are questions about
the usefulness of a particular information set for forecasting. In this paper, we take the
information set as given.

Newbold measure has been used, for example, in Barsky (1987) to explain the strong correlation

between nominal interest rates and realized inflation rates in postwar U.S. data in terms of the

ratio of 1-step-ahead errors in inflation forecasts *relative* to the long-run variability in inflation.

Our approach to predictability measurement is squarely in the tradition of Granger and

Newbold, with the important difference that we relax several constraints that limit the broad

applicability of their methods. The essence of the Granger-Newbold suggestion is that it is

natural to base a measure of predictability on the difference between the conditionally expected

loss of an optimal short-run forecast, $E\big(L(e_{t+j,t})\big)$, and that of an optimal long-run forecast,

$E\big(L(e_{t+k,t})\big)$, $j \ll k$.[8] If $E\big(L(e_{t+j,t})\big) \ll E\big(L(e_{t+k,t})\big)$, we say that the series is highly predictable at

horizon j relative to k, and if $E\big(L(e_{t+j,t})\big) \approx E\big(L(e_{t+k,t})\big)$, we say that the series is nearly

unpredictable at horizon j relative to k. Thus, we define a general measure of predictability as

$$P(L,\Omega,j,k) \;=\; 1 \;-\; \frac{E\big(L(e_{t+j,t})\big)}{E\big(L(e_{t+k,t})\big)},$$

where the information set Ω can be univariate or multivariate, as desired. The Granger-Newbold

measure emerges in the special case in which the series is covariance stationary, $L(x) = x^2$ (and

hence the optimal forecast is the conditional mean), the information set is univariate, and $k=\infty$.

The advantages of our generalization include:

(1) It allows analysis of nonstationary as well as stationary series, so long as $k<\infty$.

(2) It allows for general loss functions. The loss function $L(\cdot)$ need not be quadratic, and

need not even be symmetric; we require only that $L(0) = 0$ and that $L(\cdot)$ be strictly

---

[8] Here and throughout, E(.) denotes mathematical expectation conditional on the information
set Ω.

monotone on each side of the origin.[9] By the restrictions imposed on $L(\cdot)$, we have that $P(L(\cdot),\Omega,j,k) \in [-\infty, 1]$, with larger values indicating greater predictability. In most practical applications one expects $0<P(L(\cdot),\Omega,j,k)<1$, as the distant future is likely to be harder to forecast than the near future.[10]

(3) It allows for univariate or multivariate information sets, and economic theory may suggest relevant multivariate information sets. Implicit in the information set Ω is the choice of whether any constraints, such as linearity in past observations, are imposed when solving for the minimum-expected-loss forecast.

(4) It allows for flexibility in the choice of j and k and enables one to tailor the predictability measure to the horizons of economic interest.

Our predictability measure is closely related to Theil's (1966) U statistic, which we define for the 1-step-ahead horizon as

$$U = \frac{E\left(e_{t,t-1}^2\right)}{E\left((y_t - y_{t-1})^2\right)}.$$

To make the relationship transparent, specialize P to the quadratic, univariate, j=1 case and write it as

---

[9] In fact, we can allow for even greater generality by writing $L(y,\hat{y})$.

[10] In covariance stationary environments one can prove that $P(L(\cdot),\Omega,j,k) \in [0,1] \ \forall \ j<k$, but that result need not hold in nonstationary environments. Consider, for example, forecasting a series given by white noise deviations from a known trend, with the variance of the white noise decreasing over time.

$$P(\text{quadratic,univariate},1,k) = 1 - \frac{E\left(e_{t,t-1}^2\right)}{E\left(e_{t,t-k}^2\right)},$$

or

$$1 - P = \frac{E\left(e_{t,t-1}^2\right)}{E\left(e_{t,t-k}^2\right)}.$$

Thus, under certain conditions, 1-P is similar in spirit to Theil's U. The key difference is the numeraire. Theil's U assesses 1-step forecast accuracy relative to that of a "naive" no-change forecast, whereas P assesses 1-step accuracy relative to that of a long-horizon (k-step) forecast. In the general case,

$$P(L(\cdot),\Omega,j,k) = 1 - \frac{E\left(L(e_{t,t-j})\right)}{E\left(L(e_{t,t-k})\right)}.$$

Thus, $P(L(\cdot),\Omega,j,k)$ is effectively one minus the ratio of expected losses of two forecasts of the same object, y_t. One forecast, $\hat{y}_{t,t-j}$, is (typically) based on a rich information set, while the other forecast, $\hat{y}_{t,t-k}$, is typically based on a sparse information set.

The formula for $P(L(\cdot),\Omega,j,k)$ also makes clear that the concept of predictability is related to, but distinct from, the concept of persistence of a series. Suppose, for example, that the series y_t is a random walk. Then $P(e^2,\text{univariate},j,k) = 1 - \frac{j}{k}$, as will be shown later. But for a random walk the variance ratio at horizon j, a common measure of persistence, is[11]

---

[11] Diebold and Nerlove (1990) discuss the variance ratio and its relationships to other popular persistence measures.

$$V_j = \frac{\text{var}(y_t - y_{t-j})}{\text{var}(y_t - y_{t-1})} = j.$$

It is clear, however, that although $P(e^2,\text{univariate},j,k)$ and V_j are deterministically related in the random walk case ($P = 1 - V/k$), they are not deterministically related in more general cases.

3. Sample Predictability Measures

Predictability is a population property of a series, not any particular sample path, but predictability can be estimated from a sample path. We proceed by fitting a parametric model and then transforming estimates of the parameters into an estimate of P.[12] To keep the discussion tractable, and in keeping with the empirical analysis of subsequent sections, we use the quadratic loss function $L(e)=e^2$ for estimation, prediction, model selection, and construction of predictability measures.[13] We fit VAR(p) models, although one could easily generalize the discussion to other parametric models, such as vector ARMA(p,q) models.

Consider the N-dimensional VAR(p) process,

$$y_t = d + A_1 y_{t-1} + ... + A_p y_{t-p} + u_t$$

$$u_t \sim \text{WN}(0, \Sigma_u),$$

where d denotes possible deterministic components of the process. Assuming that u_t is independent white noise, the conditional expectation, provided that expectation exists, will be the optimal h-step-ahead forecast (that is, the forecast with minimum MSE). This will be true, even if $\det(I_N - A_1 z - ... - A_p z^p)$ has roots on the unit circle. The associated h-step-ahead forecast

---

[12] In section 6 we discuss alternative implementations of the predictability measure.

[13] Techniques for estimation, prediction, and model selection under other loss functions are discussed in Christoffersen and Diebold (1996, 1997).

MSE matrix is

$$\Sigma_y(h) \;=\; \sum_{i=0}^{h-1} \Phi_i \Sigma_u \Phi_i',$$

where $\Phi_0 = I_N$ and

$$\Phi_i \;=\; \sum_{j=1}^{i} \Phi_{i-j} A_j,$$

$i = 1, 2, \ldots$ In the stable case, the Φ_i converge to zero as $i \to \infty$ and $\Sigma_y(h)$ converges to the

unconditional variance-covariance matrix of y_t as $h \to \infty$. In the unit root case, the Φ_i do not

converge to zero, and consequently some elements of $\Sigma_y(h)$ will approach infinity as $h \to \infty$. In

other words, the forecast error variances will be unbounded and the forecast uncertainty will

become large as we predict the distant future. We construct P by simply reading off the

appropriate diagonal elements of the forecast MSE matrices for forecast horizons j and k. To

build intuition, consider a univariate AR(1) population process with innovation variance Σ_u:

$$y_t \;=\; A_1 y_{t-1} \;+\; u_t$$

Then for $A_1 = 0$ the model reduces to white noise, and short-run forecasts are just as accurate as

long-run forecasts. As a result, relative predictability is zero: $P(j,k) = 1 - \Sigma_u / \Sigma_u = 0$ for all j. In

contrast, for $A_1 = 1$ the model becomes a random walk, and relative predictability steadily

declines as the forecast horizon increases: $P(j,k) = 1 - (j\Sigma_u)/(k\Sigma_u) = 1 - j/k$.

Forecast errors from consistently estimated processes and processes with known

parameters are asymptotically equivalent. This suggests estimating P by replacing the underlying

unknown parameters by their least squares estimates; we do so throughout.[14] To determine the autoregressive lag order p we use the Akaike Information Criterion (AIC) with a suitable upper bound on the admissible lag orders. The AIC guarantees a consistent estimate of P and is less likely to underestimate the lag order in small samples than alternative criteria. The latter property is crucial in preserving the higher-order dynamics implicit in P.[15]

For stationary series, we propose building up a bootstrap approximation to the sampling distribution of $\hat{P}$.[16] To do so, we condition on the estimated model and resample with replacement from its residuals, as though the estimated residuals and parameters were the population values. Similar techniques have been used by Stine (1987) and Runkle (1987). However, we improve the coverage accuracy of bootstrap confidence intervals by several modifications. We use the first-order bias corrections of Pope (1990) and Kilian (1995) to remove the small-sample bias in the initial OLS estimate of the autoregressive coefficients prior to bootstrapping, following a suggestion by Nicholls and Pope (1988). In addition, we

---

[14] It may be possible to obtain estimators of the forecast MSE with better small-sample properties than the substitution estimator. For example, Lütkepohl (1991) discusses an approximate estimator of the forecast MSE derived from the asymptotic theory for stationary VAR(p) models, and Stine (1987) shows that first-order bias correction can considerably improve the accuracy of estimators of the forecast MSE in stationary AR(p) models. However, unlike the substitution estimator, the refinements require the assumption of stationarity. Moreover, it is not clear to what extent the refinements will translate into better estimates of P, which is a *ratio* of forecast MSEs. We defer a systematic study of the small-sample properties of alternative estimators of P to future research.

[15] See Kilian (1996a).

[16] We did not explore the possibility of constructing confidence intervals based on the delta method, because the asymptotic normality of $\hat{P}$, required for the delta method, has not yet been established. In addition, the derivation of a closed-form solution for the asymptotic standard error is likely to be complicated. Most importantly, we expect the distribution of $\hat{P}$ to be skewed in small samples and are therefore reluctant to impose symmetry.

endogenize the lag order choice by reestimating the lag order for each bootstrap replication, as in

Kilian (1996b). Finally, we also correct for bias in the bootstrap estimates as discussed in Kilian

(1995).

We use the ordinary percentile bootstrap interval, because, pending the establishment of

asymptotic normality of $\hat{P}$, there is no rationale for studentizing it. Even if $\hat{P}$ were

asymptotically normal, moreover, the percentile-t interval may not be a good choice for the P

statistic. The percentile-t method is known to perform poorly for ratio estimators such as the

correlation coefficient. We suspect that without additional variance-stabilizing transformations it

would perform poorly for our ratio estimator as well. Such transformations are not known and

would have to be simulated by bootstrap, adding another layer of burdensome simulation.[17]

Finally, we note that although the legitimacy of our predictability measure is invariant to

the possible presence of unit roots, the legitimacy of our bootstrap inference is not, because of the

discontinuity of the distribution theory for the estimated VAR at the unit circle. However, there

is reason to believe that the bootstrap may still provide a reasonable approximation in finite

samples, in which there are no such discontinuities.[18] Alternatively, one may choose to impose a

unit root in resampling, when appropriate.

In Figures 1-3, we report the results of a small Monte Carlo experiment designed to help

assess the adequacy of our bootstrap intervals. We began by fitting an AR(6) to the U.S. 10-year

Treasury bond rate. The dominant root of the estimated model was large but less than unity. We

then froze the estimated model and used it as our data-generating process. We examined sample

---

[17] See Efron and Tibshirani (1993).

[18] See Kilian (1995,1996c).

sizes of T=80 (Figure 1), T=160 (Figure 2) and T=480 (Figure 3), selected to represent 20 years

of quarterly data, 40 years of quarterly data, and 40 years of monthly data.

For each Monte Carlo trial, we let the AIC choose a lag order between 1 and 8 and

compute the implied bootstrap confidence intervals with 90% nominal coverage. The results

indicate that bias corrections and endogenous lag-order selection yield substantial improvements

in coverage accuracy compared to the standard bootstrap. However, the coverage of the bootstrap

interval still tends to fall short of the nominal probability content. The empirical coverage, for

example, is too small by up to 30% for T=80 and by up to 25% for T=160. For T=480, however,

the coverage is all but perfect. Additional improvements in coverage accuracy for the smaller

sample sizes will probably require calibration of the probability content of the interval or use of a

more efficient estimator; we leave those modifications for future research.[19]

4. Predictability of U.S. Macroeconomic Series

We examine postwar quarterly data; the sample period is 1947.2-1994.3 unless otherwise

noted. The dataset includes a broad and representative sample of macroeconomic variables,

similar to that of Nelson and Plosser (1982). We log and seasonally adjust all data except for

rates. We use univariate information sets.[20] We model trending series as autoregressions in

levels with a linear time trend included, and we model rates and ratios as autoregressions in

---

[19] We do not prove the asymptotic validity of bootstrapping the quantiles of the distribution of $\hat{P}$ in stationary autoregressions, but related results in Hall (1992) for the "smooth function model" (p. 52), which includes ratios of variances, are suggestive.

[20] Note that the univariate autoregressive representation of the series may be interpreted without loss of generality as a marginalized reduced form of a more general vector autoregressive model.

levels without a trend. We determine autoregressive lag orders using the AIC. In assessing

predictability, we fix $L(e) = e^2$ and k=40, as we vary the near-term forecast horizon j. In Figures

4-12, we plot interval estimates of P (in percent) against near-term forecast horizons j = 1, ..., 20.

Higher values of P indicate greater predictability. We construct 1000 bootstrap replications and

plot nominal 90% confidence intervals based on the 5% and 95% points of the bootstrap

distribution. We do not display point estimates because interval estimates are likely to be more

informative, given the large sampling uncertainty in many estimates.

In Figure 4 we show the estimated predictability of various measures of output. Industrial

production appears about as predictable as real GNP at short horizons, but more predictable at

long horizons. Nominal GNP and real disposable income also appear more predictable than real

GNP at longer horizons, but their estimates are considerably less precise. In contrast, the

predictability of per capita real GNP and of per capita disposable income can be rather precisely

estimated. Per capita real GNP is somewhat less predictable than real GNP, whereas per capita

real disposable income is much more predictable than real GNP or any other income measure.

In Figures 5-7 we disaggregate the expenditure side of the national income and product

accounts. In Figure 5 we show predictability of private consumption and government purchases.

Private consumption is more predictable (and more precisely estimated) than real GNP, and there

are considerable differences between consumption components. Predictability of nondurables and

services is imprecisely estimated and could be more or less predictable than aggregate private

consumption, whereas durables are fairly precisely estimated and are much less predictable than

total consumption or nondurables and services. Government purchases are about as predictable as

real GNP, but again disaggregation proves revealing: federal government purchases of goods

and services are about as predictable as GNP, but state and local government purchases are much more predictable with very low sampling uncertainty. Turning to investment in Figure 6, private investment spending is clearly one of the least predictable components of real expenditure. At long horizons it is even slightly less predictable than consumer durables. Broken down by component, changes in inventories of nonfarm businesses are extremely unpredictable, followed by residential and nonresidential investment, in that order. However, within nonresidential investment spending, structures are much more predictable than producer durables. In Figure 7 we summarize the external sector. Interestingly, although real exports and real imports are both somewhat less predictable than real GNP, especially at short horizons, the real balance of goods and nonfactor services is more predictable at all forecast horizons.

In Figure 8 we display predictability of various labor market variables for 1964.1-1991.4. The predictability of the nominal wage is much less precisely estimated than that of the real wage. The latter interval is entirely contained in the former interval, as are the intervals for real and nominal GNP. While the sampling uncertainty for the unemployment rate is not quite as large, the interval still contains that for real GNP. However, the predictability of employment is sharply estimated and similar to real GNP in shape (or slightly lower at longer forecast horizons, depending on whether household or establishment survey data are used). Hours worked tend to be less predictable than most measures of output, especially at horizons between one and three years.

In Figure 9 we show predictability of price indices and inflation. The CPI is slightly more predictable than the implicit GNP deflator, and about as predictable as the PPI for finished goods. The PPIs for intermediate goods and for crude materials are somewhat less predictable, but the

differences are not striking. Inflation (using the GNP deflator) on the other hand is virtually

unpredictable for horizons of more than one year.

In Figure 10 we display predictability of velocity and various monetary aggregates. M1

and the monetary base are the least predictable monetary aggregates, with the latter being much

more precisely estimated. They are followed, in order of increasing predictability, by

nonborrowed reserves, M2, and M3. The latter measures of money all appear more predictable

than real GNP, and M2 and M3 appear even more predictable than nominal GNP. Velocity

(based on M1) is fairly precisely estimated and highly predictable.

In Figures 11 and 12 we show the predictability estimated for various interest rates,

1971.4-1993.4. In Figure 11 we show predictability estimates for Treasury bill and bond rates.

The predictability of the fed funds rate, the three-month T-bill rate and the corresponding

commercial paper and CD rates is very similar, but predictability increases as we move from the

three-month T-bill rate to the one-year and five-year T-bond rates. Unfortunately, the estimates

also become less precise. In Figure 12 we therefore consider a broader set of secondary market

interest rates for a longer sampling period. The rates are classified by their maturity, ranging from

three months to an average of maturities exceeding 10 years. The sample period is 1953.4-

1994.4. There is clear evidence that interest rates become more predictable as the maturity

increases and that the sampling uncertainty declines. The estimates for 10 or more years

especially are very precise.

In closing this section, we step back to stress a few regularities that emerged. First, the

confidence intervals often are rather wide, reflecting the limited information available in

macroeconomic data, but not so wide as to render the exercise futile. Second, the predictability of

nominal series tends to be estimated less precisely than the predictability of real series. Third, investment and related series such as consumer durables are hard to predict. Finally, disaggregation matters a great deal; the sub-components of various aggregates often have very different predictability properties, which make economic sense. For example, measured consumption of durable goods, which, in fact, is largely investment, is very hard to predict, whereas consumption of nondurable goods is easy to predict. We also observed important differences in predictability between increasingly broad measures of money supply and between federal and local government purchases.

5. Comparing Predictability of Model Inputs, Model Outputs, and Real Data

Predictability measures are potentially useful for assessing the internal propagation mechanisms of economic models and for assessing agreement between models and data. We illustrate both uses with the indivisible labor model of Hansen (1985). In that model, the representative agent chooses labor input, h_t, and next period's capital stock, k_{t+1}, to maximize expected lifetime utility,

$$E \sum_{t=0}^{\infty} \beta^t \left(\log c_t + B (1-h_t) \right),$$

subject to the constraints,

$$c_t + i_t \leq z_t k_t^{\theta} h_t^{1-\theta}$$

$$k_{t+1} = (1-\delta) k_t + i_t$$

$$z_{t+1} = (1-\gamma) + \gamma z_t + \varepsilon_{t+1},$$

where k_0 and z_0 are given, $0 < \beta < 1$, $0 \le \delta \le 1$, and $\varepsilon_t \sim N(0, \sigma^2)$. We parameterize the model

as in Hansen (1985) and solve for the associated linear decision rules for h_t and k_{t+1} in terms of

the current period states k_t and z_t.

Propagation Mechanisms: Comparative Predictability of Model Inputs and Model Outputs

To assess the internal propagation mechanism of Hansen's model, we compare the

predictability of the model input (the exogenous technology shock) and the model outputs (the

endogenous model variables). A weak propagation mechanism, assessed in terms of

predictability, is associated with nearly identical input and output predictability, and conversely.

We calculate the predictability of the technology shock analytically, based on population

parameter values. We calculate the predictability of the model outputs numerically. We can

calculate the predictability of the model outputs to any desired degree of accuracy by simulating a

long enough realization from the model (the "model data") and then estimating predictability by

fitting an autoregressive model and constructing our substitution estimator. We compute the

predictability of the model data on output, consumption, investment, productivity, capital stock,

and hours using a simulated realization of length 10,000.

The results appear in Figure 13, in which we show the comparative predictability of the

technology shock and the model outputs. Figure 13 makes clear that many of the model outputs

have very different predictability patterns than does the model input. Model real GNP is about as

predictable as the technology shock. In contrast, model investment is less predictable than the

technology shock; model consumption and productivity are more predictable than the technology

shock; model capital stock is much more predictable than the technology shock; and model hours

are much less predictable than the technology shock.

The results indicate that the indivisible labor model has a strong internal propagation mechanism, in the sense that the predictabilities of model outputs are distinctly different from that of the technology shock and different from one another. Some model series are more predictable than the technology shock; some are less predictable; and the differences arise endogenously from the model's internal propagation mechanism. This result is surprising, insofar as other studies using other criteria have concluded that models such as Hansen's have weak internal propagation mechanisms (e.g., Cogley and Nason, 1995a).

Thus far, neither the model inputs nor the model outputs have been logged or filtered, which effectively amounts to population linear detrending of both the input and output series, because no trends are operative. In our judgment that is the right way to proceed if we are interested in assessing the propagation mechanism of the model. In many applications, however, the focus is on how well the dynamics in simulated model data match those of the cyclical component of the actual data (expressed in percent deviations from a smooth trend). A common approach is to log and HP-filter the model outputs. We show the results of doing so in Figure 14. All model outputs are now much less predictable than the technology shock, which is not surprising, because the HP filtering removes highly predictable low-frequency components. Thus it is nonsensical to compare model input and output predictabilities when only the outputs have been HP filtered. We can, however, still compare the predictability of the various model outputs: all become strikingly similar after HP filtering. To an unsuspecting observer, this finding may seem to suggest that the technology shock imparts a common pattern of predictability.

A natural conjecture, however, is that a common predictability pattern of HP-filtered

model outputs is a spurious artifact of HP filtering.[21] To explore that conjecture, we compute the

predictability of the logged and HP-filtered technology shock, whose dynamics of course cannot

have anything to do with the model's propagation mechanism. For illustrative purposes we

consider six alternative values for the persistence of the technology shock, including the value of

$\gamma = 0.95$ used in Hansen (1985). The result appears in Figure 15: for all realistic values of γ, the

predictability of the HP-filtered technology shock looks the same as the predictability of HP-

filtered model outputs! We conclude that HP filtering tends to make predictabilities look similar

and thereby masks the strong propagation mechanism in Hansen's model that is revealed in

unfiltered data.

---

[21] A number of earlier studies have documented such effects, from perspectives different from
predictability. They include King, Plosser and Rebelo (1988), Singleton (1988), Harvey and
Jaeger (1993), and Cogley and Nason (1995b).

Goodness of Fit: Comparative Predictability of Model Data and Real Data

We compute the predictability of model data, and we estimate the predictability of the actual U.S. data for 1955:III-1984:I, using the methods already described. Computation of the predictability of the model data is not subject to sampling error, while the estimation of predictability of the real data is; thus, we compute interval estimates only for the latter. We consider the model to be consistent with the data if the model measure of predictability is contained in the confidence bands estimated from the U.S. data.

The real business cycle school favors logging and HP filtering both the model data and the real data, so we begin with that strategy. In Figure 16 we show the predictability of model data and actual data, where both have been HP filtered and logged. Disregarding some minor discrepancies, the model data and real data predictabilities generally agree. Only at very short time horizons are the model data not contained in the confidence bands for the U.S. data. Of course, the relative success of the model may simply reflect the small sample size of the U.S. data and the large degree of sampling uncertainty. Furthermore, note that although the predictabilities of model data and real data generally agree, they always have the same humped shape— precisely the same humped shape that the HP filter tends to impart, as we showed earlier. Thus we're naturally suspicious that the results may be an artifact of HP filtering. To address this possibility, we continue to treat the model data and real data symmetrically by using identical detrending procedures, but we use standard linear detrending instead of HP filtering.

In Figure 17 we show estimates of the predictability of model data and U.S. data based on log-linear detrending. The predictabilities of model data and actual data summarized in Figure 17 show a pronounced divergence: only model investment and possibly productivity match the data;

model output and hours worked are not predictable enough, whereas consumption and the capital

stock are too predictable in the model. This result is robust whether or not a trend is fitted to the

model data, as shown in Figure 18.[22] We have chosen to present our results for the sample period

used in Hansen (1985). Comparing these results to our earlier findings for longer samples of

U.S. data suggests that output and hours worked in the model are roughly as predictable in the

model as in the U.S. data, whereas consumption and investment clearly are not.

6. Concluding Remarks and Directions for Future Research

We have proposed operational methods for measuring predictability and put them to work

in several contexts— measuring the predictability of a variety of U.S. macroeconomic series,

assessing the internal propagation mechanism of a simple macroeconomic model, and assessing

agreement between the model and the data. Our main intent is the introduction and illustration of

an approach to predictability measurement, not provision of a complete evaluation of a particular

macroeconomic model or the HP filter. Nevertheless, our results reveal some successes and some

failures of the model and the filter.

There are many useful directions for future research. Some are obvious, but nevertheless

important, variations on the applications reported here. To take one example, in our applications

we estimate predictability on the basis of univariate information sets, whereas the theory allows

for multivariate information sets. Empirical predictability measurement on the basis of

multivariate information sets, and comparison of the univariate estimates to various multivariate

---

[22] Unlike the real data, the model data underlying Figure 18 have not been detrended. The model is effectively a model of deviation from linear trend and therefore displays no trend. Not detrending amounts to imposing the knowledge that the population trend is zero.

estimates, will be of interest. To take a second example, we have allowed only for deterministic

linear trends in the models from which we estimate predictability. The linear deterministic trend

is certainly a great workhorse in applied econometrics, and it often compares favorably to

competitors in forecasting and macroeconometric studies.[23] Nevertheless, it will be of interest to

allow for nonlinear or stochastic trends.

Other directions for future research are wide-ranging and fundamental and, hence, more

interesting. We shall briefly discuss three.

Nonparametric Predictability Estimation

We presented our approach, based on fitting autoregressive models, as a parametric

method. This convention facilitated the exposition and allowed us to draw on established results

for bootstrap inference. However, in general, we need not assume that the fitted autoregression is

the true data-generating process; rather, it may be considered an approximation, the order of

which can grow with sample size. Thus the autoregressive model can also be viewed as a sieve in

the sense of Grenander (1981). In that sense, our approach actually *is* nonparametric.

Nevertheless, the sieve approach has a parametric flavor. For any fixed sample size, we

assess predictability through the lens of a particular autoregressive model. In the future, it may be

of interest to develop an approach with a more thoroughly nonparametric flavor by exploiting

Kolmogorov's well-known spectral formula for the univariate innovation variance,

$$\sigma^2 \; = \; \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln \; 2\pi \; f(\omega) \; d\omega\right\},$$

---

[23] See, for example, DeJong and Whiteman (1994), Diebold and Senhadji (1996), and the references therein.

where f is the spectral density function. Kolmogorov's result has been extended to univariate h-step-ahead forecast error variances by Bhansali (1992) and to multivariate h-step-ahead forecast error variances by Mohanty and Pourahmadi (1996). Several technical problems remain, however, before we can operationalize those methods in our context.

Predictability of Financial Asset Returns and Volatilities

We have focused on the application of predictability measurement to macroeconomics. It should also prove useful in finance, in which the predictability of asset returns at various horizons is a central concern. In fact, if nonpredictability was arguably the central concern of the 1960s and 1970s literature (e.g., Fama, 1970), precisely the opposite is true of the more recent literature (e.g., Fama, 1991, Campbell, Lo and MacKinlay, 1997). We have reserved application to finance for a separate paper, however, in order to devote the necessary attention to the special concerns of that literature, including multivariate information sets, very long-horizon predictability, conditional heteroskedasticity, and possibly nonquadratic loss functions.[24]

Of equal, and perhaps even greater, importance is measuring the predictability of asset return *volatility* across various horizons. Tracking and forecasting time-varying volatility is at the heart of the booming risk management industry. Little attention, however, has been given to assessing patterns of volatility predictability, in particular, the speed and pattern with which volatility predictability decays as the horizon grows.

Survey-Based Predictability Estimation

---

[24] Mean absolute error (MAE), for example, may be more useful for measuring the accuracy of forecasts of financial series, because it exists in a wider range of situations. In high-frequency financial data, for example, fat-tailed distributions are common, in which case MSE may be infinite but MAE often remains finite.

We have taken a *model-based* approach to predictability measurement, and we make inferences about predictability conditional upon a particular fitted model. Alternatively, we could take a *survey-based* approach, based on the predictions of competitive professional forecasters. Conditional upon the assumption that the reported forecasts are optimal, those data can be used for inferences about predictability.

The survey-based approach is of interest because the information sets used by actual forecasters are likely much richer than simple univariate histories. They are surely multivariate, for example, and they also contain hard-to-quantify subjective information. The survey-based approach does rely on a crucial and disputable assumption (optimality of reported forecasts), but so too does the model-based approach (adequacy of the fitted model). The key point is that the assumptions made by the two approaches are very different, so that they naturally complement one another.

A number of relevant surveys exist. They tend to focus on the major macroeconomic aggregates, such as real GDP growth, and the forecast horizons and available samples differ. The former NBER-ASA Quarterly Economic Outlook Survey, now called the Survey of Professional Forecasters and undertaken by the Federal Reserve Bank of Philadelphia, has been maintained since 1968.[25] Although the sample is long, the available forecast horizons are short— zero through four quarters ahead. The Blue Chip Indicators, in contrast, are available only from the early 1980s onward, but the available forecast horizons are substantially longer— zero through six years. We look forward to using these surveys to compute survey-based estimates of predictability and to comparing the survey-based and model-based estimates.

---

[25] Croushore (1993) provides a lucid description.

Data Appendix

All data are from Citibase. The data frequency is quarterly. All data are seasonally adjusted, except for rates.

Code	Description	Source
	Real indicators	
GNPQ	real GNP	Department of Commerce
GDPQ	real GDP	"
GYDQ	real disposable income	"
GYDPCQ	per capita real disposable income	"
GCQ	total private consumption	"
GCNSQ	nondurables and services	"
GCDQ	consumer durables	"
GPIQ	private gross fixed investment	"
GIPDQ	producers' durables	"
GISQ	structures	"
GIRQ	residential	"
GVUQ	change in nonfarm business inventories	"
GGEQ	government purchases of goods & services	"
GGFEQ	federal government	"
GGSEQ	state & local	"
GEXQ	real exports	"
GIMQ	real imports	"
GNETQ	balance of goods and nfs	"
IP	industrial production	Federal Reserve Board
LHOURS	man hours per week	Bureau of Labor Statistics (Household Survey)
LHEM	employment	"
LPNAG	employment	Bureau of Labor Statistics (Establishment Survey)
LHUR	unemployment rate	Bureau of Labor Statistics (Household Survey)
LEW	nominal wage	Bureau of Labor Statistics (Establishment Survey)
LEW77	real wage	"
POP	total population	Bureau of Labor

Statistics

Nominal Indicators

GNP	nominal GNP	Department of Commerce
GDP	nominal GDP	"
FM1	M1	Federal Reserve Board
FM2	M2	"
FM3	M3	"
FMFBA	monetary base	"
FMRNBA	nonborrowed reserves	"

Prices

PUNEW	CPI, urban consumers	Bureau of Labor Statistics
PWFSA	PPI, finished goods	"
PWIMSA	PPI, intermediate goods & supplies	"
PWCMSA	PPI, crude materials	"
GDPD	GDP implicit deflator	Department of Commerce
GNPD	GNP implicit deflator	"
FYGM3	3-month T-bill rate (secondary mkt.)	Federal Reserve Board
FYGT1	1-yr. T-bond rate	"
FYGT3	3-yr. T-bond rate	"
FYGT5	5-yr. T-bond rate	"
FYGT10	10-yr. T-bond rate	"
FYGL	>10-yr. T-bond rate (average)	"
FYFF	fed funds rate	"
FYCD3M	3-month CD rate	"
FYCP90	3-month CP rate	"

Constructed Series

Description	Definition
per capita GNP	PCGNPQ=GNPQ/POP
velocity	VEL=GNP/FM1
inflation	INFL=(GNPD-GNPD(-1))/GNPD(-1)
private consumption of nondurables and services	GCNSQ=GCNQ+GCSQ
nonresidential investment	GINQ=GIPDQ+GISQ

References

Ball, L. and Cecchetti, S.G. (1990), "Inflation and Uncertainty at Short and Long Time Horizons," *Brookings Papers on Economic Activity*, 215-245.

Barsky, R.B. (1987), "The Fisher Hypothesis and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics*, 19, 3-24.

Bhansali, R.J. (1992), "Autoregressive Estimation of the Prediction Mean Squared Error and an R^2 Measure: An Application," in D. Brillinger *et al*. (eds.), *New Directions in Time Series Analysis, I.* New York: Springer-Verlag, 9-24.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997), *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Cecchetti, S.G. (1995), "Inflation Indicators and Inflation Policy," *NBER Macroeconomics Annual 1995*, 189-219.

Christoffersen, P.C. and Diebold, F.X. (1996), "Further Results on Forecasting and Model Selection Under Asymmetric Loss," *Journal of Applied Econometrics*, 11, 561-572.

Christoffersen, P.C. and Diebold, F.X. (1997), "Optimal Prediction Under Asymmetric Loss," National Bureau of Economic Research Technical Working Paper No. 167. Forthcoming in *Econometric Theory*.

Cogley, T. and Nason, J.M. (1995a), "Output Dynamics in Real Business Cycle Models," *American Economic Review*, 85, 492.

Cogley, T. and Nason, J.M. (1995b), "Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research," *Journal of Economic Dynamics and Control*, 9, 253-278.

Cooley, T.F. and Prescott, E.C. (1995), "Economic Growth and Business Cycles," in T.F. Cooley (ed), *Frontiers of Business Cycle Research*. Princeton: Princeton University Press, 1995.

Croushore, D. (1993), "The Survey of Professional Forecasters," *Business Review*, Federal Reserve Bank of Philadelphia, November-December.

DeJong, D. and Whiteman, C.H. (1994), "The Forecasting Attributes of Trend- and Difference-Stationary Representations for Macroeconomic Time Series," *Journal of Forecasting*, 13, 279-298.

Diebold, F.X. and Mariano, R.S. (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-265.

Diebold, F.X. and Nerlove, M. (1990), "Unit Roots in Economic Time Series: A Selective Survey," in T.B. Fomby and G.F. Rhodes (eds.) *Advances in Econometrics: Co-Integration, Spurious Regressions, and Unit Roots*, 3-69. Greenwich, Connecticut: JAI Press.

Diebold, F.X., Ohanian, L. and Berkowitz, J. (1995), "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," National Bureau of Economic Research, Technical Working Paper No. 174. (Revised March 1997.)

Diebold, F.X. and Senhadji, A. (1996), "Deterministic vs. Stochastic Trend in U.S. GNP, Yet Again," *American Economic Review*, 86, 1291-1298.

Efron, B. and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Fama, E.F. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, 25, 383-417.

Fama, E.F. (1991), Efficient Capital Markets: II," *Journal of Finance*, 46, 1575-1617.

Forni, M. and Reichlin, L. (1995), "Let's Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycles," Manuscript, Department of Economics, Free University of Brussels.

Granger, C.W.J. and Newbold, P. (1986), *Forecasting Economic Time Series*, Second Edition. Orlando: Academic Press.

Grenander, U. (1981), *Abstract Inference*. New York: John Wiley.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hannan, E.J. and Poskitt, D.S. (1988), "Unit Canonical Correlations Between Future and Past," *Annals of Statistics*, 16, 7894-790.

Hansen, G.D. (1985), "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics*, 16, 309-327.

Harvey, A. and Jaeger, A. (1993), "Detrending, Stylized Facts and the Business Cycle," *Journal of Applied Econometrics*, 8, 231-248.

Jewell, N.P. and Bloomfield, P. (1983), "Canonical Correlations of Past and Future for Time Series: Definitions and Theory," *Annals of Statistics*, 11, 837-847.

Jewell, N.P., Bloomfield, P. and Bartmann, F.C. (1983), "Canonical Correlations of Past and

Future for Time Series: Bounds and Computation," *Annals of Statistics*, 11, 848-855.

Kilian, L. (1995), "Small-Sample Confidence Intervals for Impulse Response Functions," manuscript, Department of Economics, University of Pennsylvania.

Kilian, L. (1996a), "Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order," manuscript, Department of Economics, University of Pennsylvania.

Kilian, L. (1996b), "An Endogenous Lag Order Bootstrap Algorithm with an Application to Impulse Response Analysis," manuscript, Department of Economics, University of Pennsylvania.

Kilian, L. (1996c), "Impulse Response Intervals under Departures from Normality," manuscript, Department of Economics, University of Pennsylvania.

King, R.G., Plosser, C.I. and Rebelo, S.T. (1988), "Production, Growth and Business Cycles. I. The Basic Neoclassical Model," *Journal of Monetary Economics*, 21,195-232.

King, R.G. and Watson, M.W. (1996), "Money, Prices, Interest Rates and the Business Cycle," *Review of Economics and Statistics*, 78, 35-53.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*. New York: Springer-Verlag.

Mohanty, R. and Pourahmadi, M. (1996), "Estimation of the Generalized Prediction Error Variance of a Multiple Time Series," *Journal of the American Statistical Association*, 91, 294-299.

Nelson, C.R. and Plosser, C.I. (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139-162.

Nicholls, D.F. and A.L. Pope (1988), "Bias in the Estimation of Multivariate Autoregressions," *Australian Journal of Statistics*, 30A, 296-309.

Pope, A.L. (1990), "Biases of Estimators in Multivariate Non-Gaussian Autoregressions," *Journal of Time Series Analysis*, 11, 249-258.

Rotemberg, J.J. and Woodford, M. (1996), "Real Business-Cycle Models and the Forecastable Movements in Output, Hours, and Consumption," *American Economic Review*, 86, 71-89.

Runkle, D.E. (1987), "Vector Autoregression and Reality," *Journal of Business and Economic Statistics*, 5, 437-442.

Singleton, K.J. (1988), "Econometric Issues in the Analysis of Equilibrium Business Cycle Models," *Journal of Monetary Economics*, 21, 361-386.

Stine, R. A. (1987), "Estimating Properties of Autoregressive Forecasts," *Journal of the American Statistical Association*, 82, 1072-1078.

Stock, J.H. (1995), "Unit Roots and Trend Breaks," in R. Engle and D. McFadden (eds.), *Handbook of Econometrics, Volume 4*. Amsterdam: North-Holland.

Theil, H. (1966), *Applied Economic Forecasting*. Amsterdam: North-Holland.