DISCUSSION PAPER SERIES

No. 2424

MEASURING PREDICTABILITY: THEORY AND MACROECONOMIC APPLICATIONS

Francis X Diebold and Lutz Kilian

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

CEPR Discussion Papers are now available online at

www.cepr.org

MEASURING PREDICTABILITY: THEORY AND MACROECONOMIC APPLICATIONS

Francis X Diebold, Stern School of Business, New York University
Lutz Kilian, University of Michigan and CEPR

ABSTRACT

Measuring Predictability: Theory And Macroeconomic Applications*

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 covariance stationary or difference stationary processes. We propose a simple estimator, and we suggest resampling methods for inference. We then provide several macroeconomic applications. First, we illustrate the implementation of predictability measures based on fitted parametric models for several US macroeconomic time series. Second, we analyse the internal propagation mechanism of a standard dynamic macroeconomic model by comparing the 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 outline several nonparametric extensions of our approach.

Francis X Diebold
Department of Finance
Stern School of Business
New York University
44 West 4th Street,
K-MEC, Suite 9-190
New York NY 10012-1126
USA
Tel: (1 215) 998 0799
Fax: (1 215) 998 0351
Email: fdiebold@stern.nyu.edu

Lutz Kilian
Department of Economics
Lorch Hall, M-109
University of Michigan
Ann Arbor, MI 48109-1220
USA
Tel: (1 734) 764 2320
Fax: (1 734) 764 2769
Email: lkilian@umich.edu

NON-TECHNICAL SUMMARY

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 the past conveys regarding future values of this series; 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 behaviour of economic series, as well as for assessing agreement between economic models and data.

Remarkably little attention has been paid to methods for measuring predictability. Existing methods include those based on canonical correlation between past and future, and those based on comparing the innovation variance and unconditional variance of stationary series. Those methods, however, are inadequate in light of recent work stressing the possible presence of unit roots, rich and high-dimensional information sets, non-quadratic and possibly even asymmetric loss functions, and variations in forecast accuracy across horizons.

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. 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.

In this paper, we contribute to the theory of predictability measurement and apply our results in several macroeconomic contexts. 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 covariance stationary or difference stationary processes, univariate

or multivariate information sets, general loss functions, and different forecast horizons of interest. We propose parametric methods for estimating the predictability of observed series, and we suggest using bootstrap methods for inference.

We illustrate the implementation of predictability measures based on fitted parametric models for several US macroeconomic time series. We further illustrate the use of these parametric predictability measures in assessing the propagation mechanisms of dynamic macroeconomic models and in assessing agreement between economic models and the data. Our results also shed additional light on the implications of alternative detrending methods. Finally, we discuss alternative nonparametric predictability measures for financial asset returns and volatilities and survey-based estimation of predictability.

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 the past conveys regarding future values of this series; 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 the possible presence of unit roots, rich and high-dimensional information sets, non-quadratic and possibly

---

[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 (1998).

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

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 covariance stationary or difference stationary processes, univariate or multivariate information sets, general loss functions, and different forecast horizons of interest. In Section 3, we propose parametric methods for estimating the predictability of observed series, and we suggest using bootstrap methods for inference. We also illustrate the implementation of predictability measures based on fitted parametric models for several U.S.

---

[3] See, among others, Stock (1995), Diebold and Mariano (1995), Christoffersen and Diebold (1996, 1997) and Forni and Reichlin (1998).

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

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

macroeconomic time series. In Section 4, we illustrate the use of these parametric predictability

measures in assessing the propagation mechanisms of economic models, and in assessing

agreement between economic models and data.[6] Section 5 discusses alternative nonparametric

measures of predictability and directions for future research.

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 the past conveys regarding the

future values of this series; given an information set, some processes are inherently easy to

forecast, and others are difficult. This point is not purely academic. For example, a policy-

maker 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

---

[6] The exercises we undertake in section 4 are similar in spirit to Rotemberg and Woodford (1996). Rotemberg and Woodford's primary concern is with a certain class of real 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 macro-econometric methods of assessing predictability that can be applied in a variety of situations, under minimal assumptions.

[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.

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 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,

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

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-Newbold measure with $j = 1$ has been used, for example, by Barsky (1987) to assess the predictability of inflation.[8]

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

---

[8] Essentially the same measure of predictability has been proposed by Nelson (1976). The Granger-Newbold measure also is formally identical to Parzen's (1982) prediction variance horizon. Öller (1985) uses the Parzen measure to capture the information content of forecasts.

$E\big(L(e_{t+k,t})\big)$, j«k.[9] If $E\big(L(e_{t+j,t})\big)$ « $E\big(L(e_{t+k,t})\big)$, we say that the series is highly predictable at horizon j relative to k, and if $E\big(L(e_{t+j,t})\big)$ ≈ $E\big(L(e_{t+k,t})\big)$, 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=∞. The advantages of our generalization include:

(1) It is valid for both covariance stationary and difference stationary series, so long as k«∞.

(2) It allows for general loss functions. The loss function $L(\cdot)$ need not be quadratic, and need not even be symmetric; we only require that $L(0) = 0$ and that $L(\cdot)$ be strictly monotone on each side of the origin.[10] By the restrictions imposed on $L(\cdot)$, we have that for all covariance stationary or difference stationary processes $P(L(\cdot),\Omega,j,k) \in [0, 1]$, with larger values indicating greater predictability.

(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

---

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

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

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

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

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,

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

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 $P(e^2,\text{univariate},j,k) = 1 - \dfrac{j}{k}$, as will be shown later. But for a random walk the variance ratio at horizon j, a common measure of persistence, is[11]

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

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. To keep the discussion tractable, and in keeping with the empirical analysis of subsequent sections, we postulate a quadratic loss function $L(e)=e^2$ for estimation, prediction, model selection, and construction of

---

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

predictability measures.[12] It is clear that parametric measures of the degree of predictability of a

time series in general will depend on the specification of the parametric model. Here we focus

on univariate autoregressive models, although one could easily generalize the discussion to other

parametric models, such as vector ARMA models.[13]

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. In practice, we estimate P by replacing the underlying

unknown parameters by their least squares estimates. The legitimacy of this predictability

measure is invariant to the possible presence of unit roots in the autoregressive lag order

polynomial. To determine the autoregressive lag order we use the Akaike Information Criterion

(AIC) with a suitable upper bound on the admissible lag orders. The AIC is less likely to

underestimate the lag order in small samples than alternative criteria. The latter property is

---

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

[13] In section 5 we discuss alternative nonparametric implementations of the predictability measure.

crucial in preserving the higher-order dynamics implicit in P (see Kilian 1999).

Throughout this paper, inference is based on the bias-corrected bootstrap method proposed by Kilian 1998 (a,b). We account for lag order uncertainty along the lines suggested by Kilian (1998c). Recent theoretical results by Inoue and Kilian (1999a) suggest that - with the exception of some random walk models - the method of bootstrap inference we propose is asymptotically valid both for stationary and for unit root processes. In Figures 1-3, we report the results of a small Monte Carlo experiment designed to help assess the adequacy of our bootstrap intervals for the predictability measure. The results shown are for univariate information sets, quadratic loss, $k = 40$ and $j = 1, ..., 16$. 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 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 20% for T=80, and by up to 17% for T=160. For T=480 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.

We now turn to an illustrative example. Figure 4 shows point and interval estimates for the predictability of several post-war quarterly U.S. investment series from the national income and product accounts. For expository purposes, we use univariate information sets and fix $L(e) = e^2$ and k=40, as we vary the near-term forecast horizon j.[14] We plot P (in percent) against near-term forecast horizons j = 1, ..., 20. Higher values of P indicate greater predictability. All data are in logs and seasonally adjusted. We model the data as level autoregressions and, with the exception of the inventory series, allow for a linear time trend.

Private investment spending is one of the least predictable components of real expenditures. Although predictability is close to 80% at a horizon of one quarter, it drops sharply at higher horizons. Our estimates suggest that the predictability of private investment one year into the future is merely 20%. Beyond a horizon of two years it is essentially zero. Broken down by component, changes in inventories of nonfarm businesses are least predictable, followed by residential and nonresidential investment, in that order. Within nonresidential investment spending, however, structures are much more predictable than producer durables. This example highlights the fact that there are important differences in predictability across expenditure components that are potentially useful for macroeconomic modeling. For example, changes in inventories are virtually unpredictable for horizons in excess of one year. In contrast, the predictability of investment in structures remains in excess of 20% even after five years.

We also show the corresponding nominal 90% bootstrap confidence intervals. The confidence intervals often are rather wide, reflecting the limited information available in

---

[14] 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.

macroeconomic data, but not so wide as to render the exercise futile. Moreover, the degree of

sampling uncertainty may differ substantially across investment components. For example, the

predictability of structures is rather uncertain, whereas that of inventories is very precisely

estimated, especially at higher horizons.

4. 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, $0 < \beta < 1$, $0 \leq \delta \leq 1$, 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. All data are quarterly.

4.1. 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 a univariate 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 5, in which we show the comparative predictability of the

technology shock and the model outputs. We again fix $L(e) = e^2$ and set k=40, as we vary the

near-term forecast horizon j = 1, ..., 20. Figure 5 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 and hours are less predictable

than the technology shock, and consumption, productivity and the capital stock in the model are

more predictable than the technology shock.

These 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 input 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 judgement 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 6.

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: with the possible exception of the capital stock, 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 common predictability pattern of HP-filtered model

outputs is an artifact of HP filtering.[15] To explore that conjecture, we compute the predictability

---

[15] 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).

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 7: for all realistic values of γ, the predictability of the HP-filtered technology shock looks the same as the predictability of the typical HP-filtered model output! 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.

4.2. 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 8 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 the model data are 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 actual 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 are 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 9 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 9

show a pronounced divergence: only model output, investment, consumption, and capital stock

match the data; hours worked in the model are not nearly predictable enough, whereas

productivity in the model is too predictable for the first three years. This result is robust to

whether a trend is fitted to the model data, or the population trend of zero is imposed.

5. 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

estimates, will be of interest. To take a second example, we have allowed 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.[16] Nevertheless, it will be of interest to

explore models with nonlinear deterministic trends and models that allow for stochastic trends

only.

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

interesting. We shall briefly discuss three.

5.1. 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.[17]

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

---

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

[17] The asymptotic validity for the stationary case of bootstrapping predictability measures
based on the sieve approximation has been established by Inoue and Kilian (1999b).

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,

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

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.

5.2. Predictability of Financial Asset Returns and Volatilities

We have focused on the application of predictability measurement to macroeconomics.

Predictability statistics should also prove useful in finance, in which the predictability of asset

returns at various horizons is a central concern. In fact, if non-predictability 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 non-quadratic loss

functions.[18]

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 volatility predictability patterns, and in particular the speed and pattern with which volatility predictability decays as the horizon grows.

5.3. Survey-Based Predictability Estimation

We have taken a *model-based* approach to predictability measurement. Conditional upon a particular fitted model, we make inferences about predictability. 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.

---

[18] 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.

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.[19] 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.

ACKNOWLEDGMENTS

---

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

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., A.W. Lo and A.C. MacKinlay (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 F.X. Diebold (1996), 'Further Results on Forecasting and Model
    Selection Under Asymmetric Loss', *Journal of Applied Econometrics*, **11**, 561-572.

Christoffersen, P.C. and F.X. Diebold (1997), 'Optimal Prediction Under Asymmetric Loss',
    Econometric Theory, **13**, 808-817.

Cogley, T. and J.M. Nason (1995a), 'Output Dynamics in Real Business Cycle Models',
    American Economic Review, **85**, 492.

Cogley, T. and J.M. Nason (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 E.C. Prescott (1995), 'Economic Growth and Business Cycles', in T.F. Cooley
    (ed), *Frontiers of Business Cycle Research*, Princeton: Princeton University Press.

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

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

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

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

Diebold, F.X., L.E. Ohanian and J. Berkowitz (1998), 'Dynamic Equilibrium Economies: A Framework for Comparing Models and Data', *Review of Economic Studies*, **65**, 433-451.

Diebold, F.X. and A. Senhadji (1996), 'The Uncertain Unit Root in Real GNP: Comment', *American Economic Review*, **86**, 1291-1298.

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 L. Reichlin (1998), 'Let's Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycles', *Review of Economic Studies*, **65**, 453-473.

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

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

Hannan, E.J. and D.S. Poskitt (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 A. Jaeger (1993), 'Detrending, Stylized Facts and the Business Cycle', *Journal of Applied Econometrics*, **8**, 231-248.

Inoue, A. and L. Kilian (1999a), 'Bootstrapping Autoregressive Processes with Possible Unit Roots', Working Paper No. 99-09, Department of Economics, University of Michigan.

Inoue, A. and L. Kilian (1999b), 'Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in VAR(∞) Models', Working Paper No. 99-03, Department of Economics, University of Michigan.

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

Jewell, N.P., P. Bloomfield and F.C. Bartmann (1983), 'Canonical Correlations of Past and Future for Time Series: Bounds and Computation', *Annals of Statistics*, **11**, 848-855.

Kilian, L. (1998a), 'Small-Sample Confidence Intervals for Impulse Response Functions', *Review of Economics and Statistics*, **80**, 218-230.

Kilian, L. (1998b), 'Confidence Intervals for Impulse Responses under Departures from Normality', *Econometric Reviews*, **17**, 1-29.

Kilian, L. (1998c), 'Accounting for Lag Order Uncertainty in Autoregressions: The Endogenous Lag Order Bootstrap Algorithm', *Journal of Time Series Analysis*, **19**, 531-548.

Kilian, L. (1999), 'Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order', forthcoming: *Journal of Forecasting*.

King, R.G., C.I. Plosser and S.T. Rebelo (1988), 'Production, Growth and Business Cycles. I. The Basic Neoclassical Model', *Journal of Monetary Economics*, **21**,195-232.

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

Mohanty , R. and M. Pourahmadi (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. (1976), 'The Interpretation of R^2 in Autoregressive-Moving Average Time Series Models', *The American Statistician*, **30**, 175-180.

Öller, L.E. (1985), 'How Far Can Changes in General Business Activity Be Forecasted?', *International Journal of Forecasting*, **1**, 135-141.

Parzen, E. (1982), 'ARARMA Models for Time Series Analysis and Forecasting', *Journal of Forecasting*, **1**, 67-82.

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

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

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

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

Figure 1

Effective Coverage of Nominal 90 % Bootstrap Confidence Intervals
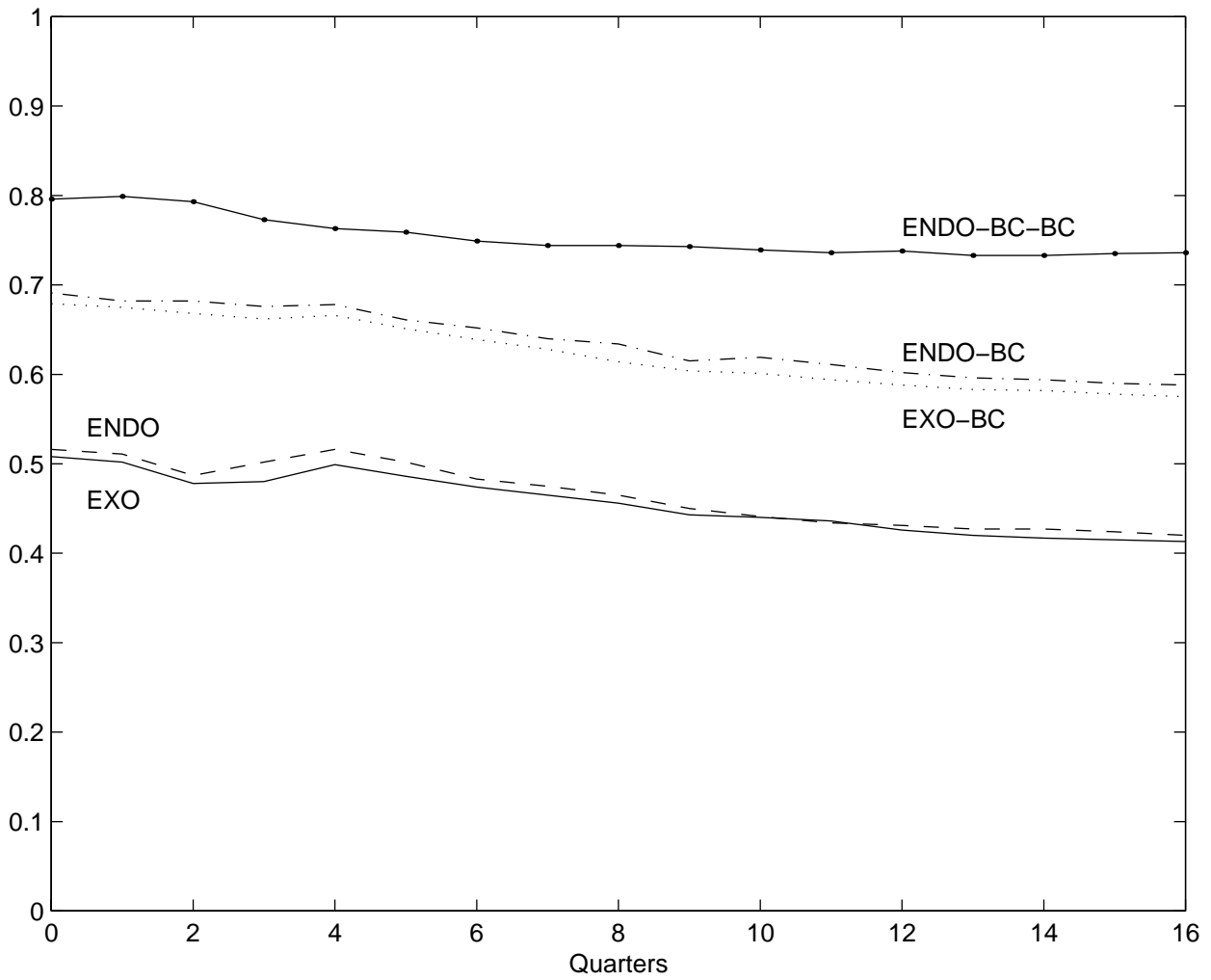
Sample Size 80
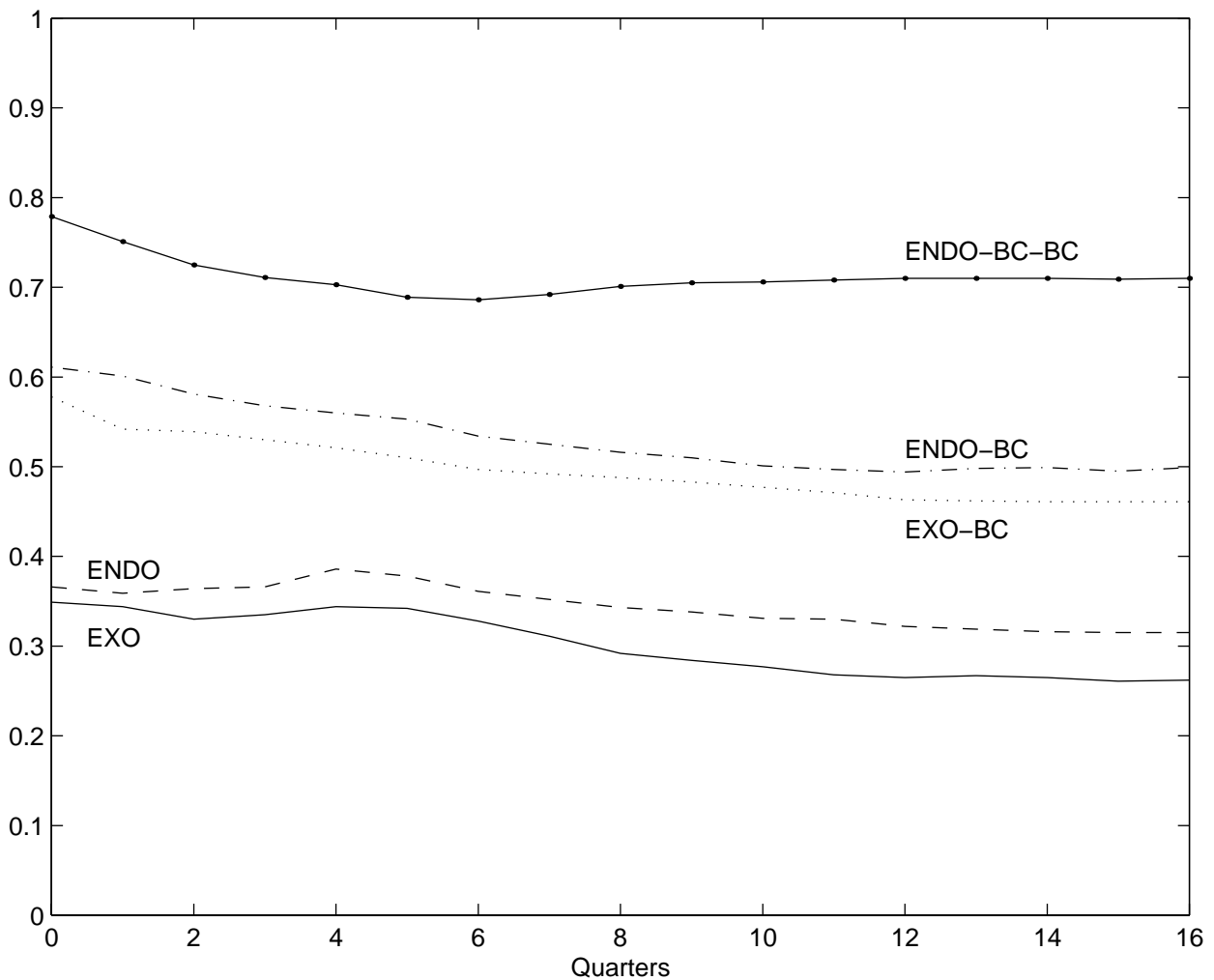
NOTES:

Monte Carlo results based on 1,000 trials with 1,000 bootstrap replications each. The DGP is based on an AR(6) approximation of the quarterly U.S. Treasury Bond Rate. In the Monte Carlo study the lag order is assumed unknown. Lag order selection is based on the AIC with $1 \leq p \leq 8$. We consider re-estimating the lag order in each bootstrap replication (*endo*) or fixing it at the value of the initial lag order estimate (*exo*). *BC* refers to bootstrap resamples conditional on bias-corrected autoregressive coefficient estimates. The *BC-BC* interval in addition corrects for bias in the bootstrap estimates as discussed in Kilian (1998).

Figure 2

Effective Coverage of Nominal 90 % Bootstrap Confidence Intervals

Sample Size 160

NOTES:

See Figure 1.

Figure 3

Effective Coverage of Nominal 90 % Bootstrap Confidence Intervals
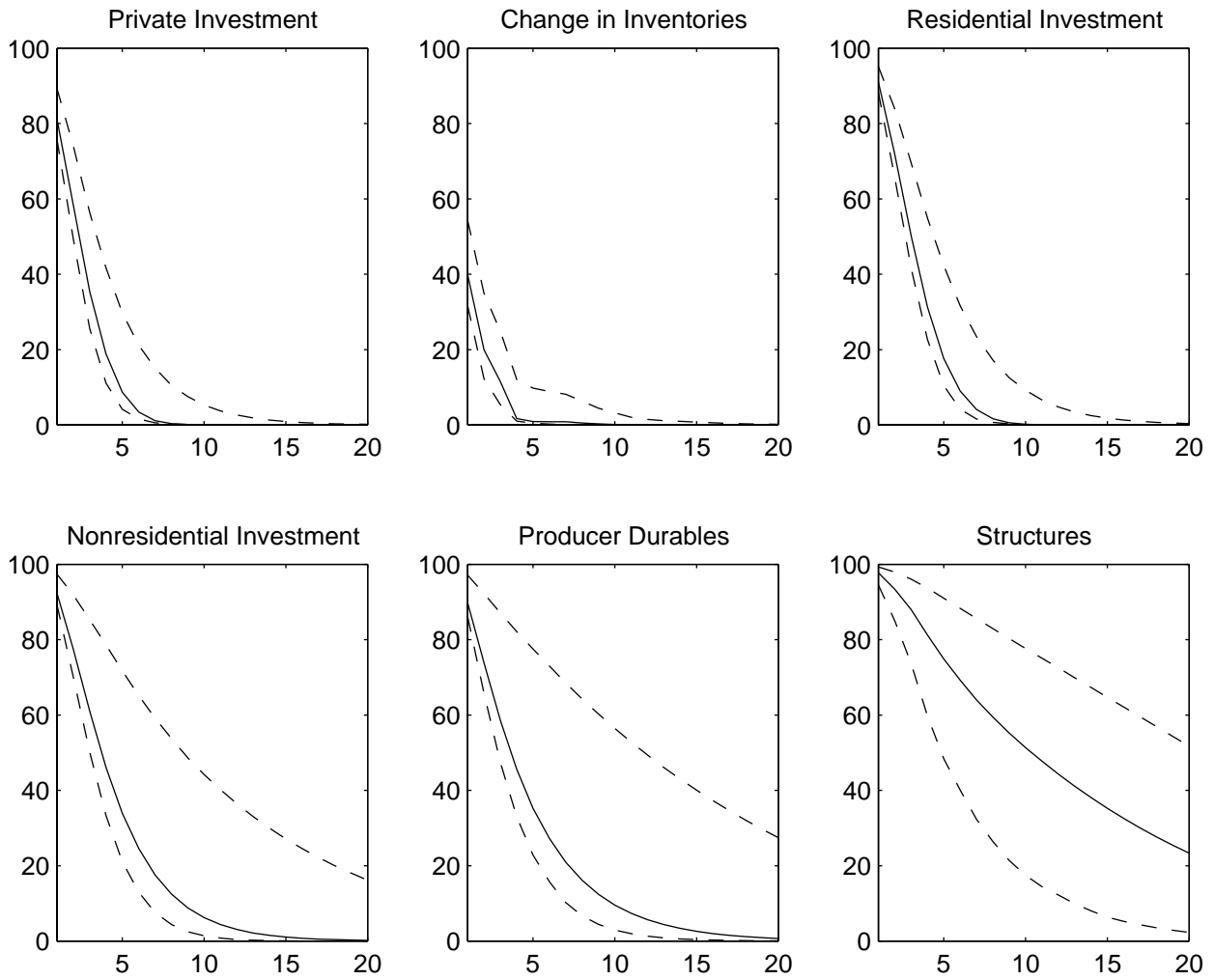
Sample Size 480
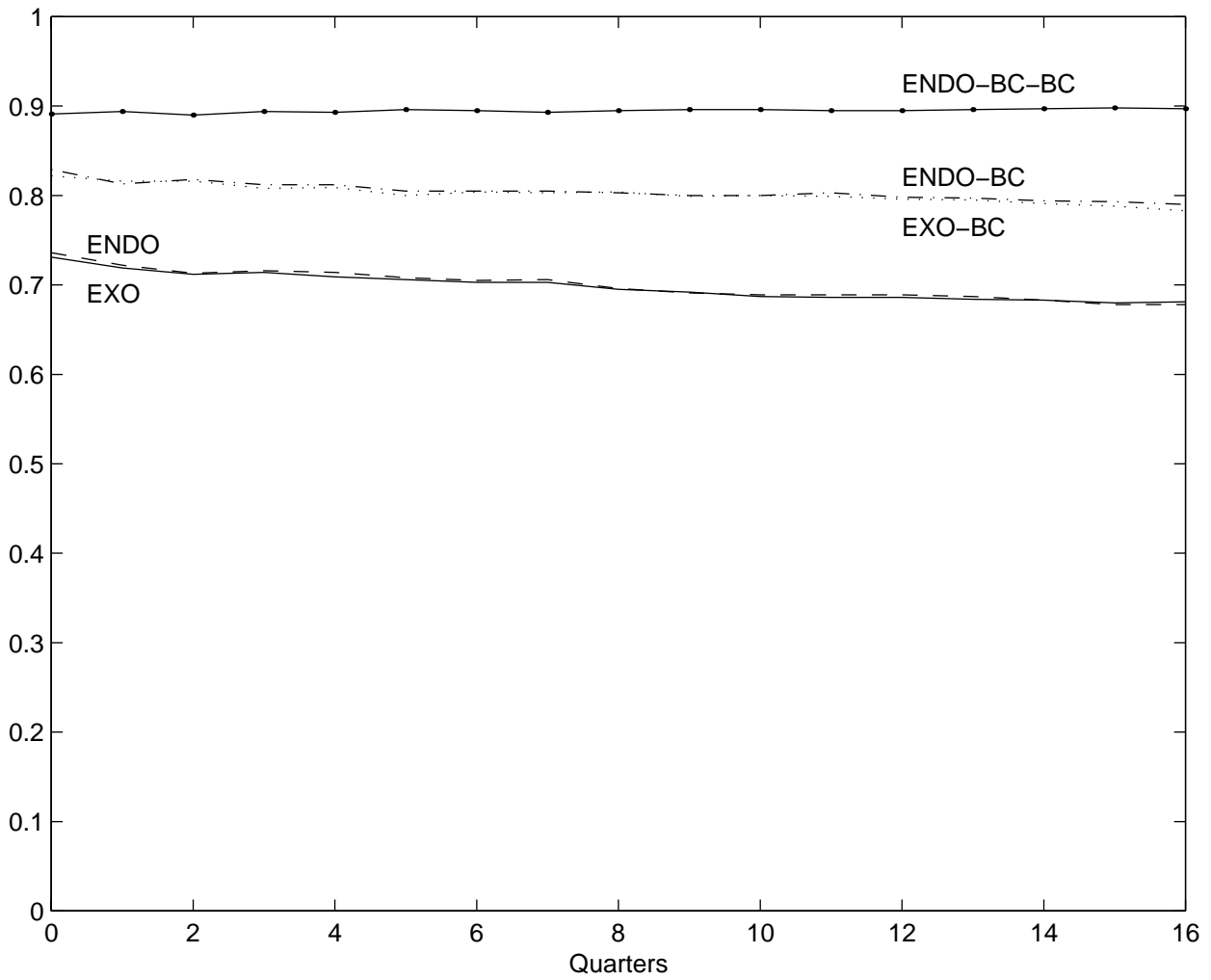
NOTES:

See Figure 1.

Figure 4

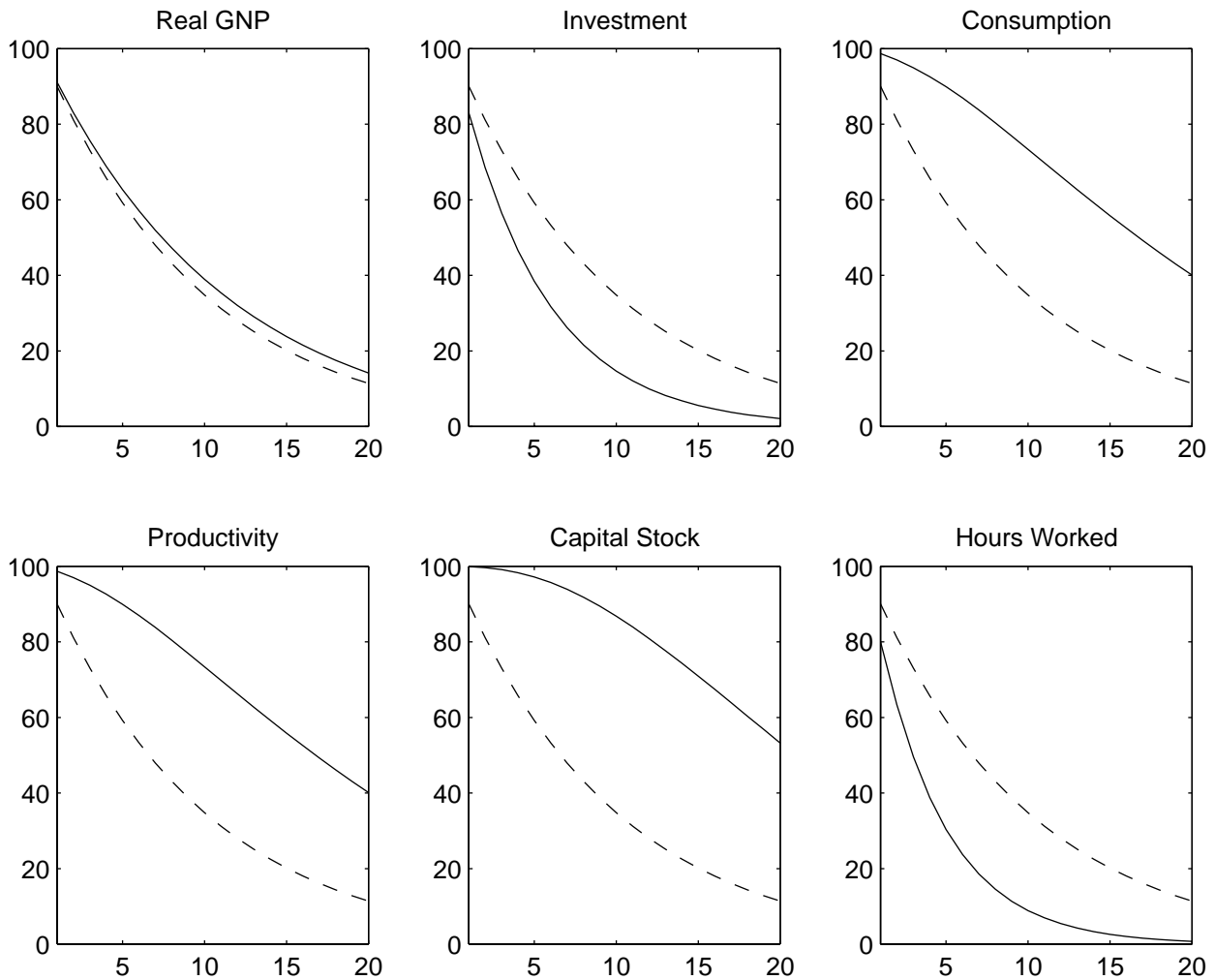# Predictability Estimates with 90 % Confidence Intervals

NOTES:

All data are from CitiBase. The sample period is 1947.II-1994.III. For a description of the methodology see text. The plots are based on k = 40 and j = 1, ..., 20.

Figure 5

The Propagation Mechanism of the Hansen (1985) Model:
Model Input (dashed line) - Model Output (solid line)

Predictability in Levels without Filtering

NOTES:

For a detailed description of the model and its parameterization see Hansen (1985). For a description of the methodology see text. The plots are based on k = 40 and j = 1, ..., 20.

Figure 6

The Propagation Mechanism of the Hansen (1985) Model: Model Input (dashed line) - Model Output (solid line)

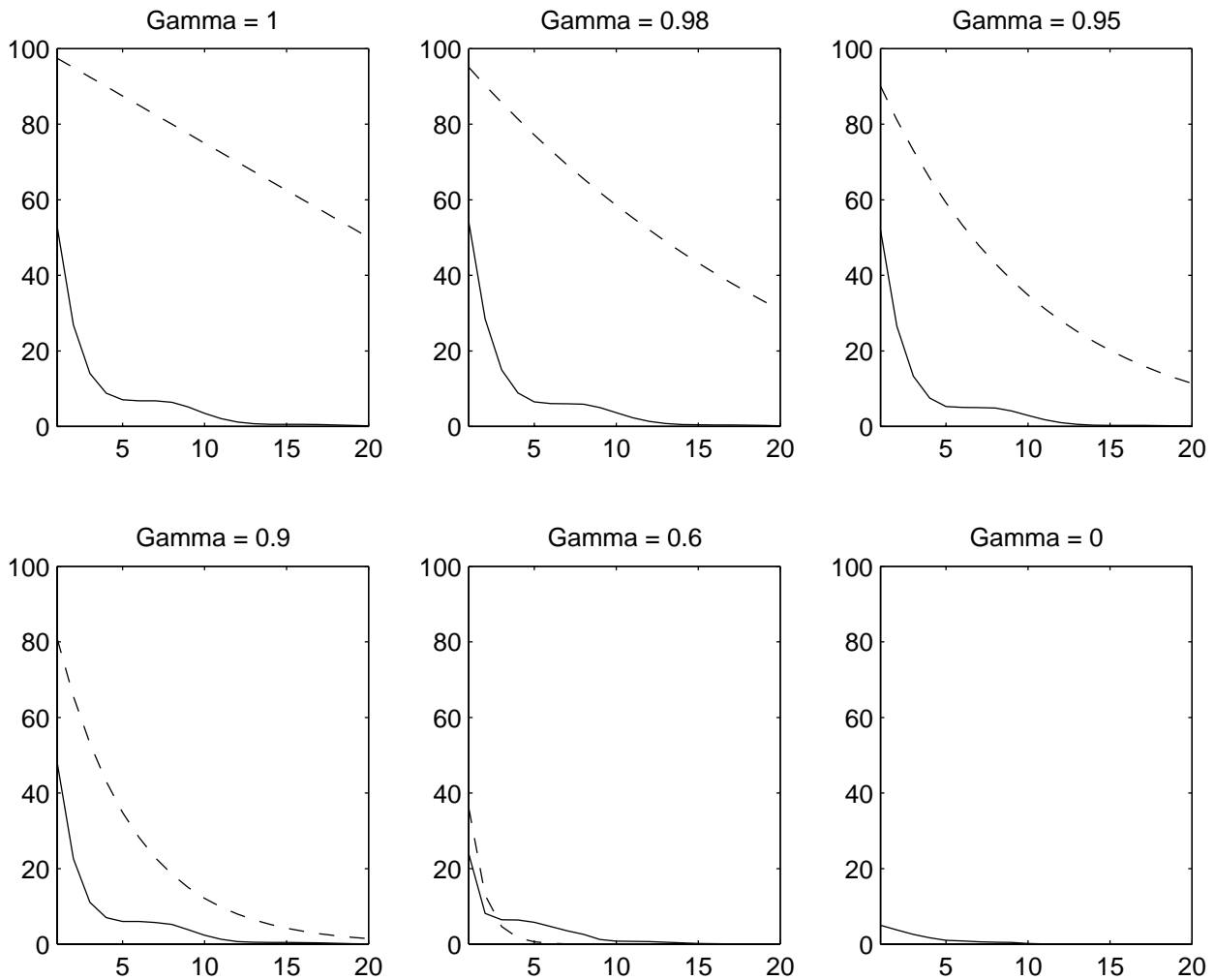## Predictability after Logging and HP-Filtering the Model Output Data
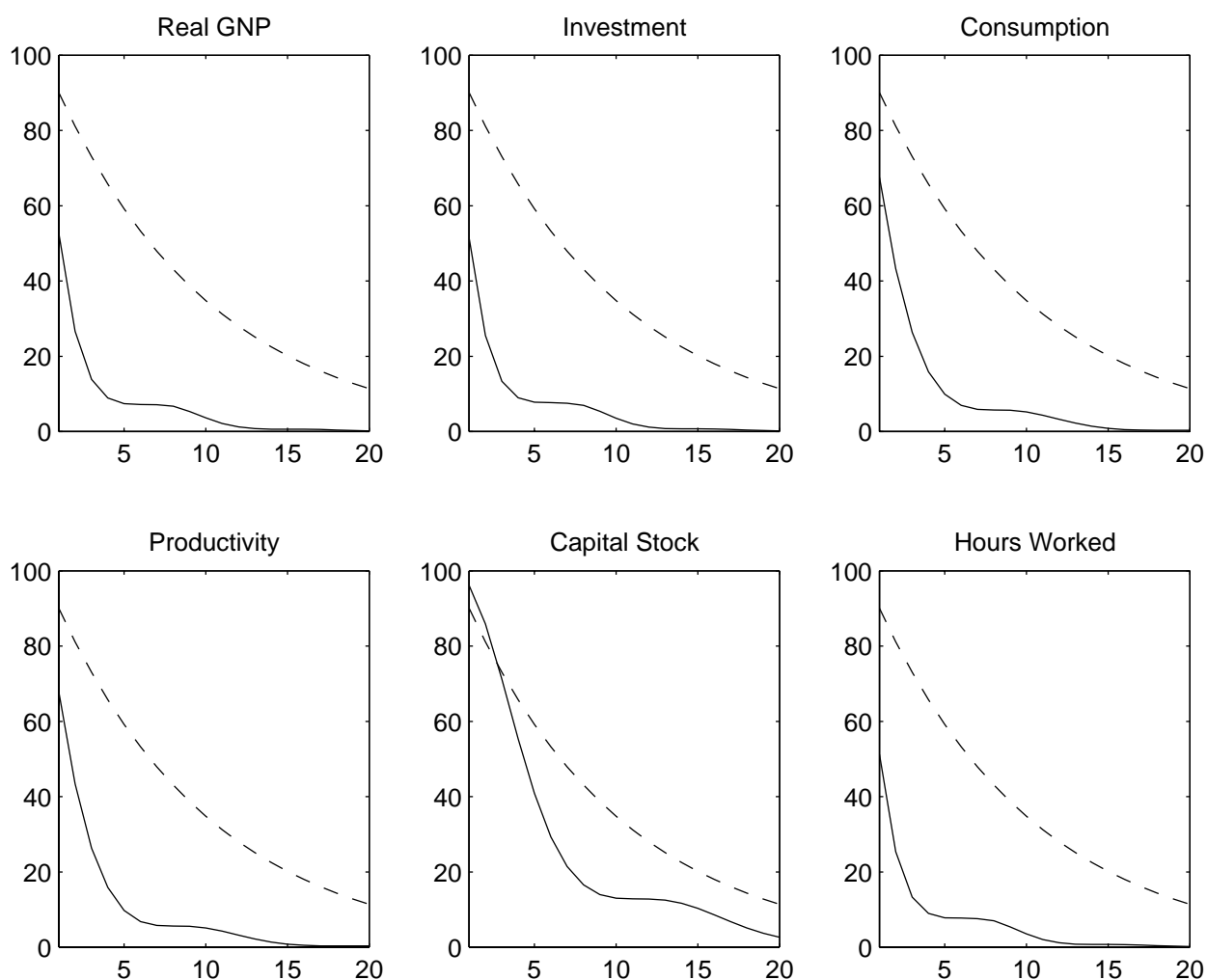
NOTES:

See Figure 5.

Figure 7

Predictability of the AR(1) Technology Shock:
Raw Data (dashed line) - HP-Filtered Data (solid line)
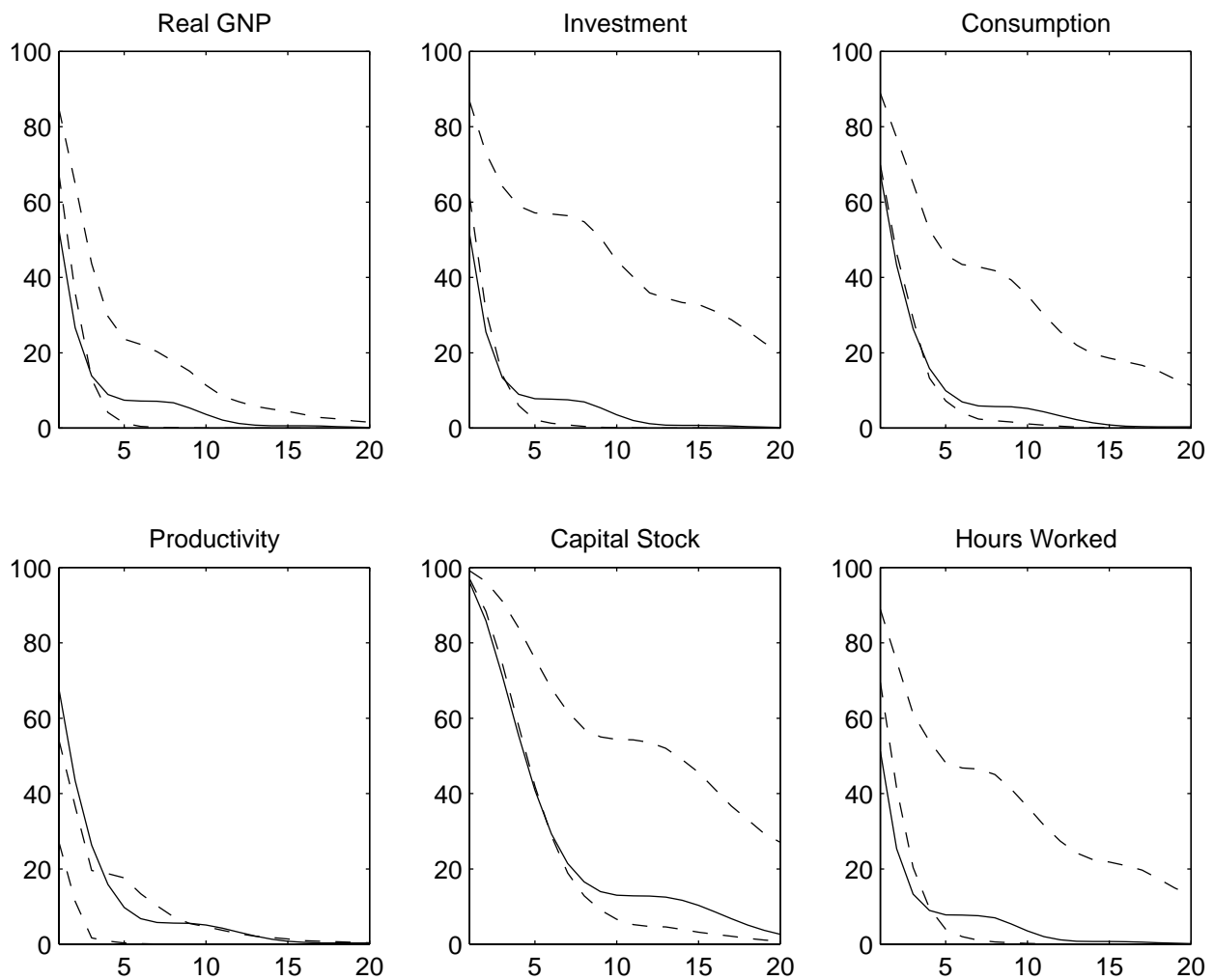
Alternative Degrees of Persistence

NOTES:

For a desription of the methodology see text. The plots are based on k = 40 and j = 1, ..., 20.

Figure 8

Model Performance in Terms of Predictability:
U.S. Data (dashed lines) - Model Data (solid line)
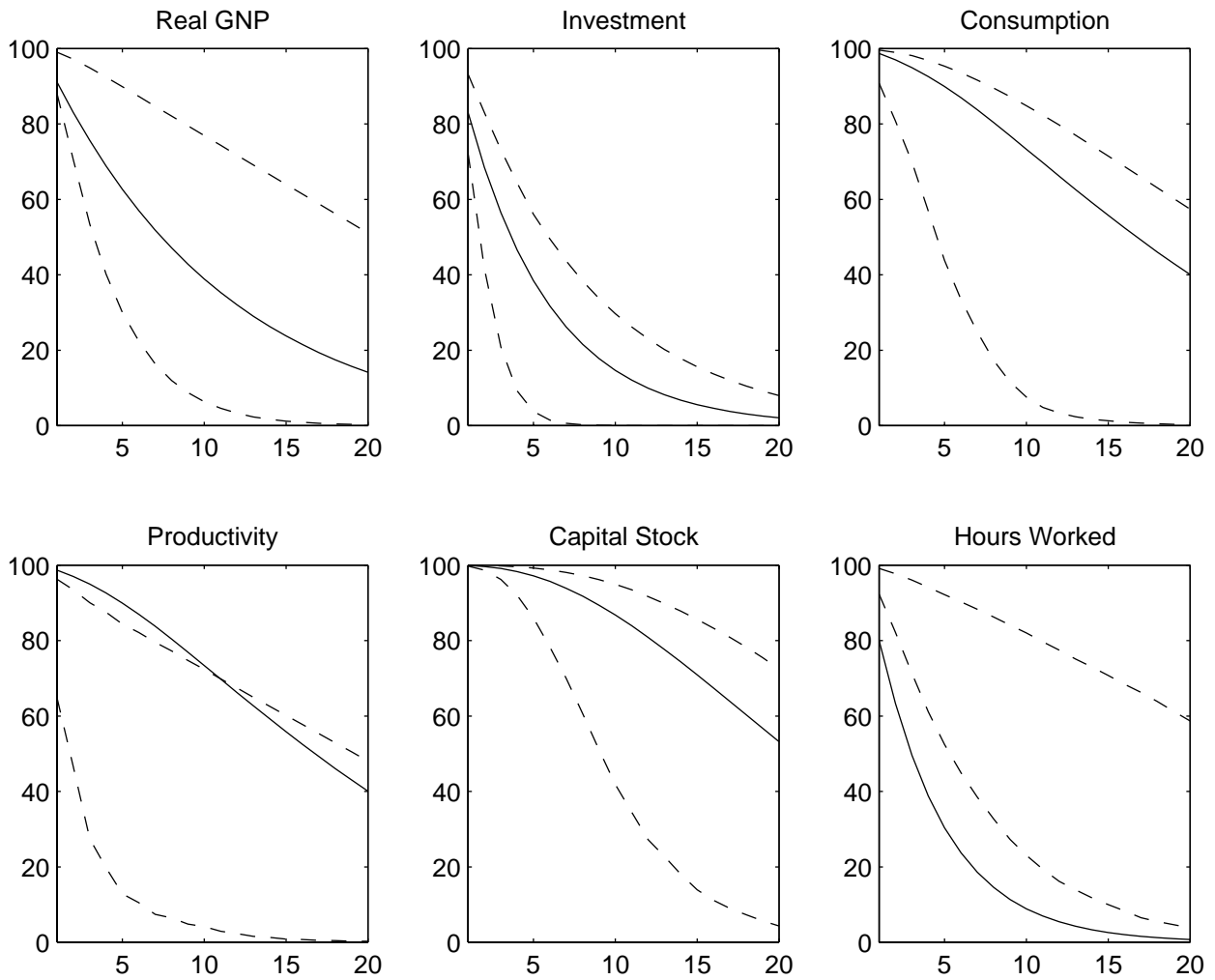
All Data Logged and HP-Filtered

NOTES:

For a detailed description of the model and its parameterization see Hansen (1985). All data are from Hansen (1985). For a description of the methodology see text. The plots are based on k = 40 and j = 1, ..., 20. The dashed lines are the 90 percent confidence intervals for the U.S. estimates.

Figure 9

Model Performance in Terms of Predictability:
U.S. Data (dashed lines) - Model Data (solid line)

All Data Logged and Linearly Detrended

NOTES:

See Figure 8.