DISCUSSION PAPER SERIES

No. 10186

CAN WE AUTOMATE EARNINGS FORECASTS AND BEAT ANALYSTS?

Ryan Ball, Eric Ghysels and Huan Zhou

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP10186.php**

ISSN 0265-8003

CAN WE AUTOMATE EARNINGS FORECASTS AND BEAT ANALYSTS?

Ryan Ball, University of Michigan
Eric Ghysels, University of North Carolina and CEPR
Huan Zhou, University of North Carolina

Discussion Paper No. 10186
October 2014

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 10186

October 2014

ABSTRACT

Can we Automate Earnings Forecasts and Beat Analysts?

Can we design statistical models to predict corporate earnings which either perform as well as, or even better than analysts? If we can, then we might consider automating the process, and notably apply it to small and international firms which typically have either sparse or no analyst coverage. There are at least two challenges: (1) analysts use real-time data whereas statistical models often rely on stale data and (2) analysts use potentially large set of observations whereas models often are frugal with data series. In this paper we introduce newly-developed mixed frequency regression methods that are able to synthesize rich real-time data and predict earnings out-of-sample. Our forecasts are shown to be systematically more accurate than analysts' consensus forecasts, reducing their forecast errors by 15% to 30% on average, depending on forecast horizon.

JEL Classification: C53, M40 and M41
Keywords: forecast combination, MIDAS regression and real-time data

Ryan Ball
Stephen M. Ross School of Business
University of Michigan
701 Tappan Street
Ann Arbor, MI 48109-1234
USA

Email: rtball@umich.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=179100

Eric Ghysels
Department of Economics
University of North Carolina
Gardner Hall CB 3305
Chapel Hill NC 27599-3305
USA

Email: eghysels@unc.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=135789

Huan Zhou
Department of Economics
University of North Carolina
CB 3305
Chapel Hill, NC 27599
USA
Email: huanz@live.unc.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=179101

Submitted 28 September 2014

1 Introduction

Earnings Per Share (EPS) is a key input in many asset pricing models, as well as a main indicator of the current and future financial health of listed companies. Not surprisingly, a lot of resources are devoted to produce accurate and timely forecasts of future earnings.

The question whether we can automate the process using econometric models has somewhat the flavor of man versus machine, like a chess player against a computer, or a self-driving car. The main motivation is more practical, however. Analyst coverage is concentrated on large firms. If we succeed in creating reliable EPS forecasts with relatively simple to implement models, we can vastly expand the scope and breath of earnings forecasting. Notably, we can consider relatively smaller firms, and also expand across international markets, provided reliable public domain data is available.

Prior research has examined various forms of univariate extrapolative time series models and concluded that they cannot match the forecast performance of professional analysts.[1] The superiority of analysts' forecasts is partially due to both their informational as well as their timing advantages. Extrapolative time series models, whether a random walk or ARIMA models, generally rely on past earnings or components of earnings (e.g., sales, expenses, cash flows, accruals), which means that forecasts for a given quarter (year) can only be made using information prior to the end of the previous quarter (year). Analysts, however, observe all publicly available (as well as sometimes non-public) information and can update their forecasts well into the forecast target quarter(year). Past studies support the existence of such information and timing advantages. Fried and Givoly (1982) concludes that analysts' annual earnings forecasts utilize a substantially larger information set, including non-earnings information as well as observations that are not available at the end of the previous year. O'Brien (1988) concurred that analysts' information on firms sales, production, and macroeconomic conditions may have resulted in better quarterly forecasts of firm earnings. Brown, Hagerman, Griffin, and Zmijewski (1987) and subsequent research identified that the superiority of analysts correlated negatively with the forecast horizon.

The empirical challenges of setting up a time series model that can match or outperform analysts are therefore two-fold: (1) include timely data, and (2) potentially use large sources of information. We propose to address both issues taking advantage of recently developed econometric methods. There is indeed a burgeoning literature on mixed frequency regression analysis and related econometric techniques. First, to facilitate real-time updating, we use Mixed Data Sampling (MIDAS) regressions to build forecasting models. The key feature of MIDAS regression models is that they allow regressors to be higher frequency than the dependent variable.[2] Hence, annual/quarterly earnings data can be combined with monthly/weekly/daily data in the same regression model. This implies we can incorporate most up-to-date financial and macroeconomic information in the model's forecasts,

[1] Recent studies, such as Bradshaw, Drake, Myers, and Myers (2012), have re-examined this issue and found that a random walk model can outperform analysts at longer time horizons.

[2] For further details, see for example recent surveys on the topic of MIDAS regressions: Andreou, Ghysels, and Kourtellos (2011) and Armesto, Engemann, and Owyang (2010) - the latter provide a very simple introduction to MIDAS regressions.

just as analysts do. To address the issue of (high frequency) data proliferation, we employ forecast combination methods. This allows us to combine forecasts for a large class of models/variables.

Besides the innovation in forecasting methods, we also utilize a large sample of firms (1474 firms in total) in our paper.[3] Recent studies, such as Bradshaw, Drake, Myers, and Myers (2012), argue that the selection of firms has a significant effect on the conclusions drawn in this field of study. Including a large number of firms enables us to appraise the performance our models against analysts in different industry subgroups and firm size subgroups.

We carried out the study in two main steps: (1) selecting a list of predictors - examining their in-sample correlation with earnings; and (2) developing a mixed frequency regression model to predict corporate earnings out-of-sample.

Our study yields some surprisingly sharp results. Utilizing the selected list of macroeconomic, financial and accounting predictors in addition to past earnings numbers, we are able to predict a significant portion of the movements in quarterly earnings and consistently outperform analysts at various forecast horizons. In particular, we find that even at very short horizons (the end of the target quarter of the forecast exercise), as far as statistical significance goes, analyst forecasts only outperform our model in a small 4 % of firms and therefore in 96 % of the cases it is either a draw or better (namely for 25 %). Moreover, the superiority of forecasting performance is more evident for cyclical industries.

Our paper relates to many fundamental research papers aiming at finding fundamental signals that can predict earnings. Ou and Penman (1989) carried out a statistical search of 68 financial statement descriptors and identified 17 or 18 to include in a panel logit model predicting earnings movements in-sample. Lev and Thiagarajan (1993) examined 12 most-commonly-cited accounting variables in analysts' earnings pronouncements in-sample, and reported the signs of the parameters estimated on these variables. Abarbanell and Bushee (1997) selected 9 out of those 12 variables and found that although significant, some of the signs of the in-sample estimated parameters are different from Lev and Thiagarajan (1993). Our paper relates to this literature by providing in-sample time series correlation analysis between earnings and a number of accounting variables, which complements the results in the aforementioned studies based on cross-sectional or panel regressions. However, the list of predictors we use is more extensive and includes macroeconomic and financial variables as well.

Another related field is the growing literature of MIDAS-based forecasting. The MIDAS regression framework was first applied to returns and volatility forecasting; see for example Ghysels, Santa-Clara, and Valkanov (2005), among others. A number of studies have also adopted MIDAS regressions in predicting macroeconomic variables, see for example, Clements, Galvão, and Kim (2008), Armesto, Hernández-Murillo, Owyang, and Piger (2009), Kuzin, Marcellino, and Schumacher (2011), Andreou, Ghysels, and Kourtellos (2013). Our paper is the first study which applies MIDAS regressions to the prediction of corporate

[3] A similar study applying forecast combination in firm earnings forecasting only included 30 firms. See Bansal, Strauss, and Nasseh (2012)

earnings.[4]

The remainder of the paper is organized as follows. Section 2 describes the earnings data as well as predictors used in our forecasting model. Section 3 introduces the econometric methods employed in this paper, namely stationary bootstrap method, MIDAS regressions, principal component forecast combination, and forecast evaluation methods. Section 4 presents the empirical results. Section 5 concludes.

2 Description of the Data

We forecast individual firms' quarterly earnings with real-time macroeconomic series as well as firm-specific financial and accounting variables. Historical vintages of both the earnings and the series used as predictors were collected from multiple databases. The following two subsections describe the sources of the raw data and the transformation carried out on each series.

2.1 Earnings Actuals and Forecasts

Earnings series are often relatively short and infrequent (quarterly at best). The median firm in the Center for Research in Security Prices (CRSP) database has a listing age of 10 years (see Loderer and Waelchli (2010)), which leaves only 40 quarterly observations for a typical firm. If split evenly to obtain an in-sample and an out-of-sample portion, there are only 20 observations for the in-sample estimation portion. While the available time series data are often limited, ideally one would need to include many predictors in order to approximate the large information set of analysts such as firm-specific financial statement variables, equity market returns and volatility, as well as macroeconomic indicators. The regression-based nature of a time series model therefore imposes some restrictions on the number of variables that can be included. To solve the potential data proliferation problem we will resort to forecast combination techniques, as will be discussed later.

We construct firms' actual and predicted earnings from the unadjusted quarterly earnings detail file retrieved from Institutional Brokers' Estimate System (I/B/E/S). I/B/E/S records firms' Earnings Per Share numbers as well as individual institutional analysts' forecasts for as many as 40,000 firms in 70 markets, together with the dates earnings and their forecasts are released. We restrict our search to listed U.S. firms and adjust the I/B/E/S Earnings Per Share actuals and forecasts by adjustment factors downloaded from the CRSP database to remove the effects of stock splits.

In order to ensure enough observations for model estimation, we exclude firms with less than 15 years of consecutive quarterly earnings actuals. Note this restriction somewhat

[4]To be more precise, Ball and Easton (2013) use MIDAS regressions to identify different elements of the accounting system by using high frequency returns to explain contemporaneous earnings. In particular, they use MIDAS regressions to explain how earnings reflects economic shocks (as proxied by daily stock returns). In the current paper MIDAS regressions are used for the purpose of out-of-sample forecasting of earnings.

biases our sample towards larger and more successful firms. Table 1 shows a breakdown of our sample by industry affiliations. As indicated at the end of the table, the total number of firms in our sample is 1474 with a large fraction of manufacturing and high tech firms.

We construct consensus forecasts by taking the median of individual analysts' forecasts made for a given forecast horizon. We will refer to the *target* quarter (TQ) as the object of interest in the predictions exercise for both analysts and our models. We consider four forecast scenarios in this paper: (1) one quarter ahead of the target quarter - implying a one-quarter ahead forecast horizon, (2) one month into the target quarter, (3) two months into the quarter and finally (4) at the end of the forecast quarter. Compared to other papers in the literature, we focus on short forecast horizons, as these are forecast horizons where institutional analysts have the most success against time series models.

2.2 List of Predictors

Predictors used in our paper fall under three categories: macroeconomic variables, firm-specific stock return and volatility, and firm financial statement variables. Macroeconomic variables other than industrial production are retrieved from Federal Reserve Economic Data (FRED) maintained by Federal Reserve Bank of St. Louis. Industrial production uses real-time data vintages provided by Federal Reserve Bank of Philadelphia. Stock returns data are downloaded from the CRSP database. Return is measured as the excess stock return over the corresponding industry portfolio.[5] Volatility is calculated as the 22-day moving average of squared daily stock returns. Financial statement variables are constructed using data from COMPUSTAT. To make the data stationary we take yearly growth rates, i.e. each quarter we calcute the growth rate with respect to the same quarter the previous year. Table 2 provides a list of the variables, their sampling frequencies used in our model and their definitions.

Macroeconomic variables reflect the state of the economy and aggregate demand conditions. These variables are especially meaningful in capturing the fluctuations in earnings due to business cycle conditions.[6] Financial statement variables are widely cited by industry analysts when making inferences regarding earnings. These fundamental signals are specific to each firm and measure its growth prospects, profit-generating ability and cost and expense management efficiency. Besides accounting variables, excess stock return and volatility are also firm-specific. These equity market variables convey the market's assessment of the value of a firm's equity, often informative of its earnings potential. Compared with accounting variables, equity market variables are more forward-looking and are updated daily, allowing us to incorporate more timely information in our prediction model.

[5]The construction of industry portfolios follow "10 Industry Portfolios" on Kenneth R. French's website. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[6]The time period of our study includes two recessions: April,2001 to November 2001 and January 2008 to June 2009, as defined by NBER.

3 Econometric Methods

This section describes an in-sample bootstrap correlation testing method and an out-of-sample forecasting model. The former can be viewed as a prelude to the selection of a good forecasting model. We begin, however, with two general observations.

First, there is a great degree of heterogeneity in different firms' responsiveness to changes in fundamental factors, the theoretical rationale of which will be discussed in detail in Section 4.1.1. Based on this observation, we adopt a strategy of (1) treating each firm individually rather than as a panel in our testing and forecasting models, and (2) then summarizing the results across all firms or subgroups of firms.

Second, the effects of the predictors on earnings can be time-varying due to business cycles or firm life-cycles. For example, macroeconomic factors may have more bearing on earnings during recessions, while firm-specific factors play a bigger role during less tumultuous business conditions. Therefore, we strive to capture a robust relation between earnings and each prediction series by applying a bootstrap method for the in-sample correlation analysis. We also make all forecasts in a rolling fashion, where forecasts are based only on series' values up to the date the forecast is made, and model parameters are re-estimated as the forecast date progresses. In addition, the dynamic instability in the forecasting ability of the predictors favors forecast combination of multiple models each using one predictor over estimating a single model with the all predictors in one regression.[7]

3.1 In-sample Correlation Analysis

Correlation analysis aims at examining whether seasonal earnings differentials $\Delta_s EPS_t$, defined as the year-to-year changes $\Delta_s EPS_t \equiv EPS_t - EPS_{t-4}$, correlate significantly with the predictors we proposed in section 2. Namely, for each predictor X (X can be the growth rate of GDP, or excess stock returns, for instance), we examine the correlation, and its sign, between ΔEPS_t and X with different numbers of lags ranging from zero to two quarters.

The statistical method we use to test $H_0 : corr(\Delta_s EPS_t, X_{t-j}) = 0$ ($j = 0, 1, 2, 3$) applies the stationary bootstrap procedure proposed in Politis and Romano (1994). The advantage of this procedure is that it allows the two tested series to exhibit serial correlation, as long as they are both stationary. The appeal of the procedure is that it does not require the explicit specification of the data generating processes of the two series (apart from some mild regularity conditions). The key inputs include the average block size, and number of bootstrap simulations. Given these inputs, the two series $\Delta_s EPS_t$ and X_{t-j} are re-sampled by blocks of random sizes and the bootstrap generates pseudo-time series $\Delta_s EPS_i^*$ and X_i^* ($i = 1, \ldots, N$ where N is the number of bootstrap replications). Sample correlation can be calculated from the simulated pseudo-time series $\beta_i \equiv corr(\Delta_s EPS_i^*, X_i^*)$. Thus β_i ($i = 1, \ldots, N$) form the empirical distribution of $corr(\Delta_s EPS_t, X_{t-j})$), with which we can test the null hypothesis that the two series are uncorrelated in population, against $H_a : corr(\Delta_s EPS_t, X_{t-j}) < 0$ or $H_a : corr(\Delta_s EPS_t, X_{t-j}) > 0$.

[7] See subsection 3.2.2 for an overview of the advantages of forecast combination.

The possible outcomes of the bootstrap test of correlation are positive, negative or insignificant. The test is carried out for each firm. We then calculate percentage of firms where the correlation is positive or negative, respectively. Any percentage that is above 5%, which is the significance level we used in the bootstrap test, is considered significant. If the percentage of firms exhibits positive correlation between their earnings and predictor x dominates the percentage of negative correlations, we consider predictor X to be positively correlated with earning differentials for an average firm. By examining various lags of X, we can also observe whether the correlation changes over time horizons, allowing the predictor to display its effect on earnings with some lags.

3.2 Out-of-sample Prediction Models

We set up a rolling-window out-of-sample forecasting model utilizing all fourteen predictors in Table 2. For a given in-sample estimation window, each predictor is used in the regression model one at the time, generating one forecast per series. The fourteen forecasts are then combined to produce the final model forecast series. The performance of our one-step-ahead forecast is evaluated against the median consensus of analyst forecasts at four forecast horizons: at the end of the forecast quarter, two months into the forecast quarter, one month into the forecast quarter and at the end of the previous quarter.

3.2.1 MIDAS Regressions

The in-sample correlation analysis in subsection 3.1 is supposed to be a reality check to see whether the earnings are correlated with different predictors in a way that is consistent with theory. This gives us some comfort that the forecasting regressions we are to discuss are not spurious. The correlation results also complement the accounting literature on earnings' prediction, as previous studies which identified factors were done on panel data, while our analysis is done in a time series setting.

There are two reasons why we take a slightly different approach to measuring earnings growth. First, for the correlation analysis we picked seasonal differences $\Delta_s EPS_t$, as it was a convenient way to deal with the seasonal fluctuations in the data.[8] Second, to make our forecasting models compatible with analyst predictions, we take quarterly growth rates, i.e. $\Delta EPS_t \equiv EPS_t - EPS_{t-1}$ and accommodate for seasonal fluctuations in the formulation of our regressions.

We start with accounting variables as predictors of earnings growth because these predictors are of quarterly publication frequency, i.e. the same as earnings. The case of high frequency data will be covered later. For each accounting variable, an Augmented Distributed

[8] We calculated the correlations between first differences of earnings and each predictor's lagged values. The results were mostly similar but not as strong results as with quarterly differencing, due to the seasonality.

Lag (ADL) model is used to generate the forecasts by this variable.

$$\Delta EPS_{t+1} = c + \sum_{i=1}^{3} \delta_i D_i + \sum_{j=1}^{p_Y^Q} \alpha_j \Delta EPS_{t-j} + \sum_{j=1}^{q_X^Q - 1} \beta_j X_{t-j}^Q + \mu_{t+1} \tag{1}$$

where D_i (i=1,2,3) are quarter dummies and X_t^Q is a series of quarterly accounting series. The number of lags p_Y^Q and q_X^Q are selected by Bayesian Information Criterion (BIC).

The timing of lags requires some further clarification, because quarterly earnings as well as accounting variables have publication lags. Most firms release their quarterly financial statements within the first month of the subsequent quarter, sometimes in the second month of the subsequent quarter. We retrieve the announcement dates from I/B/E/S database and use real-time observations, i.e. the last available quarter's numbers for both past earnings and past accounting predictors. For example, if a firm regularly releases its quarter t statements on the 15th of the first month into quarter $t+1$, then forecasts made at the end of quarter t can not be based on ESP_t, but rather ΔEPS_{t-j} for $j \geq 1$. The same applies to X_t^Q as those numbers are yet to be released as well. Hence, the timing of data on the right hand side of equation (1) are t - 1 only. However, when we discuss next regressions involving higher frequency data it should be noted that by the end of one month into the forecast quarter, the previous quarter's numbers are available.

Other predictors, whether macroeconomic or equity market performance variables, are available at monthly or even daily frequency. We could include quarterly values of these variables, which are typically the sum or average of the monthly (daily) values. There are two drawbacks of using a predictor's quarterly values when higher frequency values are available: First, it restricts the effects of the predictors to be constant across different months (days) in the same quarter. For example, if $X_t^Q \equiv X_{1,t}^M + X_{2,t}^M + X_{3,t}^M$, only one coefficient β_j in equation (1) for each quarterly lag j is estimated. One may, however, have reasons to believe that changes in the predictor that occur in different months do not have the same effect on ΔEPS_{t+1}.[9] By taking the simple average of monthly observations, one loses information. Second, updating can happen only once each quarter in the model, after all the monthly (daily) numbers are released. This means for variables without publication lag (equity market or interest rates), forecasts made with such predictors at the end of the previous quarter are the same as forecasts made two months into the forecast quarter. Hence, one foregoes the real-time flow of information throughout a quarter.

Due to the aforementioned drawbacks of using quarterly predictor values when higher frequency data are available, we use an Augmented Distributed Lag - Mixed Frequency Data Sampling ("ADL-MIDAS") Regression model to generate forecasts made with monthly macroeconomic and equity market variables. There is ample evidence that such regressions - which take advantage of real-time high frequency data - can significantly improve predictions of quarterly macroeconomic variables, using either monthly or daily data, see e.g. Schumacher and Breitung (2008), Clements, Galvão, and Kim (2008), Armesto, Hernández-

[9] One such instance is that stock returns in the month closest to the announcement of earnings may contain more information regarding earnings.

Murillo, Owyang, and Piger (2009), Kuzin, Marcellino, and Schumacher (2011), Andreou, Ghysels, and Kourtellos (2013), among many others.

We will use a double index for monthly data, namely let $X_{t,i}^M$ be the observation of month i in quarter t. When forecasts of a given quarter's earnings are made at the end of the previous quarter, the ADL-MIDAS regression model looks like the following:

$$\Delta EPS_{t+1} = c + \sum_{i=1}^{3} \delta_i D_i + \sum_{j=1}^{p_Y^Q} \alpha_j \Delta EPS_{t-j} + \beta \sum_{j=0}^{q_X^M - 1} \sum_{i=1}^{3} \omega_{j*3+i} X_{t-j,i}^M + \mu_{t+1} \tag{2}$$

This regression essentially assigns a slope parameter to each monthly lagged observation of the predictor ($\beta\omega_{j*3+i}$).[10]

When forecasts are made some time into the forecast quarter, information beyond period t becomes available. We integrate this information with an ADL-MIDAS with Leads model that provides the advantage of real-time updating. For example, with one extra month of within quarter information, we have:

$$\begin{aligned} \Delta EPS_{t+1} &= c + \sum_{i=1}^{3} \delta_i D_i + \sum_{j=1}^{p_Y^Q} \alpha_j \Delta EPS_{t-j} \\ &\quad + \beta \sum_{j=0}^{q_X^M - 1} \sum_{i=1}^{3} \omega_{j*3+i} X_{t-j,i}^M + \tilde{\omega}_1 X_{t+1,1}^M + \mu_{t+1} \end{aligned} \tag{3}$$

In equation (3), we can in principle use $k = J_X^M = 1$ (as in the above equation) or 2, 3 number of lead months.[11] When k equals three, the forecasts are made at the end of the forecast quarter, a case often referred to as "nowcasting". Note that we always align analyst forecasts with the real-time specification in equation (3).

To sum up, each of the fourteen predictors in Table 2 is included in a separate forecasting regression, generating a series of rolling-window out-of-sample forecasts. When the predictor X is only available on a quarterly basis, equation 1 will be used; while for monthly-available series X, either equation (2) or equation (3) will be adopted depending on the forecast horizon. When all fourteen forecasts are made, we then apply a principal forecast combination approach outlined in the following subsection to yield a model forecast series.

[10] Regression (2) is referred to as "U-MIDAS", i.e. Unconstrained MIDAS in the literature, following Foroni, Marcellino, and Schumacher (2013), as opposed to imposing some structure to the weighting scheme ω as is typical in MIDAS regressions.

[11] For the variables with a publication lag, i.e. industrial production and inflation, the number of leads equals k minus publication lag measured in months. We retrieved the announcement dates of these two macroeconomic variables and determined that the numbers for a given month are usually released at the middle of the next month, thus we assumed that the publication lag of these two variables is 1.

3.2.2 Principal Component Forecast Combination

Forecast combination has been accepted in the literature as an effective way to summarize information provided by many predictors. Timmermann (2006) points out that compared to estimating a forecast model with all predictors in one regression, carrying out the regression one predictor at a time and using forecast combination methods is more robust to model misspecification and measurement errors. The combined forecast also performs better in the presence of structural breaks and model instability. In addition, as also noted by Timmermann (2006), many studies have shown that forecast combination is superior to the best-performing individual forecasts.

There are many ways to form a combined forecast from a given number of individual forecasts. Basically, one needs to select a set of dynamic weights assigned to individual forecasts $\omega_{i,t}$ yielding a weighted average forecast combination:

$$cf_{t+h|t} = \sum_{i=1}^{I} \omega_{i,t} f_{i,t+h|t} \tag{4}$$

where $I = 14$ in our application, the number of individual series used in each of the individual MIDAS regressions. Diebold and Lopez (1996) surveyed the literature on forecast combination methods and categorized them into two groups, "variance-covariance" methods and "regression-based" methods.[12]

In our paper, we follow the principal component forecast combination method proposed by Chan, Stock, and Watson (1999) and Stock and Watson (2004). This method is an extension of the regression-based forecast combination method where one uses a few principal components extracted from the panel of individual forecasts, with the number of principal components determined by the ICp3 criterion proposed in Bai and Ng (2002).

To form a combination forecast($cf_{t+h|t}$) at time point t from individual forecasts ($f_{i,t+h|t}$, $i = 1, \ldots,$ I), we first extract principal components of the panel of forecasts $f_{i,s+h|s}$ (i=1:I, s=1,...t), then use the history of the principal components and realizations to estimate the following regression:

$$y_{s+h} = \sum_{j=1}^{N} \hat{\lambda}_j PC_{j,s+h|s} + v_{s+h} \tag{5}$$

where $s = 1, \ldots, t-h$, and finally we use the estimated coefficients to generate $f_{t+h|t}$:

$$cf_{t+h|t} = \sum_{j=1}^{N} \hat{\lambda}_j PC_{j,t+h|t} \tag{6}$$

We denote the coefficient matrix of the principal component analysis by $C_{I\times I}$. There-

[12] Granger and Ramanathan (1984) has shown that the optimal variance-covariance combining weight vector has a regression interpretation as the coefficient vector of a linear projection of realizations onto the individual forecasts.

fore, the weight assigned to individual forecast series $f_{i,t+h|t}$ can be calculated by equation: $\sum_{j=1}^{N} \hat{\lambda}_j * C_{i,j}$. We include an analysis on the weights in Section 4.2.

Now that the 14 forecast series have been combined into one, we continue to describe how we evaluate the performance of this combined model forecast series.

3.2.3 Forecast Evaluation Methodology

The out-of-sample forecast performance of the combined model forecasts needs to be appraised against predictions made by other benchmark models and the consensus of analysts and measured in terms of a loss function. Which one to use is an empirical question. Our choice of loss function is in part driven by the observation that there are often outliers in analysts' earnings forecast error series. Various studies have identified and tried to explain the presence such outliers. An important paper on this topic is Abarbanell and Lehavy (2003), where the authors identified that there is a small occurrence of extreme negative values in analysts' earnings forecast error series due to firms' recognition of unexpected accruals. To be robust in the presence of outliers in earnings forecast error series, we apply two criteria in evaluating the forecast performance: the median absolute scaled error ratio (MASER) and a rank-based test using the median scaled error.

Let us denote the forecast error sequence from our model (which is based on a forecast combination scheme) by e_m and the error sequence from the benchmark as e_b. Then we define MASER $\equiv median(|\frac{e_m}{\Delta EPS}|)/median(|\frac{e_b}{\Delta EPS}|)$, which is the ratio of relative error medians. A value smaller than one suggests that our model outperforms the benchmark. Since there are slightly less than 1500 firms in the dataset, instead of reporting a MASER value for each firm, we report the summary quantiles of cross-sectional MASER distribution across all firms. The quantiles give an overview of how the model performs against benchmark.

Finally, we use a rank-based test - the Mann-Whitney U Test - which is a non-parametric test for the null hypothesis $H_0 : median(|\frac{e_m}{\Delta EPS}|) = median(|\frac{e_b}{\Delta EPS}|)$.When tested against $H_a : \ median(|\frac{e_m}{y}|) < median(|\frac{e_b}{y}|)$,rejection of the null hypothesis means the model significantly outperforms the benchmark under the chosen significance level. Similarly, rejection of the null when the alternative is $H_a : median(|\frac{e_m}{\Delta EPS}|) > (<) \ median(|\frac{e_b}{\Delta EPS}|)$ implies our model under- (out-)performs against the benchmark. We report the percentages of firms where model outperforms and underperforms analysts, respectively. These test results complement the MASER analysis by examining the statistical significance of the MASER values.

4 Empirical Results

Since we carried out the study in two steps, we describe the results in two steps as well. Subsection 4.1 discusses the empirical correlation between corporate earnings and each of the predictor variables. Subsection 4.2 presents the out-of-sample forecasting performance of our models compared to a number of benchmark forecasts at different forecast horizons. A further breakdown by industry and firm size sub-samples is also carried out.

4.1 In-sample Correlation Analysis

Table 3 presents the results for the correlation analysis. The signs of the correlations conform with economic theory and past literature in general. The subsections below are a variable-by-variable interpretation of the results.

4.1.1 Macroeconomic Variables

Industrial production is a monthly business cycle indicator. Rising industrial production signals economic expansion, and consecutive decreases in production are one of the criteria of a recession. The results in Table 3 show a uniformly positive correlation between changes in industrial production and earnings.

Inflation is a predictor with complex effects on earnings. On one hand, earnings, being nominal dollar figures, should be positively affected by inflation. However, the accounting practice of depreciating fixed assets based on historic cost means that during high inflation periods, the tax benefits of depreciation are lower, and therefore the de-facto corporate tax rate is higher. The results in Table 3 indicate that inflation has a longer term negative effect on earnings, although its contemporaneous quarter's effect is ambiguous.

The default spread is obtained by subtracting ten-year AAA-rated corporate bond yields from a corresponding BAA-rated one. An increase in default spread signals more risk in the overall economy and a deterioration in credit quality. This means that firms face both less favorable macroeconomic conditions and higher borrowing costs. Both channels work to stunt earnings growth. Our results show that for the firms we sampled, default spread is (weakly) negatively correlated with earnings, although there appears to be a gestation lag of a few quarters.

The term spread is the difference between the ten-year Treasury bond and three-month T-bill yields. The market often uses Treasury yields as the benchmark to determine required interest rates on different debt securities. Since firms are, to a greater or lesser extent, net borrowers of long-term funds, an increase of term spread can increase the burden of interest payments and therefore decreasing earnings. The results in Table 3 confirm the negative impact of widening term spreads on earnings.

Treasury bill rates represent the cost of short-term borrowing; thus we would expect an increase in T-bill rates to a raise firm's interest payment expense, and therefore decreasing earnings. However, from the business cycle point of view, higher T-bill rates may be the result of growing demand for funds. For example, Rose (1994) noted that T-bill rates typically rise during economic expansions and fall during recessions. Under this assumption, higher T-bill rates should indicate higher future firm earnings. The results of our empirical analysis show that the correlation between T-bill rates and earnings are indeed complex. An increase in T-bill rates on average suppresses earnings growth at near-term, but causes earnings to rise in the longer-term. The reversion in sign of the correlations across time horizons could be the joint effect of the aforementioned two channels.

Oil prices, similar to interest rates, exhibit their effect on earnings through different

channels. From the cost point of view, since crude oil and its derivative products are used as raw materials for production in many industries, the cost channel suggests a negative correlation between oil price and firm earnings. At the same time, however, oil price is a strong indicator of economic prosperity. Oil price tends to rise when the economy is strong, thus from the demand channel, an increase in oil price elevates earnings. In Table 3, we see a strong positive contemporaneous correlation between oil price and earnings but the sign of the correlation become somewhat ambiguous as lag horizons increase, possibly due to the cost channel.

The VIX is the implied volatility of the S&P 500, sometimes referred to as the "fear index". A sharp increase in the VIX typically coincides with the onset of a recession. As we can see in the results table, firms earnings decrease one quarter after an increase in the VIX.

4.1.2 Equity Market Variables

A firm's stock returns and its volatility are two equity market indicators we include in this study. Positive excess returns, as the correlation analysis results suggest, is a very strong signal for good earnings numbers, while an increase in a firm's stock market volatility correlates negatively with earnings across all time horizons.

4.1.3 Financial Statement variables

We selected five accounting variables based on the previous literature and the availability of data. These variables are referred to as "fundamental signals", quite often cited by analysts when making earnings-based stock recommendations.

Decreases in capital expenditure are often perceived negatively by analysts. When managers have concerns over the adequacy and liquidity of a firm, they may stop or slow down investing in long-term projects. Our results support this point of view and show that a decrease in capital expenditure bodes poorly for future earnings.

An inventory increase that outruns sales increases may be a negative signal that shows a firm is having difficulty marketing its products. However, since inventory contains unfinished products too, and in some industries, there is a significant build-up of inventories before launching a new product, the correlation of inventory and earnings can vary firm-by-firm due to different inventory-holding motives. Our empirical results suggest that for most firms the correlation between the two are significantly different from zero, and a slight majority of firms show a negative response of earnings to an increase in inventory levels, which may suggest that the first channel prevails for these firms.

Profitability is the ratio of sales minus cost of goods and services over sales, also called "gross margin". A higher gross margin consistently indicates higher future earnings.

Selling, General and Administrative Cost (SG&A Cost) that outruns sales usually means the firm may have performed poorly in cost control. However, a temporary increase in SG&A Cost may be due to profit-generating motives such as a firm's decision to make a sales or advertisement campaign. The results in Table 3 show conflicting evidence on the direction

of the effect of SG&A cost increases, with the usual negative interpretation true for more firms than the profit generating motives.

Receivables that increase faster than sales also bode poorly for firms' performance, indicating that the firm may have trouble selling its products, and thus have to enter into more credit extensions. Disproportionate increases in receivables may also lead to higher doubtful provisions for receivables, decreasing future earnings. Our correlation analysis shows that excess growth in receivables is more negatively associated with earnings growth.

The results we obtain through time series testing are consistent with economic theory and the conclusions in previous studies. When aggregating across firms, however, we do see a lot of firm-level heterogeneity on how earnings respond to changes in these variables, especially when there is more than one channel for the tested variable to take effect. Such observed heterogeneity is informative on our choice of a time series over a panel forecasting model.

4.2 Out-of-sample Prediction Performance

For each firm, at each forecast horizon, we estimate the rolling-window out-of-sample forecasting models described in section 3.2. The principal components of the individual model/series forecasts are used to generate a combined, i.e. final, forecast. We evaluate the latter against several benchmark model forecasts made at the same forecast horizon.

4.2.1 Relative to Various Benchmark

At each forecast horizon, we evaluate the combined forecasts from the ADL-MIDAS (with Leads) models against three benchmark forecasts and present the results in Tables 4 and 5.

The first benchmark is a simple extrapolative time series model, namely:

$$\Delta EPS_{t+1} = c + \sum_{i=1}^{3} \delta_i D_i + \sum_{j=1}^{p_Y^Q} \alpha_j \Delta EPS_{t-j} + \epsilon_{t+1} \tag{7}$$

Compared to the simple extrapolative model, the ADL-MIDAS (with Leads) model, described in equations (2) and (3), includes real-time updating using various predictors.

The second benchmark is the quarterly ADL model, where all the predictors are treated as quarterly observations and the model follows the specifications of equation (1). In this benchmark model, non-earnings information is used, although in a less optimal way, as discussed in the methodology section. Note that use again forecast combinations to obtain a single combined prediction, similar to the ADL-MIDAS modeling approach.

The third benchmark, which is also the hardest-to-beat one, is the consensus analysts' forecasts. For each the ADL-MIDAS (with Leads) models we align properly the most up to date analyst predictions. Hence, the analysts' forecasts are of the same timing as ADL-MIDAS model. More specifically, the consensus analyst forecast was calculated as the median

of all the forecasts released on or prior to the forecast date of the ADL-MIDAS models (with leads).

Table 4 summarizes the distribution of the median absolute scaled error ratios between the ADL-MIDAS model and each benchmark model for various forecast horizons. A ratio smaller than one suggests that the new proposed model has a smaller forecast error. Furthermore, Table 5 summarizes the results of the Mann-Whitney tests. The percentage of firms where the ADL-MIDAS approach outperforms the benchmark is presented in the columns "OPF", while the percentage of firms where the proposed model under-performs the benchmark is presented under "UPF".

Let us start with one quarter ahead of the target quarter (TQ) in Table 4. We report the three quartiles of the distribution. We note that the upper 75th tail of the distribution is respectively 0.87, 1.01 and 0.99 against the three benchmark models. Hence, for at least three quarters of the firms the real-time ADL-MIDAS forecast combination approach outperforms any of the three benchmarks considered. This observation also is valid for the two middle panels, covering respectively two and one month into the TQ. As the horizon shrinks (one month in the case of two months into TQ) we see that the edge against analysts consensus forecasts deteriorates, while it remains the same for the other two model-based benchmark models. Yet, even against analysts, the real-time ADL-MIDAS forecast combination approach yields MASER ratios below one for well over half of the firms at the end of TQ, the shortest horizon we consider.

In Table 5 we report that, as far as statistical significance goes, we see that even at the shortest horizon, 25 % of the firms our proposed ADL-MIDAS model outperforms analysts consensus forecasts. In fact, analyst forecasts only outperform our model in a small 4 % of firms. Hence, for the bulk of firms - 71 % (the compliment of 25 % and 4 %) it is a draw and for 96 % it is either a draw or better.[13]

4.2.2 Firm Characteristics

We further investigated the forecast performance of the proposed ADL-MIDAS model within specific subgroups. We focus entirely on one benchmark, namely the consensus forecast of analysts. With this subgroup analysis, we try to examine the relationship between forecast performance and firm characteristics. In the results Tables 6 through 9, the firm group "All Firms" refers to all of the 1,474 firms in our sample. The results under "All Firms" are the same as the last rows in Table 4 and Table 5 and are marked in bold to serve as comparison.

In Tables 6 and 7, we split the firms into five subgroups based on quantiles of firm size measured by average quarterly sales. Forecast performance is summarized within each subgroup, and the results show that our proposed model performs slightly better for smaller firms. For example, in Table 7 we see that analysts outperform our real-time ADL-MIDAS forecast combination scheme only between 0 % and 2 % of the small firm cases - depending on the horizon considered. This pattern is consistent with previous studies. Bradshaw, Drake,

[13]The significance level of the Mann-Whitney test is 5%.

Myers, and Myers (2012),for example, identified that time series models perform better with smaller or younger firms than with larger and more mature firms. However, the observed disparities in forecast performance among different size percentile groups are small, possibly due to the fact that our sample is biased towards larger firms. More specifically, the firms in our "small"-size groups may still be relatively large in the universe of COMPUSTAT firms due to the constraints we imposed on the availability of historical data.

Other than size subgroups, we also categorized the firms by their industry affiliations and summarized the results within different industry subgroups.[14] Tables 8 and 9 present the forecast performance stratified by industry subgroups. We ranked the industries in the table by our model's performance at the end of the target quarter. We achieve better forecast performance with energy, high-tech, manufacturing and consumer durable goods firms than wholesale retail, health, consumer non-durable, utility and telecommunication firms. Our conjecture is that such a pattern emerges because of the varying degrees of sensitivity to business cycles. Cyclical industries are affected more by macroeconomic conditions, which we seem to better exploit in our model as it includes many macroeconomic indicators.

4.2.3 Forecast Combination Weights

The success of our models relative to various benchmarks can be attributed to the application of newly-developed econometrics techniques, namely mixed frequency regressions and forecast combination methods. The models not only use large sources of information, but also synthesize and update, just like analysts. To illustrate the workings of the forecast combination process and therefore shed some light on the dynamics of the model, we examine the weights assigned to each category of predictors in this subsection.

The forecast combination procedure estimates weights in a rolling fashion, giving more consideration to the variables that performed better in the preceding periods. Since weights are estimated separately for each firm and every time period a forecast was made, we opt to show the general dynamics, i.e. the average weights across all firms. Figure 1, 2, and 3 display such weights plotted against the dates on which they were estimated to yield the out-of-sample model forecasts. We use the red dashed lines to show the scenario of equal-weighting. For example, since there are in total 14 predictors, with 7 of them being macroeconomic variables, the weights assigned to this category should be 0.5 (7 over 14) under an equal weighting scheme. The figures show that around the most recent financial crisis, there can be seen a surge in weights assigned to the macroeconomic variables, and a drop in those on the firm-specific accounting indicators. The weights gradually recover towards their pre-crisis levels after 2010. This pattern suggests that our models correctly pick up the economy-wide factors during tumultuous business conditions.

To further understand the advantages of the ADL-MIDAS models with cyclical firms, we report the average weights (across time) given to macroeconomic variables in each industry subgroup. The results in Table 10 are ranked by the weights used in the "End of Target Quarter" forecast scenario, but the rankings are roughly the same in the other three cases.

[14] We use the industry classification of the Famma-French 10 industry portfolios.

We observe that for firms in cyclical industries, higher weights are applied towards the forecasts based on macroeconomic predictors. This is consistent with our conjecture that our models exploit the effects of economy-wide factors on earnings well.

5 Conclusions

Our paper examines the time series correlation between earnings and various macroeconomic, equity market, and financial statement variables. We analyzed the signs of the correlation and ensured that they are consistent with theory and economic intuition. Utilizing these variables as predictors, we use recently developed advances in the econometric analysis of mixed frequency data to formulate real-time forecasting models in a data-rich environment. In particular, ADL-MIDAS regressions are used to obtain forecasts of each firm's earnings at various short term horizons. We evaluated our model against a number of benchmark models including the consensus of analysts' forecasts, and show that we are able to achieve superior performance with a substantial portion of the firms, and match analysts performance with the rest of the firms.

The forecasting framework devised in this study could also be utilized in the future to predict other components on the corporate income statement, such as for example sales.

References

Abarbanell, J., and B. Bushee, 1997, "Fundamental analysis, future earnings, and stock prices," *Journal of Accounting Research*, 35(1), 1–24.

Abarbanell, J., and R. Lehavy, 2003, "Biased forecasts or biased earnings? The role of reported earnings in explaining apparent bias and over/underreaction in analysts earnings forecasts," *Journal of Accounting and Economics*, 36(1), 105–146.

Andreou, E., E. Ghysels, and A. Kourtellos, 2011, "Forecasting with mixed-frequency data," in *Oxford Handbook of Economic Forecasting*, ed. by C. M., and D. Hendry, pp. 225–245.

Andreou, E., E. Ghysels, and A. Kourtellos, 2013, "Should macroeconomic forecasters use daily financial data and how?," *Journal of Business & Economic Statistics*, 31(2), 240–251.

Armesto, M., K. Engemann, and M. Owyang, 2010, "Forecasting with mixed frequencies," *Federal Reserve Bank of St. Louis Review*, 92, 521–536.

Armesto, M., R. Hernández-Murillo, M. Owyang, and J. Piger, 2009, "Measuring the information content of the beige book: a mixed data sampling approach," *Journal of Money, Credit and Banking*, 41(1), 35–55.

Bai, J., and S. Ng, 2002, "Determining the number of factors in approximate factor models," *Econometrica*, 70(1), 191–221.

Ball, R. T., and P. Easton, 2013, "Dissecting earnings recognition timeliness," *Journal of Accounting Research*, 51, 1099–1132.

Bansal, N., J. Strauss, and A. Nasseh, 2012, "Can we consistently forecast a firms earnings? Using combination forecast methods to predict the EPS of Dow firms," *Journal of Economics and Finance*, pp. 1–22.

Bradshaw, M., M. Drake, J. Myers, and L. Myers, 2012, "A re-examination of analysts superiority over time-series forecasts of annual earnings," *Review of Accounting Studies*, 17(4), 944–968.

Brown, L., R. Hagerman, P. Griffin, and M. Zmijewski, 1987, "An evaluation of alternative proxies for the market's assessment of unexpected earnings," *Journal of Accounting and Economics*, 9(2), 159–193.

Chan, Y. L., J. Stock, and M. Watson, 1999, "A dynamic factor model framework for forecast combination," *Spanish Economic Review*, 1(2), 91–121.

Clements, M., A. B. Galvão, and J. Kim, 2008, "Quantile forecasts of daily exchange rate returns from forecasts of realized volatility," *Journal of Empirical Finance*, 15(4), 729–750.

Diebold, F., and J. Lopez, 1996, "Forecast evaluation and combination," .

Foroni, C., M. Marcellino, and C. Schumacher, 2013, "Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials," *Journal of the Royal Statistical Society: Series A* (forthcoming).

Fried, D., and D. Givoly, 1982, "Financial analysts' forecasts of earnings: a better surrogate for market expectations," *Journal of Accounting and Economics*, 4(2), 85–107.

Ghysels, E., P. Santa-Clara, and R. Valkanov, 2005, "There is a risk-return trade-off after all," *Journal of Financial Economics*, 76(3), 509–548.

Granger, C., and R. Ramanathan, 1984, "Improved methods of combining forecasts," *Journal of Forecasting*, 3(2), 197–204.

Kuzin, V., M. Marcellino, and C. Schumacher, 2011, "MIDAS vs. mixed-frequency VAR: nowcasting GDP in the Euro Area," *International Journal of Forecasting*, 27(2), 529–542.

Lev, B., and S. R. Thiagarajan, 1993, "Fundamental information analysis," *Journal of Accounting Research*, 31(2), 190–215.

Loderer, C., and U. Waelchli, 2010, "Firm age and performance," .

O'Brien, P., 1988, "Analysts' forecasts as earnings expectations," *Journal of Accounting and Economics*, 10(1), 53–83.

Ou, J., and S. Penman, 1989, "Financial statement analysis and the prediction of stock returns," *Journal of Accounting and Economics*, 11(4), 295–329.

Politis, D., and J. Romano, 1994, "The stationary bootstrap," *Journal of the American Statistical Association*, 89(428), 1303–1313.

Rose, P., 1994, *Money and Capital Markets*. McGraw-Hill/Irwin, Boston.

Schumacher, C., and J. Breitung, 2008, "Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data," *International Journal of Forecasting*, 24(3), 386–398.

Stock, J., and M. Watson, 2004, "Combination forecasts of output growth in a seven-country data set," *Journal of Forecasting*, 23(6), 405–430.

Timmermann, A., 2006, "Forecast combinations," in *Handbook of Economic Forecasting*. Elsevier, pp. 135–196.

Table 1: Industry Composition of the Sampled Firms

Industry Subgroup	Number of Firms
Consumer NonDurables	72
Consumer Durables	43
Manufacturing	226
Energy	63
High Tec	260
Telecom	16
Wholesale Retail	181
Health	138
Utilities	76
Other	399
Total	1474

Table 2: List of Predictors

Frequency	Category	Predictor	Definition and Transformation
Monthly & Quarterly	Macro Variables	Industrial Production	Year-over-year Growth Rate
		CPI	Year-over-year Growth Rate
		Default Spread	First Differenced Yield Spread between BAA Corporate Bonds and AAA Corporate Bonds
		Term Spread	First Differenced Yield Spread between 10-year Treasury Bonds and 3-month Treasury Bills
		Tbill Rate	First Differenced 3-month Treasury Bills Yield
		Oil Price	Year-over-year Growth Rate
		VIX	First Differenced
	Stock Returns & Volatility	Excess Stock Returns	Firm's Stock Returns Minus Industry Portfolio Returns
		Stock Volatility	22-day Moving Average of Firm's Squared Daily Stock Returns
Quarterly Only	Firm Accounting Variables	Capital Expenditure	Year-over-year Growth Rate
		Inventory	Year-over-year Growth Rate of Inventory Minus Year-over-year Growth Rate of Sales
		Profitability	Year-over-year Growth Rate of (Revenue-Cost of Goods and Services)/Revenue
		SG&A Cost	Year-over-year Growth Rate of Selling,General and Administrative Cost Minus Year-over-year Growth Rate of Sales
		Receivable	Year-over-year Growth Rate of Receivable Minus Year-over-year Growth Rate of Sales

Table 3: In-sample Correlation Results

		Same Quarter		Lag One Quarter		Lag Two Quarters		Lag Three Quarters	
		Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive
Macro Variables	Industrial Production	0%	100%	0%	100%	0%	100%	0%	100%
	CPI	12%	16%	16%	10%	24%	8%	27%	7%
	Default Spread	8%	9%	10%	7%	14%	6%	16%	7%
	Term Spread	28%	5%	22%	5%	17%	5%	17%	7%
	Tbill	11%	5%	8%	6%	7%	8%	5%	8%
	Oil Monthly	6%	32%	9%	28%	12%	19%	17%	11%
	VIX	9%	11%	13%	5%	19%	4%	22%	3%
Firm Equity Variables	Excess Stock Return	2%	39%	2%	42%	2%	45%	5%	29%
	Stock Volatility	27%	4%	29%	4%	26%	6%	20%	8%
Firm Accounting Variables	Capital Expenditure	9%	47%	13%	41%	18%	37%	21%	35%
	Inventory	41%	29%	39%	31%	34%	32%	27%	34%
	Profitability	1%	79%	3%	65%	4%	55%	10%	40%
	SG&A Cost	54%	26%	43%	29%	35%	30%	21%	34%
	Receivable	34%	17%	31%	19%	29%	20%	26%	21%

Note: The entries of the table are percentages of firms where earning changes ($\Delta_s EPS_t$) are positively or negatively correlated with each predictor's contemporaneous or lagged changes (X_{t-j} with $j = 0$ for "Same Quarter", $j = 1$ for "Lag One Quarter", $j = 2$ for "Lag Two Quarters", and $j = 3$ for "Lag Three Quarters"). For given predictor X and selected lag j, the percentage of firms where $H_0 : corr(\Delta_s EPS_t, X_{t-j}) = 0$ can be rejected against $H_a : corr(\Delta_s EPS_t, X_{t-j}) < 0$ (> 0) is reported under column "Negative"("Positive"). The average block size used for the stationary bootstrap test is 12 observations, and the number of bootstrap simulations is 5000. The significance level is set at 5%.

Table 4: Out-of-Sample Predictive Performance Test MASER Distribution

	End of TQ			Two Months into TQ			One Month into TQ			Quarter ahead of TQ		
Benchmark Models	25th	Median	75th	25th	Median	75th	25th	Median	75th	25th	Median	75th
Extrapolative Model	0.62	0.74	0.86	0.64	0.75	0.87	0.64	0.76	0.88	0.65	0.76	0.87
Quarterly ADL Model	0.66	0.79	0.92	0.71	0.83	0.97	0.71	0.85	0.99	0.74	0.88	1.01
Analysts' Consensus	0.63	0.84	1.08	0.62	0.81	1.06	0.58	0.78	1.03	0.54	0.72	0.99

Note: For each firm, forecasts of quarterly earnings made at various horizons (marked by the column groups of the table) by applying the MIDAS model are evaluated against forecasts from the benchmark models at the same horizon. Dividing the median scaled forecast error of the former with that of a later yields a ratio, indicating which model forecasts better. A ratio smaller than 1 favors the MIDAS model against the benchmark model. For example, a value of 0.8 means the former reduces the median forecast error of the later by 20%. To assess whether or how much MIDAS outperforms its counterpart in general, such ratio is calculated for each firm in our sample, and the results form an empirical distribution. The numbers reported in the table are various percentile measurements of this distribution for the given forecast horizons. The benchmark models are described in Section 4.2.1. We will refer to the *target* quarter (TQ) as the object of interest in the predictions exercise for both analysts and our models and consider (1) end of the target quarter, (2) two months into the target quarter, (3) one month into target quarter and (4) a quarter ahead of the target quarter.

Table 5: Out-of-Sample Predictive Performance Mann-Whitney U Test Results

	End of TQ		Two Months into TQ		One Month into TQ		Quarter ahead of TQ	
Benchmark Models	OPF	UPF	OPF	UPF	OPF	UPF	OPF	UPF
Extrapolative Model	38%	0%	36%	0%	35%	0%	34%	0%
Quarterly ADL Model	27%	0%	23%	0%	20%	0%	18%	0%
Analysts' Consensus	25%	4%	26%	4%	30%	4%	35%	3%

Note: The entries of the table are percentages of firms where the MIDAS model outperforms or under-performs each benchmark model when forecasting at different horizons, according to the Mann-Whitney U Test. "OPF"stands for outperform while "UPF" stands for under-perform. The significance level of the Mann-Whitney U Test is set at 5%. The benchmark models are described in Section 4.2.1. We will refer to the *target* quarter (TQ) as the object of interest in the predictions exercise for both analysts and our models and consider (1) end of the target quarter, (2) two months into the target quarter, (3) one month into target quarter and (4) a quarter ahead of the target quarter.

Table 6: Out-of-Sample Predictive Performance Test MASER Distribution by Firm Size

	End of TQ			Two Months into TQ			One Month into TQ			Quarter Ahead of TQ		
	25th	Median	75th	25th	Median	75th	25th	Median	75th	25th	Median	75th
All Firms	**0.63**	**0.84**	**1.08**	**0.62**	**0.81**	**1.06**	**0.58**	**0.78**	**1.03**	**0.54**	**0.72**	**0.99**
Smallest (0-20th percentile)	0.63	0.82	0.99	0.63	0.79	1.01	0.57	0.75	0.97	0.54	0.71	0.93
Small (20-40th percentile)	0.61	0.81	1.08	0.61	0.78	1.01	0.55	0.74	0.96	0.52	0.72	0.93
Medium (40-60th percentile)	0.62	0.79	1.04	0.60	0.80	1.02	0.58	0.76	1.04	0.54	0.70	0.97
Large (60-80th percentile)	0.64	0.84	1.08	0.62	0.84	1.11	0.58	0.78	1.08	0.54	0.74	1.03
Largest (80-100th percentile)	0.67	0.93	1.26	0.65	0.87	1.20	0.65	0.82	1.14	0.56	0.76	1.06

Note: The firms in our sample are divided into five equal-sized subgroups based on sales. The median scaled forecast error ratios between the MIDAS model and consensus analysts' forecasts made at various horizons are calculated for each firm in each subgroup. Summary percentile numbers of these ratios are reported in the table for each subgroup. We will refer to the *target* quarter (TQ) as the object of interest in the predictions exercise for both analysts and our models and consider (1) end of the target quarter, (2) two months into the target quarter, (3) one month into target quarter and (4) a quarter ahead of the target quarter.

Table 7: Out-of-Sample Predictive Performance Mann-Whitney U Test Results by Firm Size

	End of TQ		Two Months into TQ		One Month into TQ		Quarter ahead of TQ	
	OPF	UPF	OPF	UPF	OPF	UPF	OPF	UPF
All Firms	**25%**	**4%**	**26%**	**4%**	**30%**	**4%**	**35%**	**3%**
Smallest (0-20th percentile)	25%	1%	28%	2%	30%	0%	37%	1%
Small (20-40th percentile)	28%	2%	28%	3%	31%	2%	38%	2%
Medium (40-60th percentile)	27%	4%	28%	4%	32%	5%	35%	2%
Large (60-80th percentile)	22%	5%	24%	4%	29%	5%	33%	3%
Largest (80-100th percentile)	20%	7%	23%	6%	26%	6%	33%	5%

Note: The firms in our sample are divided into five equal-sized subgroups based on sales. The percentages of firms in each subgroup where the MIDAS model outperforms or under-performs consensus analysts' forecasts at various horizons according to the Mann-Whitney U Test are reported in the table. "OPF"stands for outperform while "UPF" stands for under-perform. The significance level of the Mann-Whitney U Test is set at 5%. We will refer to the *target* quarter (TQ) as the object of interest in the predictions exercise for both analysts and our models and consider (1) end of the target quarter, (2) two months into the target quarter, (3) one month into target quarter and (4) a quarter ahead of the target quarter.

Table 8: Out-of-Sample Predictive Performance Test MASER Distribution by Industry

	End of TQ			Two Months into TQ			One Month into TQ			Quarter Ahead of TQ		
	25th	Median	75th	25th	Median	75th	25th	Median	75th	25th	Median	75th
All Firms	**0.63**	**0.84**	**1.08**	**0.62**	**0.81**	**1.06**	**0.58**	**0.78**	**1.03**	**0.54**	**0.72**	**0.99**
High Tech	0.60	0.75	0.98	0.59	0.74	0.98	0.54	0.72	0.93	0.52	0.68	0.91
Energy	0.58	0.76	1.02	0.55	0.73	0.92	0.50	0.62	0.81	0.48	0.57	0.76
Manufacturing	0.60	0.78	0.99	0.60	0.77	1.00	0.54	0.73	0.98	0.49	0.65	0.87
Other	0.61	0.80	1.00	0.60	0.78	0.97	0.58	0.75	0.95	0.55	0.71	0.90
Consumer Durables	0.65	0.81	0.92	0.65	0.74	0.96	0.52	0.68	0.96	0.53	0.65	0.78
Wholesale Retail	0.63	0.89	1.17	0.62	0.80	1.17	0.61	0.83	1.16	0.55	0.80	1.07
Health	0.71	0.94	1.16	0.71	0.93	1.18	0.66	0.90	1.15	0.60	0.83	1.08
Consumer NonDurables	0.74	1.00	1.27	0.73	0.95	1.28	0.73	0.95	1.29	0.59	0.88	1.21
Utilities	0.87	1.12	1.44	0.93	1.16	1.40	0.85	1.09	1.36	0.84	1.05	1.34
Telecom	0.76	1.22	1.42	0.82	1.00	1.32	0.70	1.02	1.41	0.67	0.93	1.28

Note: The firms in our sample are divided into ten subgroups based on industry affiliations. The median scaled forecast error ratios between the MIDAS model and consensus analysts' forecasts made at various horizons are calculated for each firm in each subgroup. Summary percentile numbers of these ratios are reported in the table for each subgroup. We will refer to the *target* quarter (TQ) as the object of interest in the predictions exercise for both analysts and our models and consider (1) end of the target quarter, (2) two months into the target quarter, (3) one month into target quarter and (4) a quarter ahead of the target quarter.

Table 9: Out-of-Sample Predictive Performance Mann-Whitney U Test Results by Industry

	End of TQ		Two Months into TQ		One Month into TQ		Quarter ahead of TQ	
	OPF	UPF	OPF	UPF	OPF	UPF	OPF	UPF
All Firms	**25%**	**4%**	**26%**	**4%**	**30%**	**4%**	**35%**	**3%**
Energy	40%	0%	48%	0%	59%	0%	65%	0%
High Tech	34%	3%	37%	3%	40%	2%	43%	2%
Manufacturing	30%	3%	30%	2%	35%	3%	41%	2%
Consumer Durables	27%	0%	29%	0%	38%	0%	50%	0%
Other	24%	4%	26%	4%	30%	4%	36%	2%
Wholesale Retail	20%	4%	18%	4%	24%	6%	27%	4%
Consumer NonDurables	17%	10%	16%	10%	14%	10%	25%	6%
Health	15%	6%	19%	7%	18%	4%	25%	5%
Telecom	7%	20%	13%	13%	7%	13%	7%	7%
Utilities	3%	8%	3%	4%	7%	8%	8%	7%

Note: The firms in our sample are divided into ten subgroups based on industry affiliations. The percentages of firms in each subgroup where the MIDAS model outperforms or under-performs consensus analysts' forecasts at various horizons according to the Mann-Whitney U Test are reported in the table. "OPF"stands for outperform while "UPF" stands for under-perform. The significance level of the Mann-Whitney U Test is set at 5%. We will refer to the *target* quarter (TQ) as the object of interest in the predictions exercise for both analysts and our models and consider (1) end of the target quarter, (2) two months into the target quarter, (3) one month into target quarter and (4) a quarter ahead of the target quarter.

Figure 1: Weights Assigned to Macroeconomic Variables

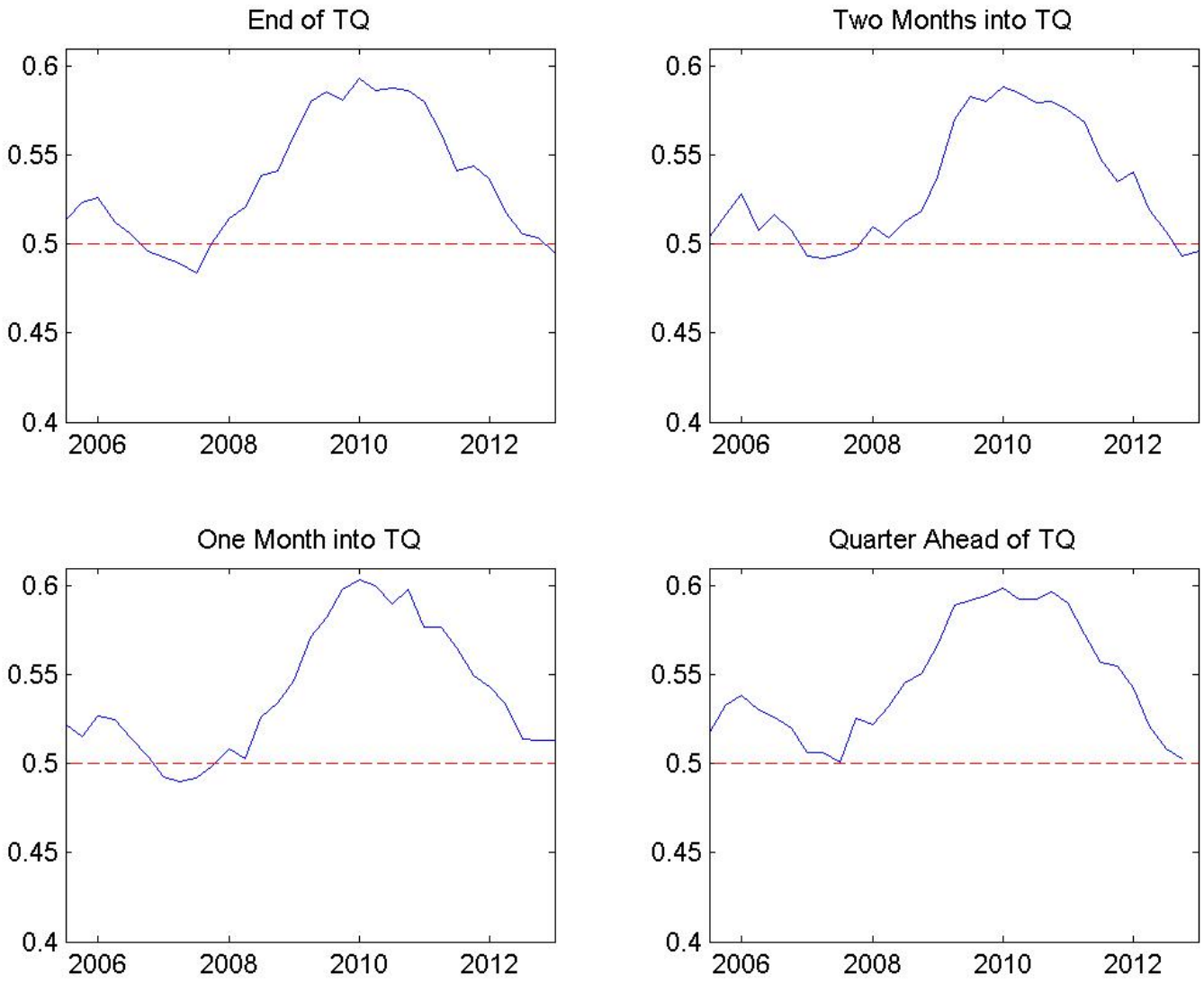

Figure 2: Weights Assigned to Firm-Specific Equity Variables

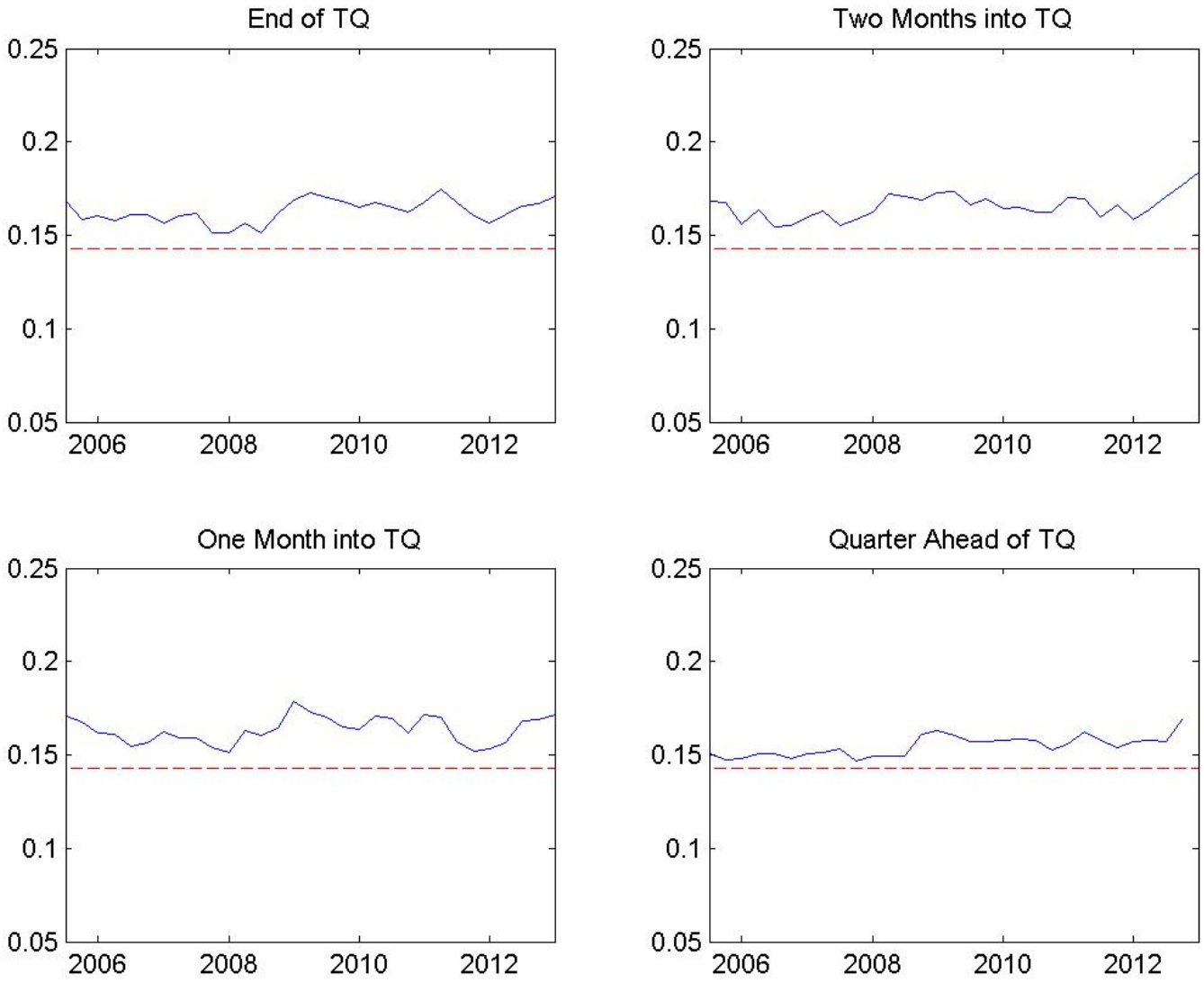

Figure 3: Weights Assigned to Firm-Specific Accounting Variables

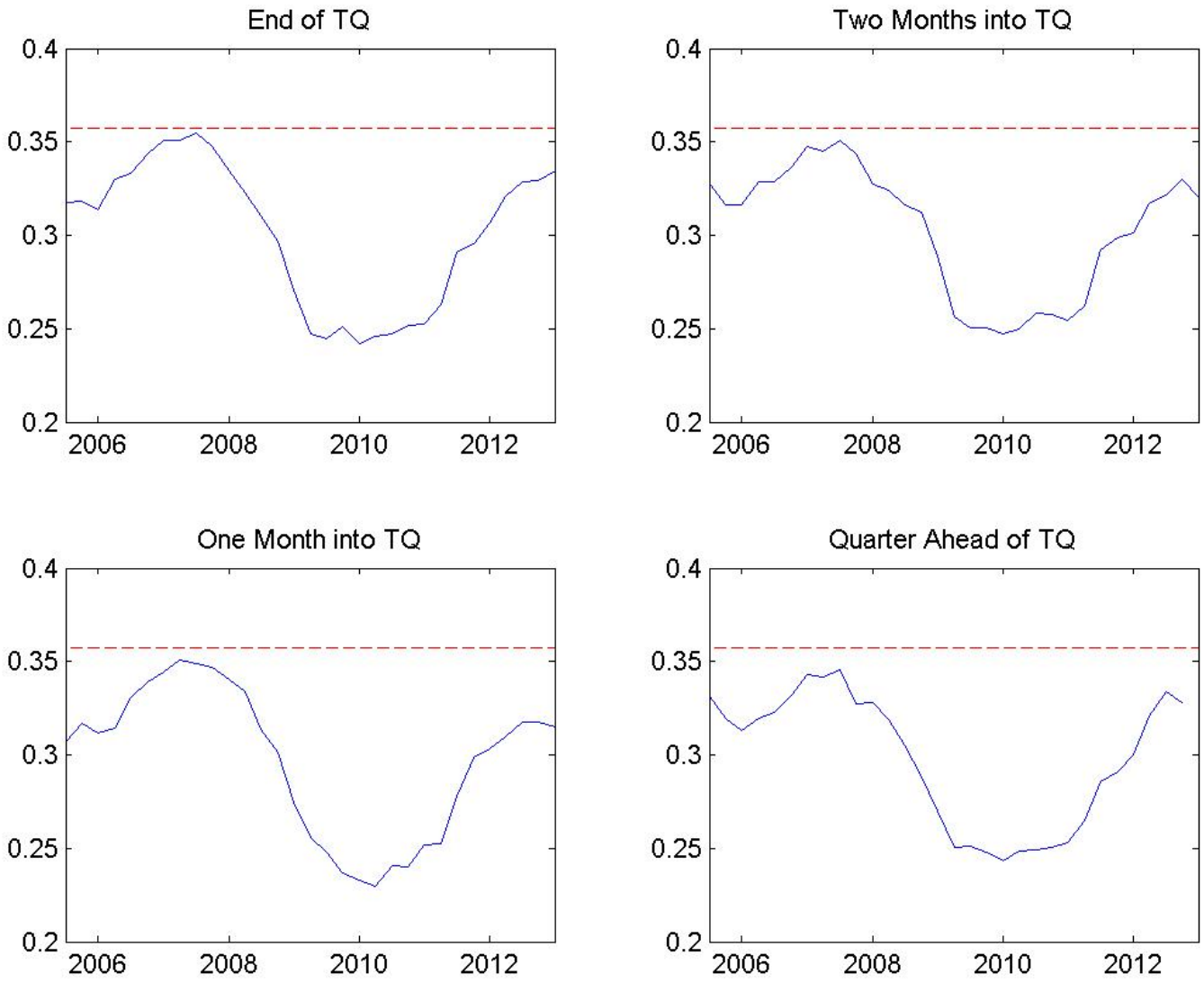

Table 10: Average Weights Assigned to Macroeconomic Variables by Industry

	End of TQ	Two Months into TQ	One Month into TQ	Quarter ahead of TQ
All Firms	**0.551**	**0.544**	**0.555**	**0.561**
Energy	0.607	0.591	0.574	0.604
High Tech	0.560	0.554	0.567	0.581
Consumer Durables	0.559	0.559	0.557	0.571
Manufacturing	0.555	0.549	0.547	0.563
Telecom	0.554	0.526	0.557	0.546
Wholesale Retail	0.553	0.540	0.553	0.553
Health	0.547	0.546	0.558	0.574
Consumer NonDurables	0.546	0.534	0.546	0.561
Other	0.533	0.530	0.548	0.545
Utilities	0.524	0.508	0.525	0.505

Note: The entries in this table are the average weights assigned to the seven macroeconomic variables in ADL-MIDAS models. Such averaging is first done across all the firms in a given subgroup, and then the entire out-of-sample period. Industry subgroups are ranked by the column "End of TQ".