

Mixing Frequencies: Stock Returns as a Predictor of Real Output Growth

Anthony S. Tay
December 2006

Paper No. 34-2006

Mixing Frequencies: Stock Returns as a Predictor of Real Output Growth

Anthony S. Tay

School of Economics and Social Sciences

Singapore Management University

90 Stamford Road, Singapore 178903

email: anthonytay@smu.edu.sg

December 2006

Abstract

We investigate two methods for using daily stock returns to forecast, and update forecasts of, quarterly real output growth. Both methods aggregate daily returns in some manner to form a single stock market variable. We consider (i) augmenting the quarterly AR(1) model for real output growth with daily returns using a nonparametric Mixed Data Sampling (MIDAS) setting, and (ii) augmenting the quarterly AR(1) model with the most recent r -day returns as an additional predictor. We find that our mixed frequency models perform well in forecasting real output growth.

Keywords: Forecasting, Mixed Data Sampling, Functional linear regression, Test for Superior Predictive Ability.

JEL codes: C53, E37

Acknowledgements: The author is grateful for financial support from the SMU Research Grant 05-C208-SMU-008 “Optimal Use of Economic Data Sampled at Different Frequencies”. The author also thanks Peter Hansen for the use of with his SPA programs. Zhang Xuefeng and Fred Yan-Qing Goh provided valuable research assistance.

Mixing Frequencies: Stock Returns as a Predictor of Real Output Growth

December 2006

Abstract

We investigate two methods for using daily stock returns to forecast, and update forecasts of, quarterly real output growth. Both methods aggregate daily returns in some manner to form a single stock market variable. We consider (i) augmenting the quarterly AR(1) model for real output growth with daily returns using a nonparametric Mixed Data Sampling (MIDAS) setting, and (ii) augmenting the quarterly AR(1) model with the most recent r -day returns as an additional predictor. We find that our mixed frequency models perform well in forecasting real output growth.

Keywords: Forecasting, Mixed Data Sampling, Functional linear regression, Test for Superior Predictive Ability.

JEL codes: C53, E37

1. Introduction

The fact that financial variables are available at high frequencies, and quickly, raises two related questions pertinent to their use in forecasting macroeconomic variables, which are only available at much lower frequencies, and are much less timely. First, is it gainful to use the financial variables at the higher frequencies, instead of aggregating down to the frequency of the macroeconomic variables? Second, can high-frequency financial data be used to update macroeconomic forecasts as and when the financial data become available? We explore these questions for the particular case where daily stock market returns are used to forecast, and update forecasts of, quarterly real output growth.

The extant out-of-sample evidence on stock returns as a predictor of output growth is actually rather weak, despite the traditional view of stock returns as a leading indicator of economic activity (see Stock and Watson 2003, for a comprehensive survey of the use of asset prices as predictors for output and inflation). A quick out-of-sample forecasting exercise provides an example: starting with the sample period 1964Q1 to 1989Q4, we recursively estimate two models, a quarterly AR(1) for real output growth, and a quarterly AR(1) for real output growth with lagged quarterly returns on the S&P 500 index as an additional predictor. We use these models to generate one-step-ahead forecasts of real output from 1990Q1 to 2005Q4. In most of the estimation sample periods, lagged return is statistically significant at the 10%, if not the 5% level of significance. Yet lagged returns do not improve out-of-sample forecasts of real output, and in fact produces slightly worse forecasts: the ratio of the out-of-sample mean square forecast error (MSFE) from the AR(1) + stock return model to the MSFE of the AR(1) model is 1.011. Repeating the exercise with current returns replacing lagged returns produces similar results: the out-of-sample MSFE for the augmented AR(1) model is larger than the MSFE for the unaugmented AR(1) model. In this example, we used the standard practice of aggregating the higher frequency variable down to the frequency of the lower frequency variable. In

this paper, we ask if alternative approaches might improve the out-of-sample predictive performance of simple forecasting models of real output growth that use stock returns as a predictor.

The question of how to use high frequency data for forecasting macroeconomic variables is essentially a question of how to weight the high frequency observations to form an aggregate predictor. That is, the question is how best to filter the data. By using quarterly stock returns, the forecaster is essentially filtering the data by adding up consecutive lots of 66 daily returns. This smooths out daily fluctuations in stock returns and emphasizes the lower frequency features of the data. Some smoothing is obviously desirable: any single, individual, daily observation of stock return is unlikely to contain much information that is useful for forecasting quarterly real output growth. But what is the optimal way to filter the data? There is no compelling reason to believe that an equal-weighted average of 66 daily returns is the best way to aggregate the data. Perhaps there is too much smoothing, perhaps there is too little. Should more weight be placed on more recent observations? It may be that the quarterly frequency for stock returns is optimal for predicting quarterly output, but there is no reason why it must be so. Finding the optimal frequency is ultimately an empirical question, which is the focus of our investigation. We investigate two models for using daily stock returns to forecast, and update forecasts of, quarterly real output growth. Both models use daily returns aggregated in some manner to form a single stock market predictor. We consider (i) augmenting the quarterly AR(1) model for real output growth with daily returns using a nonparametric Mixed Data Sampling (MIDAS) setting, and (ii) augmenting the quarterly AR(1) model with the most recent r -day return as an additional predictor.

Many recent studies have focused on the issue of mixing data of different frequencies. For instance, Mariano and Murasawa (2003) construct a monthly coincident indicator of the business cycle using quarterly GDP and monthly coincident business cycle indicators. In the forecasting context, Ghysels, Santa-Clara, and Valkanov (2006a) investigate if equity returns at very high

frequency (data observed at 5 minute intervals) are useful for forecasting future realized volatility of daily equity return, defined as increments in quadratic variation. They find that models that make direct use of 5 minute data do not produce better volatility forecasts than models that use daily information as predictors. Clements and Galvão (2006) consider monthly predictors of quarterly output growth and inflation. Both papers use the MIDAS approach introduced in Ghysels, Santa-Clara, and Valkanov (2004) and Ghysels, Sinko, and Valkanov (2006b) in their investigations. Stock returns are among the many predictors studied in Clements and Galvão (2006), but there are important differences between that work, and the work presented here. First, we use a nonparametric framework based on functional regression modelling (Ramsay and Silverman, 2005). Furthermore, our treatment of lagged quarterly output is different. Second, Clements and Galvão (2006) consider the optimal combination of monthly stock returns whereas we consider the optimal combination of daily returns. Third, in addition to the MIDAS setting, we also consider simple regression forecasting models that include stock returns at various frequencies. Finally, and most importantly, our conclusions and recommendations regarding the use of stock returns as a predictor for quarterly real output growth are substantially different from theirs.

In the next section, we describe the data, the setup of our study, the models we use, how they are estimated, and how the forecasts are evaluated. Results are presented and discussed in Section 3. Section 4 concludes.

2. Data, Study Setup, Models

Throughout this paper, quarterly real output growth refers to quarterly percentage change in seasonally adjusted annualized chained-weighted GDP. Stock returns are returns on the S&P500 stock index. We use data from the start of 1964 to the end of 2005, which gives 168 quarterly observations of real output growth, and 10,958 daily observations of stock market returns (we use

five-day weeks, with zero returns for days when the market is closed). Data from 1962Q1 to 1963Q4 are available in our sample, but are reserved for accomodating lags in our forecasting models.

2.1 *Study Setup*

We evaluate two classes of models for forecasting US quarterly real output growth. Both will be described in detail shortly, but for now we simply note that both use lagged quarterly output growth and daily stock returns as predictors; the difference between the two models is in how the mix of frequencies is handled. The performances of these two models in forecasting real output growth at various horizons are compared against a benchmark model that uses only the latest available lagged output growth. We consider forecasting at horizons of $h=0, 20, 40, 70, 90, 110,$ and 130 days, corresponding approximately to monthly horizons of zero to six months. The choice of forecast horizons covers both our desire to determine how best to combine daily return data into a quarterly forecasting model, and our desire to see if daily return data within a quarter can be helpful in updating our forecasts of real output growth for that quarter. The zero-month-ahead forecast, sometimes called a ‘nowcast’, is a forecast of real GDP growth in a particular quarter, made at the end of that quarter. This is, of course, not a trivial exercise because GDP data for any one quarter is available only about a month after the end of the quarter.

Describing models that combine quarterly and daily data requires notation that is a little more intricate than usual; we explain here the system we use in the paper. We use τ as a time index, with $\tau=1$ to $10,985$ denoting the ends of days 1 through $10,985$. We use t_k , $k=1$ to 168 , to indicate the end-of-quarters: t_1 is the day-index of the last day of 1964Q1, t_2 is the day-index of the last day of 1964Q2, and so on. We have $t_1=65$, $t_2=130$, $t_3=196$, ..., $t_{176}=10,985$. Each quarter contains 64 to 66 days, not counting weekends. We use $x(\tau)$ to denote the one-day return over day τ , and $x(\tau, r)$ to denote the r -day return measured at time τ . Thus, $x(\tau, 1) \equiv x(\tau)$, and the total return on the index over day τ and the previous $r-1$ days is

$$x(\tau, r) = \sum_{i=0}^{r-1} x(\tau - i, 1) \quad (1)$$

Observed values of output growth are denoted by $y(t_k)$, $k = 1, 2, \dots, 168$.

2.2 Models

The first model we consider is a non-parametric implementation of the MIDAS model introduced in Ghysels et al. (2004, 2006b). In this implementation, we treat τ as running continuously from 0 to 10,985, and $x(\tau)$ as noisy observations of a continuous-time process. This is incorporated into a h -day ahead forecasting model for quarterly real output growth as

$$y(t_k) = \beta_0 + \beta_1 y^* + \int_0^L \beta(s) \tilde{x}_{k,h}(s) ds + \varepsilon(t_k) \quad (2)$$

where $\tilde{x}_{k,h}(s) \equiv x(t_k - h - s)$, and y^* is an appropriate lag of quarterly real output growth. We take $L = 270$, i.e., we include just over a year of daily returns data in each of our forecasting models. Because of the timing of the release of quarterly real output growth, the latest available lagged quarterly real output growth y^* is $y(t_{k-1})$ for horizons $h = 0, 20$, and 40 , $y(t_{k-2})$ for horizons $h = 70, 90$, and 110 , and $y(t_{k-3})$ for horizon $h = 130$.

Before proceeding, we note that no attempt is being made in model (2) to capture the structural aspects of the stock market - real output relationship. It is merely a projection of real output growth on available lagged output growth and stock returns data, for the purpose of forecasting. It is a forecasting model, and will be evaluated as such.

Model (2) is an example of a functional regression model. Although there have been few applications of such models in economics (e.g. Ramsay and Ramsey, 2002), such models have proven useful in other fields. Methods for estimating such models, and examples of their application, are discussed in detail in Ramsay and Silverman (2002, 2005). Software is available for estimating general functional regressions; we use Prof. James Ramsay's MATLAB implementation, downloaded from <http://ego.psych.mcgill.ca/misc/fda/>. The idea behind the estimation method is to represent $\beta(s)$ and $\tilde{x}_{k+1,h}(s)$ in terms of basis functions

$$\beta(s) = \sum_{i=1}^m b_i \theta_i(s) \quad \text{and} \quad \tilde{x}_{k,h}(s) = \sum_{j=1}^n c_{k,h,j} \psi_{k,h,j}(s)$$

where $\theta_i(s)$, $i=1, \dots, m$, and $\psi_{k,h,j}(s)$, $j=1, \dots, n$, are the basis functions used to represent $\beta(s)$ and $\tilde{x}_{k,h}(s)$ respectively, and where b_i and $c_{k,h,j}$ are the corresponding weights on these basis functions. We use B-spline bases of order 4 for both $\beta(s)$ and $\tilde{x}_{k,h}(s)$, with knots placed at every five lags, that is, at lags $s = 0, 5, 10, \dots, 270$. Model (2) is estimated by minimizing the residual sum of squares subject to a roughness penalty. Suppose we are estimating (2) over a sample of T observations. In matrix notation, we can write $\mathbf{x}(s) = \mathbf{C}\boldsymbol{\psi}(s)$ and $\beta(s) = \boldsymbol{\theta}(s)' \mathbf{b}$, where $\boldsymbol{\theta}(s)$ and \mathbf{b} are the $n \times 1$ vectors of the basis functions $\theta_i(s)$ and their weights b_i respectively, $\boldsymbol{\psi}(s)$ is the vector of basis functions $\psi_j(s)$, and \mathbf{C} is the $T \times m$ matrix of the coefficients $c_{k,h,j}$ on these basis functions. Then

$$\begin{aligned} \mathbf{y} &= \beta_0 + \beta_1 \mathbf{y}^* + \int_0^L \mathbf{x}(s) \beta(s) ds + \boldsymbol{\varepsilon} \\ &= \beta_0 + \beta_1 \mathbf{y}^* + \int_0^L \mathbf{C}\boldsymbol{\psi}(s) \boldsymbol{\theta}(s)' \mathbf{b} ds + \boldsymbol{\varepsilon} \\ &= \beta_0 + \beta_1 \mathbf{y}^* + \mathbf{C}\mathbf{J}_{\psi\theta} \mathbf{b} + \boldsymbol{\varepsilon} \\ &= \mathbf{Z}\boldsymbol{\zeta} \end{aligned}$$

where \mathbf{y} and \mathbf{y}^* are $T \times 1$ vectors of observations of quarterly real output $y(t_k)$ and its appropriate lag respectively, $\mathbf{J}_{\psi\theta} = \int_0^L \boldsymbol{\psi}(s) \boldsymbol{\theta}(s)' ds$, $\mathbf{Z} = [\mathbf{1} \quad \mathbf{y}^* \quad \mathbf{C}\mathbf{J}_{\psi\theta}]$, and $\boldsymbol{\zeta}$ is the vector $[\beta_0 \quad \beta_1 \quad \mathbf{b}]'$. To ensure sufficient smoothness in $\beta(s)$, the coefficient vector $\boldsymbol{\zeta}$ is estimated by minimizing the penalized sum of squares

$$(\mathbf{y} - \mathbf{Z}\boldsymbol{\zeta})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\zeta}) + \lambda \mathbf{b}' \mathbf{R} \mathbf{b} = (\mathbf{y} - \mathbf{Z}\boldsymbol{\zeta})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\zeta}) + \lambda \boldsymbol{\zeta}' \mathbf{R}_0 \boldsymbol{\zeta}$$

where $\mathbf{R} = \int_0^T [D^2 \boldsymbol{\theta}(s)][D^2 \boldsymbol{\theta}(s)'] ds$, and where \mathbf{R}_0 is the matrix \mathbf{R} augmented with two leading rows and columns of zeros. Then the minimizing value of $\boldsymbol{\zeta}$ is

$$\hat{\boldsymbol{\zeta}} = (\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0)^{-1} \mathbf{Z}'\mathbf{y}$$

with variance $\text{var}[\hat{\boldsymbol{\zeta}}] = \sigma_\varepsilon^2 (\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0)^{-1} \mathbf{Z}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0)^{-1}$, where σ_ε^2 can be estimated from the residuals. The smoothing parameter λ is chosen by cross-validation over the estimation period, where we search over $\lambda = 10^2$ to $\lambda = 10^7$. We choose a new smoothing parameter by cross-validation

each time the estimation period is rolled forward. We refer to forecasts from (2) as “functional regression forecasts”.

The second model we consider is a linear regression forecasting model of the form

$$y(t_k) = \beta_0 + \beta_1 y^* + \beta_2 x(t_k - h, r) + v(t_k), \quad (3)$$

that is, where we use lagged r -day returns on the stock market together with an appropriate lag of real output growth. This model is a restricted version of the functional forecasting model (2), forcing $\beta(s)$ to be zero over some days, and constant over other days. This is a natural alternative to (2) for several reasons. First, taking r -day returns is a very natural, easily interpretable way of combining r one-day returns. Second, model (3) is easy to apply, and directly addresses the question of whether there is an optimal frequency of stock returns for forecasting quarterly real output growth. We consider the performance of model (3) for $r = 1, 5, 10, 15, \dots, 270$, over each of the horizons $h = 0, 20, 40, 70, 90, 110, 130$. That is, at each horizon, we evaluate 55 versions of (3), corresponding to 55 values of r , the level of aggregation of stock returns.

2.3 Forecast Evaluation

We evaluate both models on their out-of-sample forecast performance, with all parameters estimated using a rolling scheme. In the first instance, the models are estimated over the estimation period 1964Q1 – 1989Q4 (104 quarters). Then forecasts from our two models and the benchmark model are computed for 1990Q1. The estimation period is then rolled forward by one quarter, to 1964Q2 – 1990Q1, and forecasts computed for quarterly real output in 1990Q2. This is repeated until forecasts for 2005Q4 are made, giving 64 forecasts from each model.

The quarterly real output growth data is plotted in Figure 1, with the initial estimation sample period and the forecast sample period marked out. We will evaluate our forecasts over the full forecast sample period, as well as over each of three forecast subperiods: 1990Q1 to 1994Q4 (20 quarters), 1995Q1 to 2000Q4 (24 quarters), and 2001Q1 to 2005Q4 (20 quarters). The purpose for

evaluating the forecasts over three subperiods is because each of the three subperiods contain different features of real output growth. The first subperiod starts just before the NBER-dated contraction of 1990Q3 – 1991Q1, and the subsequent recovery. Real output growth is fairly stable over the second subperiod, with minor fluctuations around a rate of about 4 percent. Real output growth in the third subperiod falls in early 2001, increases sharply in 2003 to 7.5 percent, and then falls back to levels averaging approximately three percent. We take the view, in evaluating forecasts, that a good forecast ought to capture the important features of the target variable. Thus, we are more interested in the forecast performance over the first and third subperiods. In the second period, the target variable fluctuates steadily over a significant length of time. Any two reasonable forecasting models should predict the same, and so there would be little to choose between them over this subperiod. Furthermore, and importantly, a *bad* forecasting model that *always* predicts minor fluctuations over a constant rate will also appear to be a reasonable forecasting model. A comparison between the good and the bad forecasting models will reveal little over such a subperiod.

We will evaluate, at each forecast horizon h , the 56 models (the functional forecasting model (2), plus the 55 linear regression models (3)) by comparing their MSFEs and Mean Absolute Forecast Error (MAFEs) against the MSFE and MAFE for the forecasts from the appropriate benchmark model

$$y(t_k) = \beta_0 + \beta_1 y^* + \eta(t_k) \quad (4)$$

where, as before, y^* is $y(t_{k-1})$, $y(t_{k-2})$, or $y(t_{k-3})$, depending on the horizon h . Statistical significance is evaluated using the test for Superior Predictive Accuracy (SPA) developed in Hansen (2005), for comparing multiple alternative forecasts against a benchmark forecast. The SPA test is closely related to the ‘reality check’ test of White (2000), and like the latter is designed to account for the mining over alternative forecasting models. The SPA test, however, controls for poor alternatives,

and is therefore less sensitive to the inclusion of these alternatives, and has been demonstrated to be more powerful in simulations and in application (Hansen, 2005; Hansen and Lunde, 2005).

4. Results and Discussion

The estimated coefficient functions for the functional forecasting model for $h=0$, 70, and 130 are displayed in Figures 2(a), (b), and (c) respectively. Only the coefficient functions estimated over the initial estimation period are shown. The estimated coefficient function for the functional forecasting model with $h=0$ shows that most of the weight is placed on returns in the previous two quarters ($s = 70$ to 200 approximately), and that these are statistically significant (Figure 2a). Little weight is placed on returns in the current quarter ($s = 0$ to 70, approx.). Some weight is placed on returns in the fourth quarter ($s > 200$) but these are statistically insignificant. This estimated weight function does not change much as the estimation sample period rolls forward. The estimated weight function is fairly similar to what is obtained when quarterly real output growth is regressed onto lagged output growth, current quarterly stock return, and three lags of quarterly stock returns. Over the period 1964Q1 – 1989Q4, this regression gives

$$y_t = 0.025 + 0.180 y_{t-1} + 0.018 x_t + 0.079 x_{t-1} + 0.137 x_{t-2} + 0.039 x_{t-3} \quad (5)$$

$$\begin{array}{cccccc} [4.953] & [1.816] & [0.388] & [1.699] & [2.893] & [0.800] \end{array}$$

where we use standard notation for regressions where all variables are measured in the same frequency, with x_t denoting quarterly stock returns. The numbers in the square brackets are t-statistics. As in the functional regression model, the largest weights are for returns in the previous two quarters.

For $h = 70$ (3-month horizon), the estimated coefficient function on stock returns for the functional forecasting model places weights that decline with the lag (Figure 2b). The weights on the first three quarters of included returns are significant. For $h = 130$ (6-month horizon), statistically

significant weight is placed on the first quarter of included daily returns only (Figure 2c). Again, these estimated weight functions are broadly similar to those obtained from corresponding quarterly-data regressions. The estimated weight functions suggest that it is stock returns within one year of the end of the target quarter that are relevant for forecasting real output growth.

In Figure 3, the performance of the functional forecasting models are visually compared with forecasts from the benchmark model (4). Three groups of plots are presented: Figure 3(a) compares the 0, 20, and 40 day horizon forecasts against forecasts from (4) when $y^* = y(t_{k-1})$; Figure 3(b) compares the 70, 90, and 110 day horizon forecasts against forecasts from the benchmark model when $y^* = y(t_{k-2})$; Figure 3(c) compares the 130 day horizon forecasts against forecasts from (4) when $y^* = y(t_{k-3})$. Different benchmark models are used for forecasts at different horizons to account for the fact that at longer horizons, the most recent lags of real output growth are not yet available. The top-left panels of Figure 3(a) and (b), and the left panel of Figure 3(c), all labeled ‘benchmark’, show the forecasts from the benchmark model (dots) plotted over the solid line representing the realizations of quarterly real output growth. The other panels show the functional regression forecasts at the various horizons, against the realizations of real output growth. Clearly, the benchmark forecasts do not appear to pick out the important features of the data; the benchmark model for the six-month horizon model is particularly weak (Figure 3c). In all cases, the forecasts from the functional regression models do a better job (at least visually) at tracking the data in the first subperiod 1990Q1 – 1994Q4, and especially in the last forecast subperiod, 2001Q1 – 2005Q4.

The MSFE and MAFE comparisons of the functional regression forecasts with the benchmark forecasts are given in Table 1. The numbers in the first panel of the table are ratios of the functional regression models’ out-of-sample MSFE to the corresponding benchmark models’ out-of-sample MSFE. A value less than one indicates that the functional regression model improves on the forecasts of the benchmark model. The second panel shows the corresponding ratios of MAFEs. The

numbers are in agreement with the visual comparisons. Some improvements in MSFEs and MAFEs are seen in the first forecast subperiod, but these are very small. There is generally no improvement in the second forecast subperiod, but as pointed out earlier, little can be drawn from this subperiod. More noteworthy is the fact that large improvements in out-of-sample MSFEs and MAFEs are obtained in the last forecast subperiod, with reductions in MSFE of over 20 percent to over 30 percent. The fact that large improvements are observed in one subperiod and not in others is consistent with observations elsewhere that the usefulness of financial assets for predicting output growth varies over time (e.g., Stock and Watson, 2003).

The corresponding figures and statistics for the linear regression forecast models are given in Figures 4(a), (b), and (c), and in Table 2. At each horizon, we evaluate forecasts from 55 models (corresponding to 55 values of r), but we only report the statistics for the best performing model. Visual comparisons in Figure 4 generally lead to the same conclusions as those obtained from studying Figure 3. There are, however, some quantitative differences. In Table 2, we see small improvement in forecast performance in the first subperiod, and no improvement in MSFEs and MAFEs in the second subperiod. There does not appear to be important differences between the ratios for these two subperiods and the corresponding ratios in Table 1. However, very large reductions in MSFEs and MAFEs are observed for the third forecast subperiod, much larger reductions than those observed in Table 1. In terms of MSFE, the reductions over baseline range from just under 37 percent to over 50 percent.

The fact that the improvements in the third forecast subperiod are larger than those obtained from the functional forecasting model is particularly interesting because the restrictions imposed in the linear regression models (3) are generally false in-sample (see for instance equation (5)). This result is, nonetheless, consistent with results (in other contexts) that show that the imposition of substantial a priori structure can lead to improved forecasts. Stock and Watson (2004) find, in a study

covering seven developed countries, that an equally weighted combination of forecasts of real output growth based on individual predictors performs better than forecast combinations that heavily weight recent performance or allow for substantial time variation in the weights. Diebold and Li (2006) consider a forecasting model of the yield curve that imposes a restricted version of the Nelson-Siegel (1987) functional form on the yield curve at each period, with a simple AR(1) specification on the (three) parameters of this functional form. They find that long horizon forecasts from this model are more accurate than forecasts from the usual benchmark models.

It is also interesting to note which models perform best at the various horizons. At the 0-month horizon, the best model is one that uses one-year returns on the stock market. With each one month increase in the forecast horizon, up to the 4-month forecast horizon, the length of the return included in the best model falls by approximately one month. This supports the assertion made earlier that it is returns over the one-year period prior to the end of the target quarter that is relevant for forecasting output.

We now carry out a slightly different, more formal, comparison of the forecasts, using the SPA test of Hansen (2005). We carry out one test per forecast horizon. For each horizon, we ask if any of the 56 models (one functional forecasting model, and 55 linear regression forecasting models) produces forecasts that are significantly better than the baseline model appropriate to the horizon. At each horizon, let the forecasts from the 56 alternative models be \hat{Y}_{mt} , where $m = 1, \dots, 56$, and the evaluation period is $t = 1, \dots, n$, and let the baseline model forecasts be \hat{Y}_t . Then the hypothesis is $\boldsymbol{\mu} \leq \mathbf{0}$ where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{56} \end{pmatrix} = E \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_{56}(t) \end{pmatrix}$$

and $X_{m,t} = L(Y_t, \hat{Y}_t) - L(Y_t, \hat{Y}_{mt})$. The function $L(.,.)$ is the loss criteria. We use mean square loss $L(Y_t, \hat{Y}_{mt}) = (Y_t - \hat{Y}_{mt})^2$, and absolute loss $L(Y_t, \hat{Y}_{mt}) = |Y_t - \hat{Y}_{mt}|$. The test statistic is

$$T_n^{sm} = \max_m \frac{n^{1/2} \bar{X}_m}{\hat{\sigma}_m}.$$

Estimates of $\hat{\sigma}_m^2 = \widehat{\text{var}}(n^{1/2} \bar{X}_m)$ and consistent p-values for this test, as well as upper and lower bounds for the p-values, are obtained using a bootstrap implementation described in Hansen (2005). We use 1000 bootstrap replications throughout. The bootstrap implementation also requires the choice of a parameter q to account for dependency in the relative loss series. We report results for $q = 0.5$ and $q = 0.25$.

The results of this test are reported in Table 3. Results are reported for both MSFE and MAFE loss metrics, for $q = 0.25$ and $q = 0.5$, for the three forecast subsamples. The number that appears under ‘Best’ refers to the number of days over which returns are aggregated in the best performing model. If ‘func’ appears in this column, this indicates that the best performing model is the functional forecasting model. For the third forecast subperiod, we observe that as the forecast horizon increases by a month, the number of days over which returns are aggregated in the best performing model falls by about a month, very similar to an earlier observation. The exception is the 6-month horizon. We note, however, that the second best performing model in this case is the model with 125-day returns. Three “p-values” are reported for each loss function. The p-values reported under SPA_c are the asymptotically consistent p-values for the test statistic. The numbers reported under SPA_l and SPA_u are the lower and upper bounds for the p-values. The results show that the benchmark model is outperformed in almost all cases in the third forecast subsample, at either the 5% or 10% level of significance. The exceptions are the 2-month-ahead forecasts with $q = 0.5$, and the 5-months-ahead forecasts, where weak statistical significance is obtained only for $q = 0.25$ with MSFE as the loss criterion. The 2- and 5- month forecast horizons are the horizons at which a new

observation of lagged real output growth becomes available, which may explain the fall in statistical significance at these horizons. Nonetheless, the results in Table 2 suggest that the gains from including stock returns at an appropriately aggregated level into the forecasting model is substantial at all horizons. The results also highlight the value of using latest available observations (of low frequency) stock returns to update forecasts of real output.

4. Conclusion

Financial data are in general available at much higher frequencies than macroeconomic data. We asked how high frequency stock market data might best be used when the objective is to forecast real output growth. In particular, we asked how daily data should be combined in a model for forecasting quarterly real output growth. We explored two different methods, one where daily data enters into the quarterly forecasting model as a functional predictor, and another where daily stock return data is simply aggregated into multi-day returns and put into a linear regression forecasting model.

All the results suggest that stock returns are useful predictors for real output growth, particularly in recent years: (i) our forecasting models that include stock returns have succeeded in tracking the main features of real output growth; (ii) our short-horizon forecasting models show that stock returns are useful for updating forecasts of real output growth. Our forecasting models all suggest that it is returns over the year leading up to the end of the target quarter that contains information useful for forecasting output growth. Although estimating the weights in-sample leads to forecasts that are better than a benchmark forecast model using only lagged output growth, results from our linear regression forecasting models that assume equal weight over the daily returns perform even better, even though restriction of equal weights is generally rejected in-sample.

The results suggest a simple way of using the latest available stock return data to forecast, or to update forecasts of, real output growth, and that is to use aggregated stock returns as a predictor, including data in the year leading up to the end of the target quarter: to predict real output growth 2 months prior to the end of the target quarter, use aggregate returns over the latest 10 months; to predict real output growth 10 months prior to the end of the target quarter, use the latest available 2-month returns, etc.. More generally, the results in the paper suggest that the practice of coercing all variables in a forecasting model to have the same frequency might not be an optimal procedure. Mixing frequencies can lead to better forecasts.

References

- Clements, M.P. and Galvão, A.B. (2006), "Macroeconomic Forecasting with Mixed Frequency Data: Forecasting US Output Growth and Inflation," manuscript.
- Diebold, F. X. and Li, C. (2006), "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics* 230, 337-364.
- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004), "The MIDAS Touch: Mixed Data Sampling Regression Models," manuscript.
- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006a), "Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies," *Journal of Econometrics* 131, 59-95.
- Ghysels, E., Sinko, A., and Valkanov, R. (2006b), "MIDAS Regressions: Further Results and New Directions," forthcoming in *Econometric Reviews*.
- Hansen, P. R. (2005) "A Test for Superior Predictive Ability," *Journal of Business and Economic Statistics* 23, 365-380.
- Hansen, P. R. and Lunde, A. (2005), "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?" *Journal of Applied Econometrics* 20, 873-889.
- Mariano, R. S. and Murasawa, Y. (2003), "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," *Journal of Applied Econometrics* 18, 427-433.
- Nelson, C. R. and Siegel, A. F. (1987), "Parsimonious Modeling of Yield Curves," *Journal of Business* 60, 473-489.
- Ramsay, J. O. and Ramsey, J. B. (2002), "Functional data analysis of the dynamics of the monthly index of nondurable goods production," *Journal of Econometrics* 107, 327-344.
- Ramsay, J. O. and Silverman, B. W. (2002), *Applied Functional Data Analysis*, Springer Series in Statistics, Springer, New York.

- Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, 2nd Ed., Springer-Verlag, New York.
- Stock, J. H. and Watson, M. W. (2003), "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature* 41, 788-829.
- Stock, J. H. and Watson, M. W. (2004), "Combination Forecasts of Output Growth in a Seven-Country Data Set," *Journal of Forecasting* 23, 405-430.
- White, H. (2000), "A Reality Check for Data Snooping," *Econometrica* 68, 1097-1126.

Table 1 Performance of Functional Forecasts vs Benchmark Forecasts

Horizon	0 mth	1 mth	2 mth	3 mth	4 mth	5 mth	6 mth
MSFE Ratio							
Full Sample	0.910	1.026	0.940	0.908	0.875	0.871	0.899
1990Q1:1994Q4	0.906	0.839	0.930	0.857	0.898	0.908	1.099
1995Q1:2000Q4	1.021	1.346	1.053	1.146	0.934	0.890	0.757
2001Q1:2005Q4	0.728	0.744	0.763	0.676	0.768	0.791	0.783
MAFE Ratio							
Full Sample	0.963	1.060	1.021	0.992	0.959	0.936	0.985
1990Q1:1994Q4	0.933	0.926	0.958	0.945	0.985	0.922	1.070
1995Q1:2000Q4	1.067	1.297	1.130	1.131	1.047	0.990	0.975
2001Q1:2005Q4	0.842	0.868	0.937	0.864	0.807	0.879	0.887

Note: Numbers in the upper panel are ratios of mean squared forecast error (MSFE) from the functional forecasting models (2) to the MSFE of the corresponding benchmark models (4). The numbers in the lower panel are the corresponding ratios of mean absolute forecast errors (MAFE)

Table 2 Performance of Regression Forecasts vs Benchmark Forecasts

Horizon	0 mth	1 mth	2 mth	3 mth	4 mth	5 mth	6 mth
Best model based on full sample MSE ratio	270	255	230	195	175	170	150
MSE Ratio							
Full Sample	0.872	0.870	0.882	0.820	0.798	0.824	0.772
1990Q1:1994Q4	1.078	1.041	0.979	0.903	0.912	0.884	0.877
1995Q1:2000Q4	0.878	0.873	0.926	1.002	0.881	0.852	0.700
2001Q1:2005Q4	0.572	0.624	0.670	0.466	0.528	0.704	0.709
MAE Ratio							
Full Sample	0.958	0.966	0.982	0.917	0.897	0.925	0.917
1990Q1:1994Q4	1.060	1.027	0.998	0.998	0.998	0.962	0.975
1995Q1:2000Q4	0.988	0.984	1.026	0.996	0.948	0.955	0.899
2001Q1:2005Q4	0.776	0.855	0.892	0.703	0.696	0.838	0.863

Note: Numbers in the upper panel are ratios of mean squared forecast error (MSFE) from the regression forecasting models (3) to the MSFE of the corresponding benchmark models (4). The numbers in the lower panel are the corresponding ratios of mean absolute forecast errors (MAFE)

Table 3 Tests for Superior Predictive Ability

Metric	q	0 Month Horizon				1 Month Horizon				2 Month Horizon				3 Month Horizon			
		Best	SPA _l	SPA _c	SPA _u	Best	SPA _l	SPA _c	SPA _u	Best	SPA _l	SPA _c	SPA _u	Best	SPA _l	SPA _c	SPA _u
<i>1990Q1:1994Q4</i>																	
MSE	0.5	func	0.444	0.656	0.701	func	0.246	0.357	0.357	155	0.529	0.692	0.745	125	0.274	0.300	0.300
	0.25	func	0.277	0.412	0.449	func	0.150	0.188	0.193	155	0.367	0.499	0.526	125	0.287	0.326	0.326
MAE	0.5	func	0.141	0.184	0.194	func	0.162	0.211	0.214	100	0.386	0.482	0.482	75	0.306	0.344	0.344
	0.25	func	0.083	0.105	0.111	func	0.043	0.063	0.064	100	0.203	0.234	0.274	75	0.168	0.189	0.189
<i>1995Q1:2000Q4</i>																	
MSE	0.5	270	0.387	0.578	0.639	255	0.275	0.449	0.513	1	0.441	0.635	0.717	200	0.609	0.825	0.847
	0.25	270	0.264	0.403	0.466	255	0.181	0.296	0.361	1	0.381	0.520	0.653	200	0.569	0.761	0.794
MAE	0.5	270	0.884	0.997	0.997	225	0.763	0.953	0.964	1	0.102	0.256	0.319	175	0.674	0.863	0.869
	0.25	270	0.879	0.996	0.998	225	0.728	0.938	0.969	1	0.085	0.210	0.316	175	0.680	0.856	0.857
<i>2001Q1:2005Q4</i>																	
MSE	0.5	255	0.041	0.041	0.041	235	0.038	0.038	0.038	225	0.116	0.128	0.137	185	0.042	0.042	0.042
	0.25	255	0.020	0.020	0.020	235	0.015	0.015	0.015	225	0.068	0.068	0.073	185	0.034	0.034	0.034
MAE	0.5	255	0.037	0.038	0.038	235	0.005	0.005	0.005	225	0.150	0.181	0.200	195	0.069	0.069	0.071
	0.25	255	0.023	0.026	0.026	235	0.001	0.001	0.001	225	0.101	0.121	0.137	195	0.024	0.024	0.024

Note: The number that appears under ‘Best’ refers to the number of days over which returns are aggregated in the best performing model in the model class (3). If ‘func’ appears in this column, this indicates that the best performing model is the functional forecasting model (2).

Table 3 contd. Tests for Superior Predictive Ability

Metric	q	4 Month Horizon				5 Month Horizon				6 Month Horizon			
		Best	SPA _{<i>l</i>}	SPA _{<i>c</i>}	SPA _{<i>u</i>}	Best	SPA _{<i>l</i>}	SPA _{<i>c</i>}	SPA _{<i>u</i>}	Best	SPA _{<i>l</i>}	SPA _{<i>c</i>}	SPA _{<i>u</i>}
<i>1990Q1:1994Q4</i>													
MSE	0.5	100	0.254	0.340	0.340	85	0.042	0.043	0.043	65	0.286	0.432	0.432
	0.25	100	0.232	0.305	0.305	85	0.040	0.040	0.040	65	0.307	0.413	0.421
MAE	0.5	230	0.178	0.211	0.211	85	0.001	0.001	0.001	65	0.317	0.345	0.357
	0.25	230	0.109	0.125	0.125	85	0.000	0.000	0.000	65	0.273	0.300	0.308
<i>1995Q1:2000Q4</i>													
MSE	0.5	185	0.230	0.293	0.293	210	0.285	0.483	0.485	190	0.045	0.048	0.048
	0.25	185	0.088	0.116	0.116	210	0.118	0.212	0.214	190	0.016	0.017	0.017
MAE	0.5	180	0.454	0.681	0.683	185	0.531	0.697	0.744	165	0.320	0.371	0.388
	0.25	180	0.388	0.553	0.587	185	0.472	0.655	0.694	165	0.288	0.328	0.337
<i>2001Q1:2005Q4</i>													
MSE	0.5	165	0.076	0.076	0.076	145	0.117	0.117	0.137	15 ¹	0.045	0.045	0.045
	0.25	165	0.038	0.038	0.038	145	0.084	0.084	0.093	15 ¹	0.020	0.020	0.020
MAE	0.5	165	0.001	0.001	0.001	145	0.183	0.234	0.243	15 ¹	0.047	0.047	0.047
	0.25	165	0.000	0.000	0.000	145	0.149	0.192	0.197	15 ¹	0.020	0.020	0.020

Note: The number that appears under ‘Best’ refers to the number of days over which returns are aggregated in the best performing model in the model class (3). If ‘func’ appears in this column, this indicates that the best performing model is the functional forecasting model (2).

¹ The second best model is the model (3) with 125-day return.

Figure 1 Quarterly Real Output Growth 1964Q1 to 2005Q4

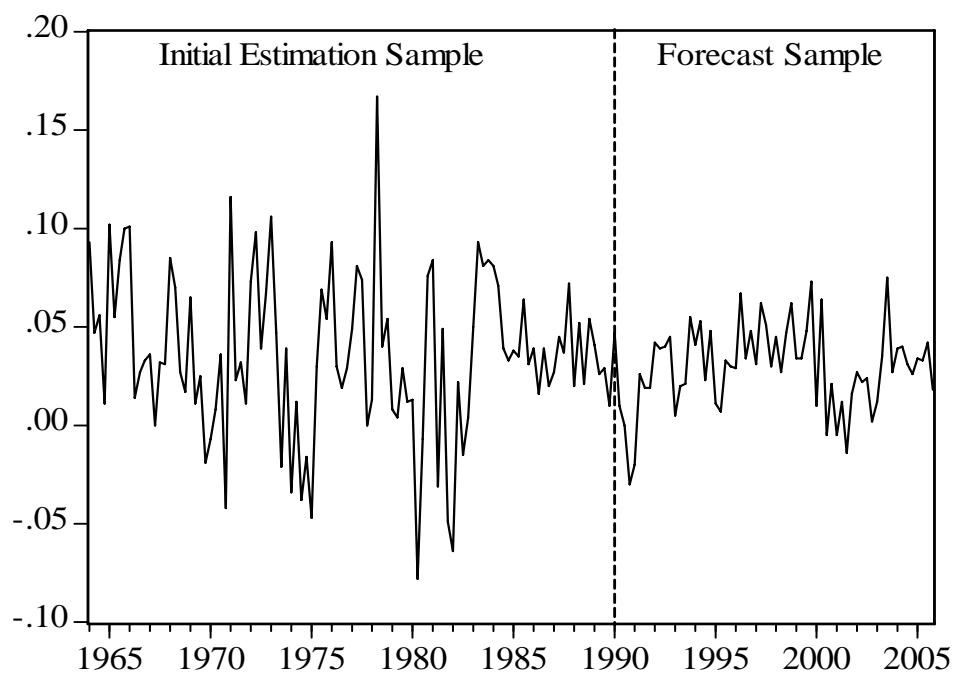
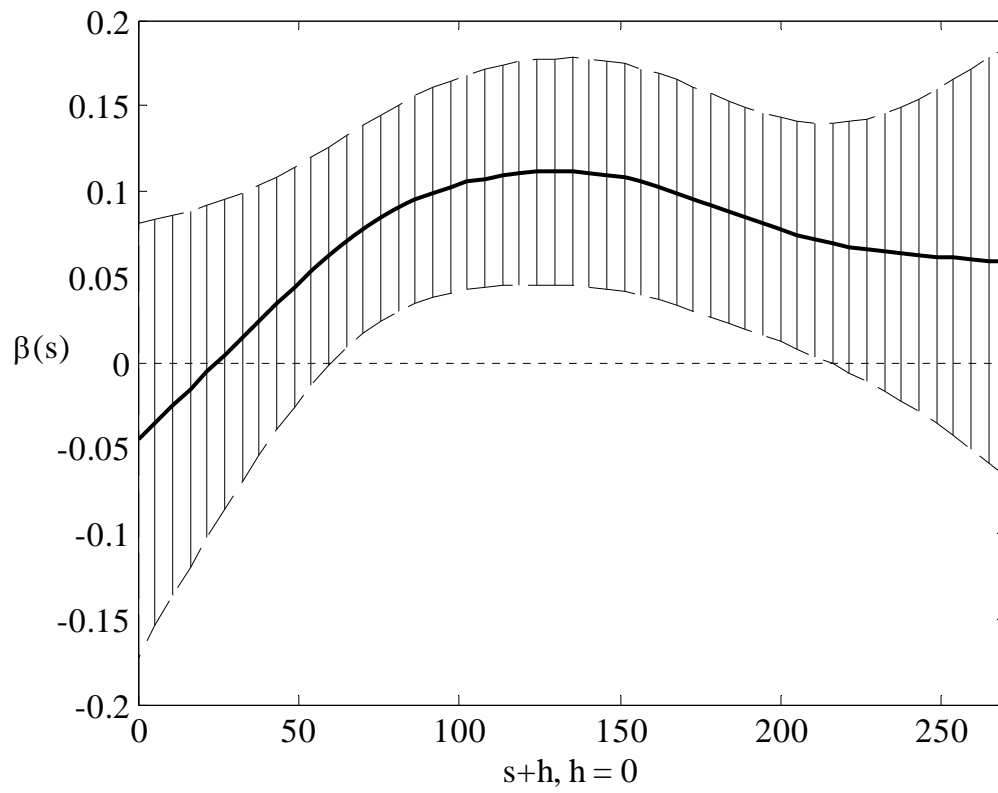
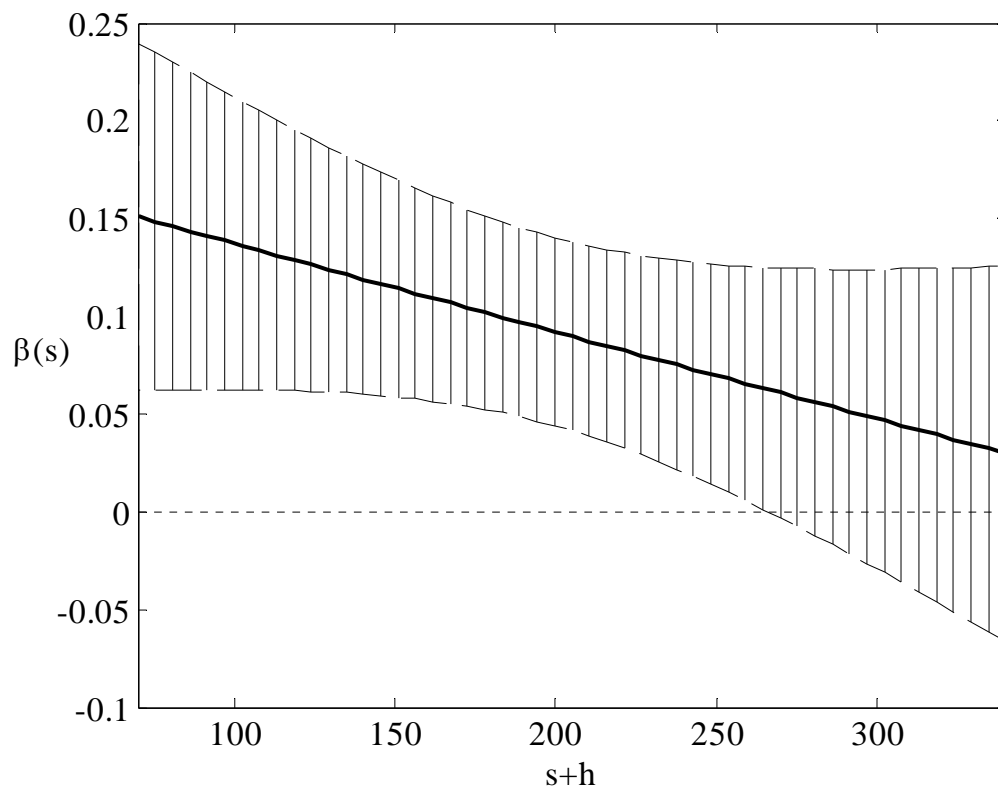


Figure 2(a) Estimated Coefficient Function for Daily Stock Returns (0 month horizon)



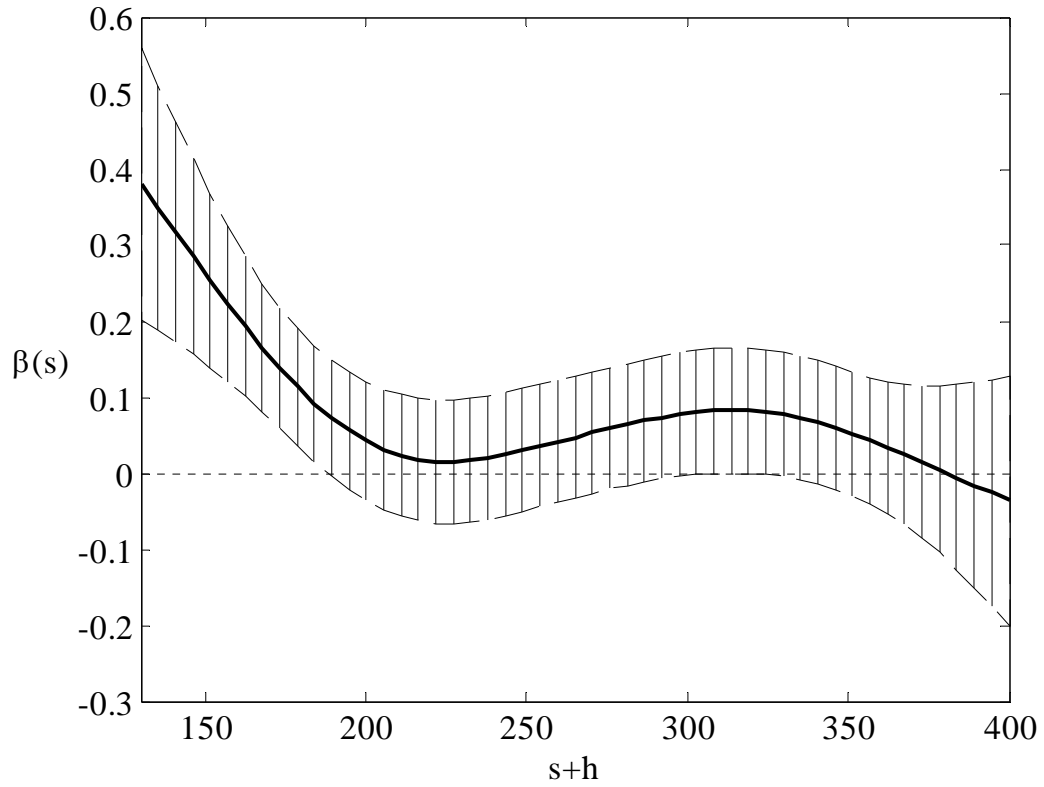
Note: $\beta(s)$ is the estimated coefficient function on daily stock returns for model (2), with s from 0 to 270, $h = 0$.

Figure 2(b) Estimated Coefficient Function for Daily Stock Returns (3 month horizon)



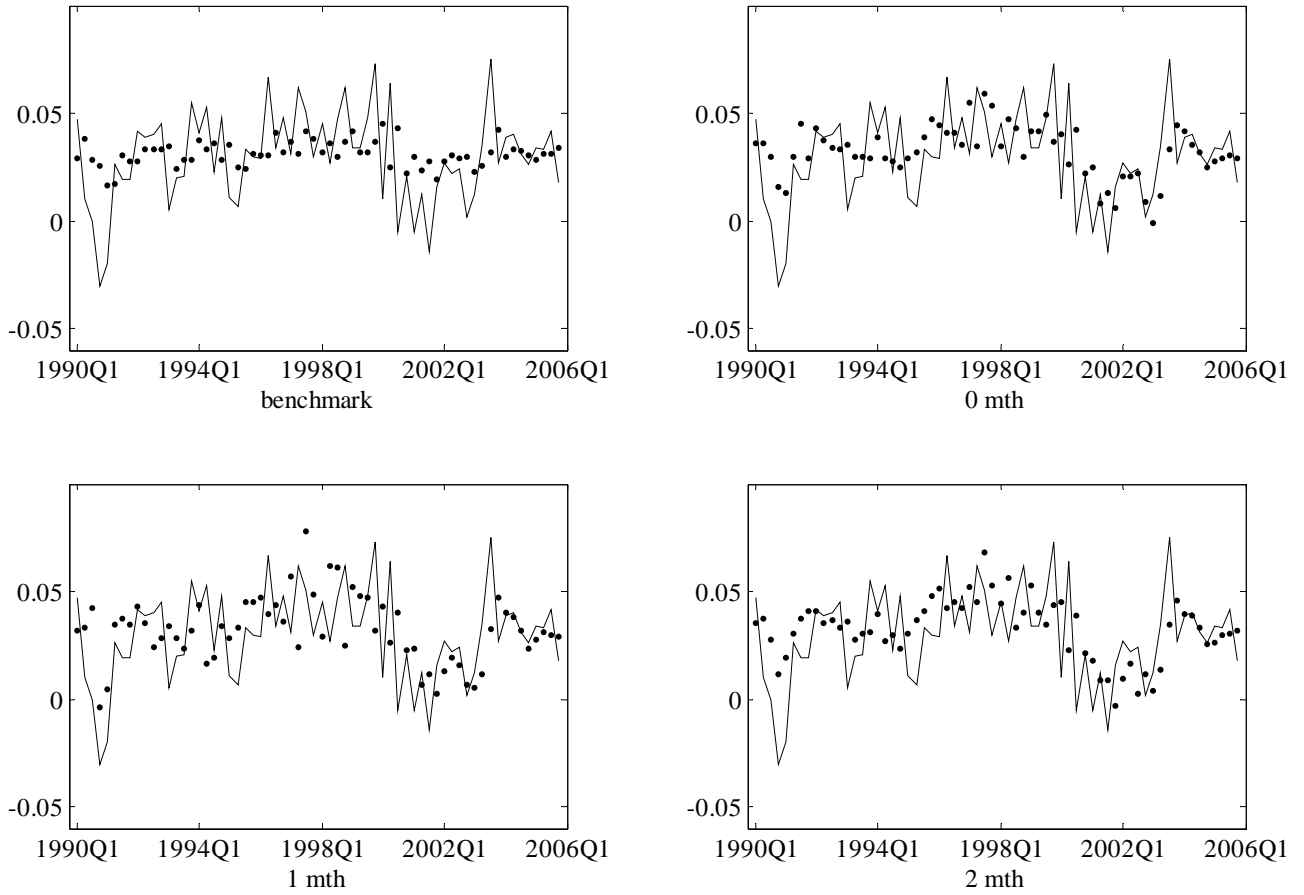
Note: $\beta(s)$ is the estimated coefficient function on daily stock returns for model (2), with s from 0 to 270, $h = 70$.

Figure 2(c) Estimated Coefficient Function for Daily Stock Returns (6 month horizon)



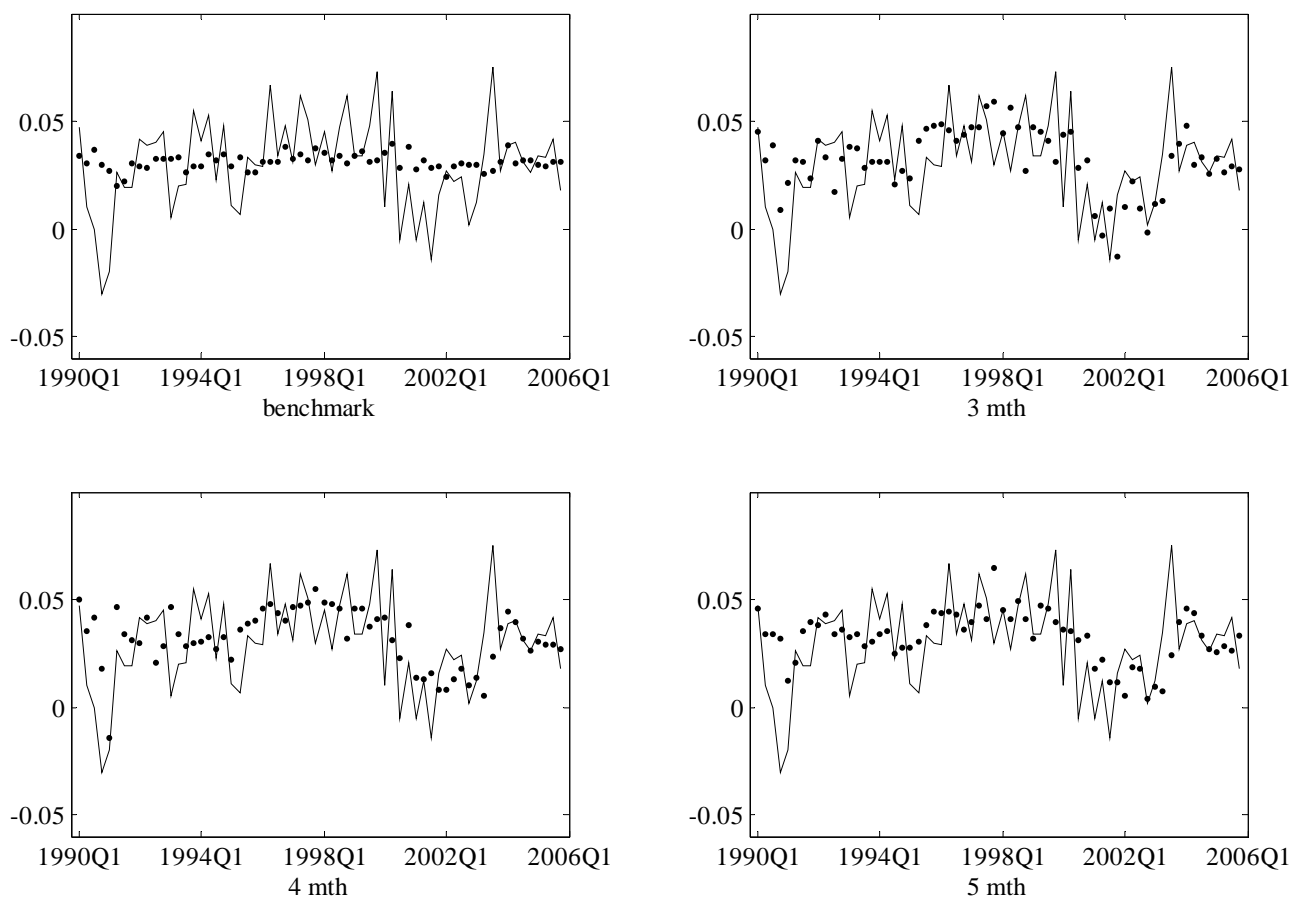
Note: $\beta(s)$ is the estimated coefficient function on daily stock returns for model (2), with s from 0 to 270, $h = 130$.

Figure 3(a) Realized Output with Functional Forecasts, Various Horizons



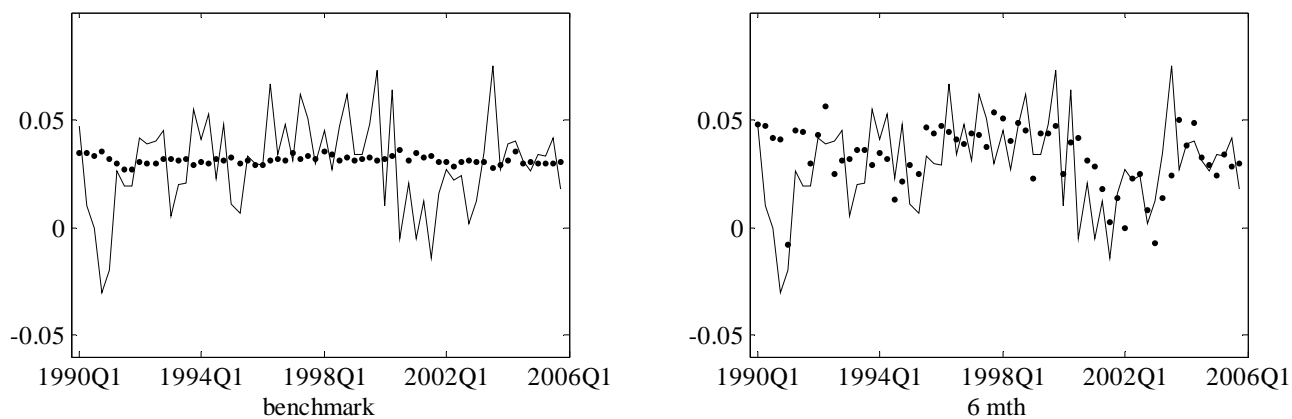
Note: Time series of forecasts (dots) from various models plotted against realizations of real output growth (solid line). The figure labeled 'benchmark' show forecasts from model (4) with $y^* = y(t_{k-1})$. The other figures show forecasts from model (2) for $h = 0, 20,$ and 40 (0 month, 1 month, and 2 month horizons, resp.).

Figure 3(b) Realized Output with Functional Forecasts, Various Horizons



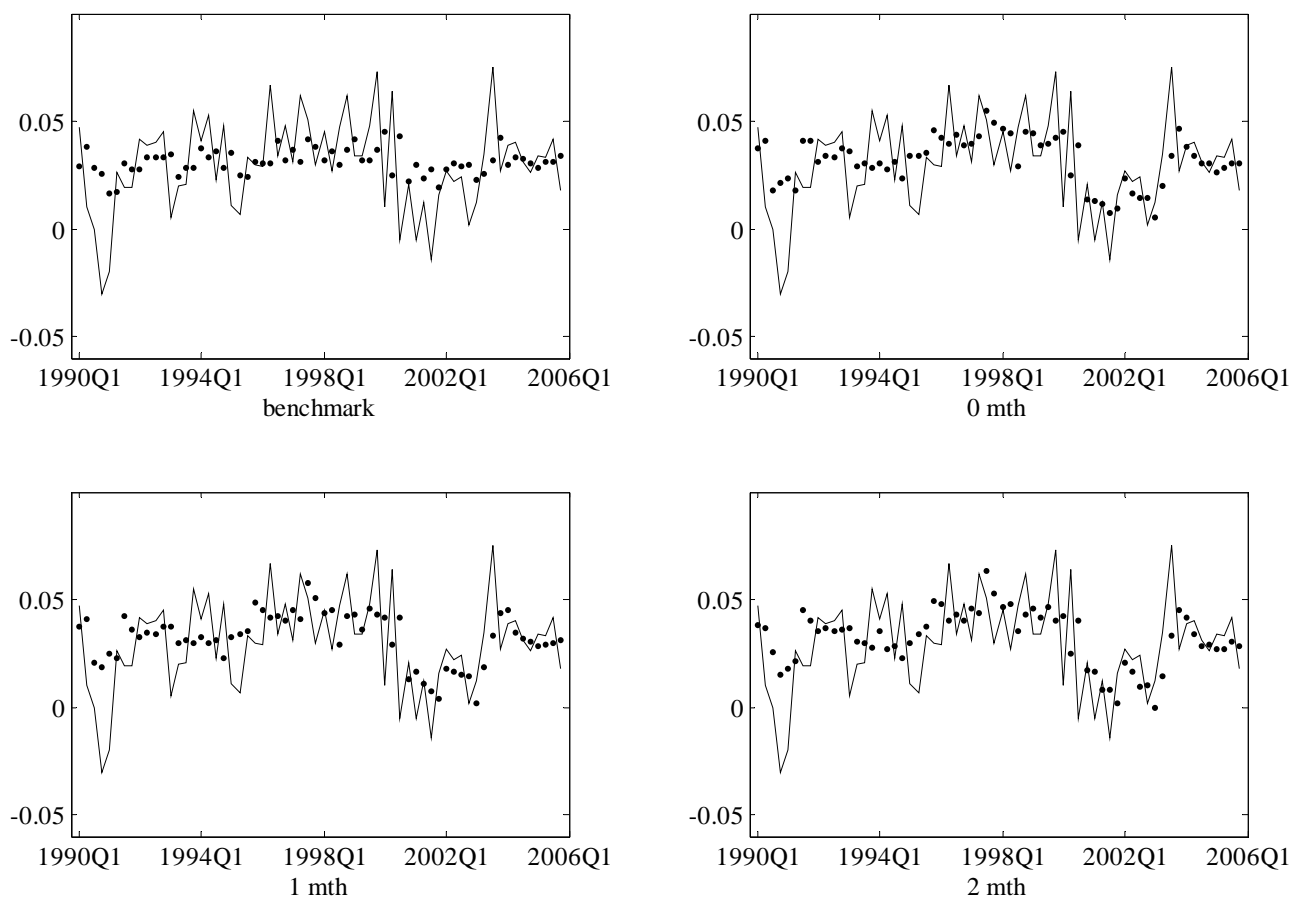
Note: Time series of forecasts (dots) from various models plotted against realizations of real output growth (solid line). The figure labeled ‘benchmark’ show forecasts from model (4) with $y^* = y(t_{k-2})$. The other figures show forecasts from model (2) for $h = 70, 90, \text{ and } 110$ (3 month, 4 month, and 5 month horizons, resp.).

Figure 3(c) Realized Output with Functional Forecasts, Various Horizons



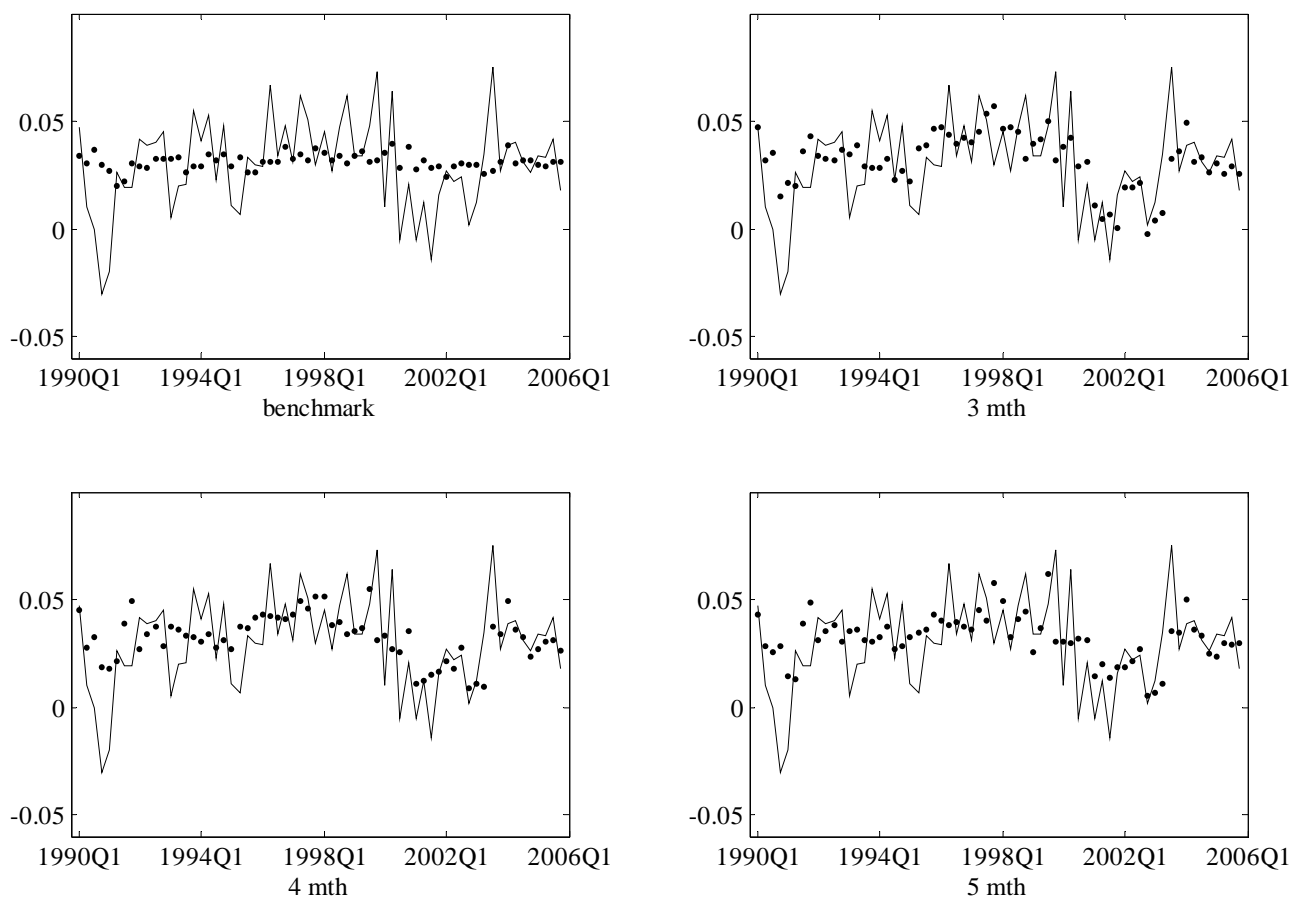
Note: Time series of forecasts (dots) from various models plotted against realizations of real output growth (solid line). The figure labeled 'benchmark' show forecasts from model (4) with $y^* = y(t_{k-3})$. The other figure show forecasts from model (2) for $h = 130$ (6 month horizon.)

Figure 4(a) Realized Output with Linear Regression Forecasts, Various Horizons



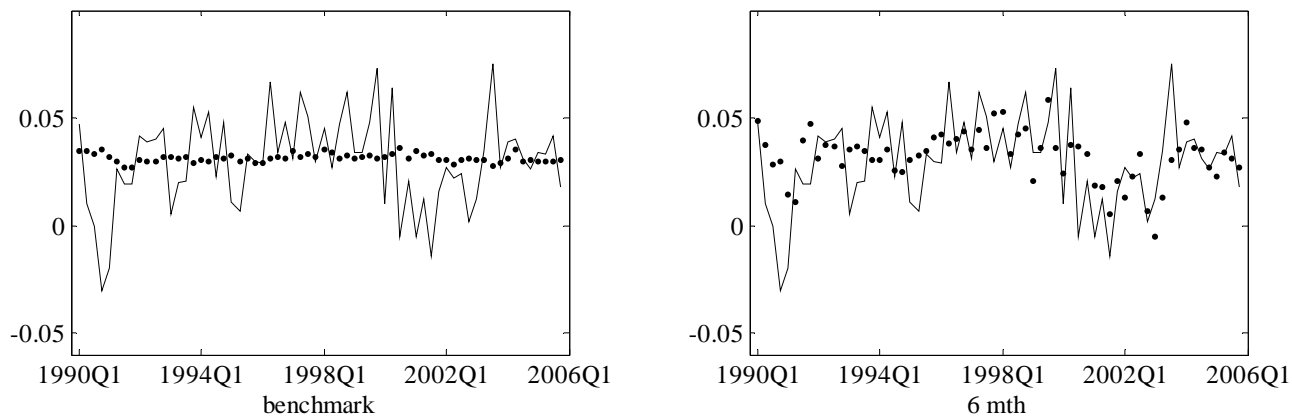
Note: Time series of forecasts (dots) from various models plotted against realizations of real output growth (solid line). The figure labeled 'benchmark' show forecasts from model (4) with $y^* = y(t_{k-1})$. The other figures show forecasts from model (3) for $h = 0, 20$, and 40 (0 month, 1 month, and 2 month horizons, resp.).

Figure 4(b) Realized Output with Linear Regression Forecasts, Various Horizons



Note: Time series of forecasts (dots) from various models plotted against realizations of real output growth (solid line). The figure labeled 'benchmark' show forecasts from model (4) with $y^* = y(t_{k-2})$. The other figures show forecasts from model (3) for $h = 70, 90, \text{ and } 110$ (3 month, 4 month, and 5 month horizons, resp.).

Figure 4(c) Realized Output with Linear Regression Forecasts, Various Horizons



Note: Time series of forecasts (dots) from various models plotted against realizations of real output growth (solid line). The figure labeled 'benchmark' show forecasts from model (4) with $y^* = y(t_{k-3})$. The other figure show forecasts from model (3) for $h = 130$ (6 month horizon.)