JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **24**: 1187–1206 (2009)
Published online 16 April 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jae.1075

FORECASTING US OUTPUT GROWTH USING LEADING INDICATORS: AN APPRAISAL USING MIDAS MODELS

MICHAEL P. CLEMENTS[a]* AND ANA BEATRIZ GALVÃO[b]

[a] *Department of Economics, University of Warwick, UK*
[b] *Department of Economics, Queen Mary, University of London, UK*

SUMMARY

We evaluate the predictive power of leading indicators for output growth at horizons up to 1 year. We use the MIDAS regression approach as this allows us to combine multiple individual leading indicators in a parsimonious way and to directly exploit the information content of the monthly series to predict quarterly output growth. When we use real-time vintage data, the indicators are found to have significant predictive ability, and this is further enhanced by the use of monthly data on the quarter at the time the forecast is made. Copyright © 2009 John Wiley & Sons, Ltd.

1. INTRODUCTION

The value of leading indicators of US output growth has recently been called into question, in terms of both the performance of individual indicator series, and the only relatively modest improvements that arise from employing combinations of indicators. The recent appraisal by Stock and Watson (2003) of the performance of leading indicators around the time of the 2001 recession concludes that indicators which might have been expected to perform well on the basis of past performance failed to do so (e.g., building permits and consumer confidence), and that some indicators did perform better than expected (namely, stock market indicators). The failure of individual indicators to perform consistently well over a number of recessions is problematic when it is difficult to anticipate in advance which should be watched. One solution is to combine the forecasts from the individual-indicator models. For the period 1999Q1 to 2002Q3, Stock and Watson (2003) find some modest improvements of the order of 5% on MSFE from this strategy relative to an autoregression in output growth.

Our interest is in whether leading indicators can be given a new lease of life by: (i) combining a number of indicators in a single model rather than considering their performance individually; (ii) exploiting the fact that some leading indicators are available at a monthly frequency, and using this monthly information directly (rather than aggregating it to derive quarterly indicator series); and (iii) using the monthly data that will sometimes be available on the quarter when the forecasts of output growth are made. That is, are there better ways of using leading indicator information for predicting output growth? A modelling approach within which we can assess the benefits of all these three aspects is the MIDAS (MIxed Data Sampling) approach of Ghysels *et al.* (2004, 2007), which consists of time-series regressions that allow the regressand and regressors to be

* Correspondence to: Michael P. Clements, Department of Economics, University of Warwick, Warwick CV4 7AL, UK. E-mail: M.P.Clements@warwick.ac.uk

sampled at different frequencies. The MIDAS regression is a parsimonious way of allowing lags of explanatory variables, so that it is feasible to include multiple indicators in a single model. With few exceptions, the approach has so far been applied to modelling high-frequency financial data (e.g., Ghysels *et al.*, 2006, predicting the weekly and monthly future volatility of equity returns using daily realized volatility and intraday squared and absolute returns, and Clements *et al.*, 2008, for exchange rates).[1]

There are other ways of using monthly indicator data to predict quarterly output growth series, such as factor model approaches (Giannone *et al.*, 2008; Schumacher and Breitung, 2008) and 'bridge models' (e.g., Rünstler and Sédillot, 2003; Golinelli and Parigi, 2005; Zheng and Rossiter, 2006). We focus exclusively on MIDAS models, although comparisons of forecasts from MIDAS models with some of these alternative approaches would clearly be of interest.

Our forecast comparisons are influenced by the recent literature on the effects of data vintage on model specification and forecast evaluation (e.g., Robertson and Tallman, 1998; Orphanides, 2001; Croushore and Stark, 2001, 2003; Koenig *et al.*, 2003). In contrast, Stock and Watson's (2003) evaluation uses data from a specific data vintage (2003 Q1 vintage). In addition to using the data vintage available at each point in time, we evaluate the suggestion by Koenig *et al.* (2003) to use 'real-time vintage data', as opposed to the 'end-of-sample vintage data' typical of the conventional approach to real-time estimation and forecasting. Our adoption of a 'real-time' perspective counts against the use of the composite leading indicator (CLI, produced by the Conference Board), as the CLI has been subject to revisions and reweightings of the components over time, partly in response to perceived poor forecast performance (see, for example, Diebold and Rudebusch, 1991). Consequently, our primary data series are the individual component indicators of the CLI.

In summary, our main findings are that: (a) MIDAS is a useful vehicle for combining a small group of indicators for forecasting; (b) the use of information on the current quarter improves forecasts; (c) combination in modelling with MIDAS is better than combination of forecasts when predicting the direction of change of output growth; (d) the use of real-time vintage data (as explained below) relative to end-of-sample vintage data improves forecast performance; and (e) evidence of the predictive ability of the indicators (in comparison with an autoregression) is stronger when the aim is to forecast final data rather than first-released data, although first-released data can generally be forecast more accurately.

The plan of the rest of our paper is as follows. Section 2 describes the models we use. Section 2.1 describes the basic model of Ghysels *et al.* (2004, 2007), and Section 2.2 how it can be adapted for modelling macroeconomic data by including an autoregressive component as in Clements and Galvão (2008). Sections 2.3 and 2.4 note simple extensions to allow multiple indicators and the use of current-quarter data. Section 3 begins by briefly describing the data, and Section 3.1 describes the way in which real-time data are used in our real-time forecasting exercises. The end of Section 3.1 reports our first set of results, which support the use of real-time vintage data over end-of-sample vintage data, and subsequent comparisons are constructed with this in mind. Section 3.2 describes the formal tests we carry out as to whether the leading indicators (LIs) have predictive content, and reports the results. Section 3.3 compares the findings to the results of forecast combination, and also compares the two when the goal is to predict the direction of change of output growth. The following subsection presents results for when monthly LI information is available for the

[1] Clements and Galvão (2008) apply it to forecasting output growth using coincident indicators to obtain 'nowcasts' and short-term forecasts, and Galvão (2007) extends the approach to smooth transition MIDAS models for US output growth. Ghysels and Wright (2008) is also an application to macro-data.

DOI: 10.1002/jae

current quarter (Section 3.4). Finally, we use the same out-of-sample period as Stock and Watson (2003) to assess whether MIDAS models improve output growth forecasts in comparison with a combination of forecasts of single LI regressions. Section 4 offers some concluding remarks.

2. MIDAS REGRESSION APPROACH

2.1. Basic MIDAS Model

The MIDAS approach of Ghysels *et al.* (2004, 2007) models the response of the dependent variable to the higher-frequency explanatory variables as a highly parsimonious distributed lag, as a way of preventing the proliferation of parameters that might otherwise result. Modelling the coefficients on the lagged explanatory variables as a distributed lag function allows for long lags with only a small number of parameters needing to be estimated.

The basic MIDAS model for a single explanatory variable, and h-step-ahead forecasting, is given by

$$yt = \beta_0 + \beta_1 B\ (L^{1/m};\theta)\ x_{t-h}^{(m)} + \varepsilon_t \tag{1}$$

where $B\ (L^{1/m};\theta) = \sum_{j=1}^{K} b(j;\theta)\ L^{(j-1)/m}$, and $L^{s/m}x_{t-1}^{(m)} = x_{t-1-s/m}^{(m)}$. Here, t indexes the basic time unit (in our case, quarters), and m is the higher sampling frequency, and as shown $L^{1/m}$ operates at the higher frequency. The 'Exponential Almon Lag' of Ghysels *et al.* (2004, 2007) parameterizes $b(j;\theta)$ as

$$b(j;\theta) = \frac{\exp(\theta_1 j + \theta_2 j^2)}{\sum_{j=1}^{K} \exp(\theta_1 j + \theta_2 j^2)}$$

For quarterly output growth and monthly indicators as explanatory variables, $m = 3$. We choose to set $K = 12$ so that the model incorporates the last year of monthly data. Thus, $B(L^{1/3};\theta) = \sum_{j=1}^{12} b\ (j;\theta)L^{(j-1)/3}$, and $L^{s/3}x_{t-1}^{(3)} = x_{t-1-s/3}^{(3)}$. An h-step-ahead forecasting model would be specified as

$$y_t = \beta_0 + \beta_1 B\ (L^{1/3};\theta)x_{t-h}^{(3)} + \varepsilon_t \tag{2}$$
$$y_t = \beta_0 + \beta_1[b\ (1;\theta)\ x_{t-h}^{(3)} + b(2;\theta)\ x_{t-h-1/3}^{(3)} + b(3;\theta)\ x_{t-h-2/3}^{(3)} + b\ (4;\theta)x_{t-h-1}^{(3)} + \ldots] + \varepsilon_t$$

The model is specified and estimated on data up to T (say) for the dependent variable, but only $T - h$ for the RHS variables. Data up to T on the RHS are then used with the estimated coefficients to generate a forecast of y_{T+h}. In (2), strictly $\beta = (\beta_0\ \beta_1)'$, θ and ε_t should be indexed by the forecast horizon h. This is the direct forecasting approach reviewed by Bhansali (2002) and Marcellino *et al.* (2006), and considered by Clark and McCracken (2005) in the context of testing for predictive accuracy between rival models. The obvious advantage for models with explanatory factors (other than autoregressive lags) is that models to forecast these variables are not required.

2.2. Autoregressive MIDAS Models

For macroeconomic variables, such as US output growth, autoregressive models often yield forecasts which are competitive with models that include explanatory variables. As an example,

DOI: 10.1002/jae

in the study by Stock and Watson (2003), the combined forecast of the individual-indicator model forecasts is only around 5% better than that of an autoregression. This suggests including autoregressive dynamics in models which incorporate leading indicators as explanatory variables, as in the models of Stock and Watson (2003). However, simply adding a quarterly lag of y to (say) a one-step-ahead forecasting model, i.e.

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 B\ (L^{1/3}; \theta)\ x_{t-1}^{(3)} + \varepsilon_t$$

is not in general appropriate, as noted by Ghysels *et al.* (2007). The reason becomes clear if we write the model as

$$y_t = \beta_0\ (1-\lambda)^{-1} + \beta_1 (1-\lambda L)^{-1} B(L^{1/3}; \theta) x_{t-1}^{(3)} + \tilde{\varepsilon}_t$$

where $\tilde{\varepsilon}_t = (1-\lambda L)^{-1}\varepsilon_t$. It is apparent that the polynomial on $x_{t-1}^{(3)}$ is the product of a polynomial in $L^{1/3}$, $B\ (L^{1/3}; \theta)$, and a polynomial in L, $\sum \lambda^j L^j$, which generates a 'seasonal' response of y to $x^{(3)}$, irrespective of whether $x^{(3)}$ displays a seasonal pattern. We adopt the suggestion of Clements and Galvão (2008) to include autoregressive lags in MIDAS models via a common factor restriction (see, for example, Hendry and Mizon, 1978):

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 B\ (L^{1/3}; \theta)\ (1-\lambda L)\ x_{t-1}^{(3)} + \varepsilon_t$$

The multi-step analogue is

$$y_t = \beta_0 + \lambda y_{t-h} + \beta_1 B(L^{1/3}, \theta)(1-\lambda L^h)\ x_{t-h}^{(3)} + \varepsilon_t \tag{3}$$

which we term the MIDAS-AR. Further details, including estimation of the model, are provided by Clements and Galvão (2008).

2.3. Multiple Leading Indicator Models

As explained in the Introduction, our interest is in the predictive power of models that include multiple indicators, given the unreliability of single-indicator models. The MIDAS specification (1) can be extended in a straightforward way to allow for multiple regressors. The main advantage of MIDAS in this context is that multiple indicators can be included without a large increase in the number of parameters that need to be estimated. An M-MIDAS model that combines the information of n_l monthly leading indicators to predict output growth, h steps ahead, would be written as

$$y_t = \beta_0 + \sum_{i=1}^{n_l} \beta_{1i} B_i (L^{1/3}; \theta_i) x_{i,t-h}^{(3)} + \varepsilon_t \tag{4}$$

where the component indicators are indexed by $i = 1, \ldots, n_l$. Each leading indicator requires the estimation of only two parameters to describe the lag structure (θ_i) and one to weight their impact on y_t (β_{1i}). The β_{1i} parameters define the weights attached to the leading indicators, and are specific to the forecast horizon. Allowing an autoregressive component as in (3) is straightforward.

DOI: 10.1002/jae

2.4. Forecasting with Information on the Current Quarter

Because macroeconomic forecasts are often produced several times each quarter, it is of interest to consider forecasts when monthly data pertaining to the current quarter are partially available. For example, in the model below, s is the forecast horizon in months, so for example $s = 2$ denotes a forecast horizon of 2 months, corresponding to the use of information on x up to and including the first month of the quarter being forecast. When $s > 3$ forecasts are generated of quarters beyond the current quarter but using information on the current quarter (unless s is a multiple of 3):

$$y_t = \beta_0 + \sum_{i=1}^{n_l} \beta_{1i} B_i(L^{1/3}; \theta_i) x_{i,t-s/3}^{(3)} + \varepsilon_t \quad (5)$$

3. EMPIRICAL FORECAST COMPARISONS

We evaluate the forecast performance of MIDAS regressions using the component LIs of the Conference Board's Composite Leading Indicator (CLI), as used by Stock and Watson (2003). The data are monthly from 1959 : 1 to 2003 : 12, and the 10 component series are listed in Table I. To construct quarterly LI data, we average the raw monthly series to obtain quarterly series, and then apply the transformations indicated in Table I, where differencing refers to quarterly differencing. For the monthly data, the same transformations are applied (but we use monthly differencing).

As well as an AR model, we consider as a competitor to the MIDAS approach an autoregressive distributed lag (ADL) model that uses quarterly data on the indicators. A quarterly-frequency ADL(1, p) model that employs x to forecast y h steps ahead is

$$y_t = \beta_0 + \rho y_{t-h} + \sum_{i=0}^{p-1} \beta_i x_{t-h-i} + \varepsilon_t \quad (6)$$

Models are estimated for each h. When the lag orders match up, the ADL model is nested within the MIDAS model (for example, set $\theta_1 = \theta_2 = 0$ in equation (4), when $p = K/3$). In these circumstances, the MIDAS model relaxes the implicit restriction in (6) that the monthly values of the indicator within a particular quarter are assigned the same weight. For a given maximum lag

Table I. Description of leading indicators

Name	Description	Transformation	Available in Real-Time monthly vintages
Spread	Term spread (10-year Federal Funds)	Level	√
Stock	Stock price index (500 common stocks)	Δ ln	√
Hours	Average weekly hours in manufacturing	ln	√
Claims	New claims for unemployment insurance	Δ ln	
Building	Building permits	Δ ln	√
Vendor	Vendor performance diffusion index	ln	
ordersc	Orders—consumer goods and materials	Δ ln	
ordersn	Orders—non-defence capital goods	Δ ln	
Expect	Consumer confidence index (Michigan)	ln	√
M2	Real money supply M2	Δ ln	

DOI: 10.1002/jae

($p = K/3$), we select lags and indicators to enter in the ADL model for each forecast horizon using SIC and the general-to-specific procedure of Birchenhall *et al.* (1999). We consider specifications using 1 year ($p = 4; K = 12$) and 2 years ($p = 8; K = 24$) of past information.

3.1. Real-Time Data and Forecasting Performance

Our basic set of explanatory variables consists of 10 leading indicators (LIs), as described in Table I. Of these, three are not revised (spread, stock and expect), and hours and building are readily available in monthly vintages from the real-time datasets of Croushore and Stark (2001). Real-time vintages of data on GDP growth are available from the same source. Using these data, we undertake two sets of exercises. The first uses real-time data on output growth, and makes use of all 10 indicators, but using final-vintage data on all the indicators. The second restricts the set of indicators to the five available in real time (either because they are not revised, or because the appropriate data vintages are available)[2] and thus it is a proper real-time exercise.

The advantage of using real-time data is that it is possible to replicate the model estimates and forecasts that would have been produced at each point in time. We use vintages of output growth from 1992Q2 to 2003Q1 to generate forecasts of output growth for one, two and four quarters ahead, so, for each forecast horizon, there are 40 forecasts for the period 1992Q2–2003Q4.[3] We also have monthly data vintages on hours and building from, respectively, 1971 : 8 and 1970 : 3.

We will compare the results of taking the 'actuals' to be the 'true values' of output growth, which are generally regarded as being best approximated by the final vintage (2005Q2, in our case), as well as assuming the aim is to forecast first-released output growth. The first release of output growth in the real-time dataset is available at the middle of the following quarter.

Our approach is to recursively estimate our forecasting models as we move the forecast origin forward through the sample. There are two ways of using past data when estimating a regression model in real time. The first one is to use the end-of-sample vintage (EndVint), so, at the forecast origin t, one uses the output growth data obtained from the current vintage made available at $t + 1$ to estimate the forecasting model. This is coupled with the final vintage data on the 10 indicators, or for the five real-time indicators, the data available in the monthly vintage for the first month after the end-of-quarter t (for example: data from the April data vintage if the forecast origin is the first quarter of the year). The second approach was proposed by Koening *et al.* (2003) and is referred to as the use of real-time vintage data (RTVint). This is motivated by a potential problem with the first approach, which is that any given vintage of data will be a mixture of first announcements, data revised only a few times (the more recent values), as well as earlier values which will typically have been revised a large number of times. Estimating the parameters of a model based on a sample consisting of this mixture of data may not result in the best estimates from a forecasting perspective, depending on the nature of the revisions to the LHS and RHS variables.

The real-time vintage data approach consists of using only first-released data to specify and estimate the model at each forecast origin. Let y_t^{t+v} denote output growth in period t from data vintage $t + v$, where $v \geq 1$. Thus the first released value of output growth in t is y_t^{t+1}. The vector of first-released output growth available at time $t + 1$ is $[y_2^3, \ldots, y_{t-2}^{t-1}, y_{t-1}^t, y_t^{t+1}]'$. To estimate an

[2] Real money supply (M2) is excluded from the second exercise as, although it is available in real time, only quarterly (as opposed to monthly) vintages are provided.
[3] That is, we have to generate one-step-forecasts of 1992 : 2 to 2002 : 1, two-step-ahead forecasts of 1992 : 3 to 2002 : 2, etc.

DOI: 10.1002/jae

AR(1), for example, the lagged values of y are also only first-released data, so the right-hand-side data vector is $\left[y_1^2, \ldots, y_{t-3}^{t-2}, y_{t-2}^{t-1}, y_{t-1}^{t}\right]'$. In the case of the MIDAS model, we would also use the first releases from the monthly datasets. To see precisely how this works, suppose $K = 2$ for the sake of exposition, and consider a single monthly LI, $x^{(3)}$, and a model for forecasting one quarter ahead. The last information available at the forecasting horizon on the indicator is $x_t^{(3),t+1/3}$, or $x_t^{t+1/3}$ excluding the superscript, since we are only considering MIDAS models with monthly data on x. Then the estimation of the MIDAS model would draw on the data vectors given by $\left[\ldots, x_{t-2}^{t-5/3}, x_{t-1}^{t-2/3}\right]'$ and $\left[\ldots, x_{t-7/3}^{t-2}, x_{t-4/3}^{t-1}\right]'$. So if t corresponds to the fourth quarter of the year, we are regressing output growth in Q4 on output growth in Q3 from the previous data vintage (y_{t-1}^t) and on the monthly indicators in September and August from October and September vintages (i.e., $x_{t-1}^{t-2/3}$ and $x_{t-4/3}^{t-1}$). Although our models use longer lags of x than the two lags we consider, no new issues arise.

Koenig *et al*. (2003, pp. 627–628) recommend instead taking the lags of the explanatory variables from the same data vintage, so that instead of using the data vectors $\left[\ldots, x_{t-2}^{t-5/3}, x_{t-1}^{t-2/3}\right]'$ and $\left[\ldots, x_{t-7/3}^{t-2}, x_{t-4/3}^{t-1}\right]'$, as explained above, these would be replaced by $\left[\ldots, x_{t-2}^{t-2/3}, x_{t-1}^{t+1/3}\right]'$ and $\left[\ldots, x_{t-7/3}^{t-2/3}, x_{t-4/3}^{t+1/3}\right]'$, while using $\left[\ldots, y_{t-2}^{t}, y_{t-1}^{t+1}\right]'$ as the lagged dependent variable. We find that whether we restrict attention to first-release data on the explanatory variables or take the lags from the same vintage has little effect (as shown in Table II), and our use of real-time vintage data makes use of the former.

Koenig *et al*. (2003) find that using real-time vintage data produces superior forecasts of output growth at short horizons using single-indicator distributed-lag models, and Clements and Galvão (2008) show that their approach applied to MIDAS models produces relatively more accurate 'nowcasts' and short-horizon forecasts.

Table II compares the accuracy of forecasts from these two approaches (EndVint and RTVint). Accuracy is measured by root mean square forecast errors (RMSFE). The results for four forecasting models are presented. The models are: an AR(1) estimated to produce 'direct forecasts' (to match the way forecasts are produced by the other models); a MIDAS with an AR term and using all 10 (or five) LIs (and $K = 24$); a MIDAS model without the AR term and all 10 (or five) LIs; an ADL with maximum $p = 8$ and lag/indicator selection as in Birchenhall *et al*. (1999). Table II(A) records the results for the 10 indicators; Table II(B) the results for the five indicators measured in real time using first-release data; and Table II(C) the results for the five indicators measured in real time but taking lags from the same vintage. For each table, the first panel records RMSFEs relative to when the (same) model is estimated using final data. Consider first the exercise using all 10 LIs. We find gains to the use of real-time vintage output growth data for all models other than the AR. For the AR, the two real-time approaches are broadly equivalent. The second panel of Table II(A) compares the RMSFEs of the four models to the AR benchmark, for both ways of using real-time data. Encouragingly, both MIDAS specifications are more accurate than the AR (by around 10%) when using real-time vintage output growth data, although the same is not true of the ADL. These results are suggestive of the greater accuracy of MIDAS, and some tests of the statistical significance of these differences are run in the next section.

For the full real-time exercise based on the subset of five of the LIs, the results are qualitatively similar, and further strengthen the case for estimating the MIDAS models on RTVint as opposed to final vintage data (first panel of Table II(B)), as well as the gains to using MIDAS relative to the AR

DOI: 10.1002/jae

Table II. Performance of forecasting models with end-of-sample vintages and real-time vintages

Table II(A). Using quarterly vintages of data only for output growth; all 10 leading indicators

	AR		M-MIDAS-AR		M-MIDAS		ADL	
	RTVint.	EndVint.	RTVint.	EndVint.	RTVint.	EndVint.	RTVint.	EndVint.
Benchmark: indicated model estimated with final vintage data								
$h = 1$	1.012	0.996	**0.860**	1.044	**0.943**	**1.070**	**0.863**	**1.062**
$h = 2$	**1.113**	**1.105**	0.974	**1.135**	**0.918**	1.171	1.002	1.040
$h = 4$	**1.062**	1.046	**0.895**	0.980	**0.906**	**0.945**	1.048	**0.928**
Benchmark: AR with the indicated use of real-time data								
$h = 1$			**0.886**	1.091	**0.918**	1.057	1.033	1.292
$h = 2$			**0.930**	1.090	**0.891**	1.144	1.298	1.355
$h = 4$			1.252	1.390	1.305	1.381	1.367	1.228

Table II(B). Using quarterly vintages of data for output growth and the set of five real-time leading indicators

	AR		M-MIDAS-AR		M-MIDAS		ADL	
	RTVint.	EndVint.	RTVint.	EndVint.	RTVint.	EndVint.	RTVint.	EndVint.
Benchmark: indicated model estimated with final vintage data								
$h = 1$	1.012	0.996	**0.853**	1.019	**0.861**	1.036	1.013	1.042
$h = 2$	**1.113**	**1.105**	0.955	1.035	**0.944**	1.007	**0.947**	**1.109**
$h = 4$	**1.062**	1.046	**1.184**	0.995	**1.190**	0.983	1.049	1.035
Benchmark: AR with the indicated use of real-time data								
$h = 1$			**0.859**	1.042	**0.847**	1.034	1.048	1.094
$h = 2$			0.972	1.060	0.974	1.046	1.004	1.184
$h = 4$			1.247	1.062	1.269	1.064	1.215	1.215

Table II(C). Using quaterly vintages of data for output growth and the set of five real-time leading indicators: real time-vintage data computed as in Koenig *et al*. (2003)

	AR	M-MIDAS-AR	M-MIDAS	ADL
Benchmark: indicated model estimated with final-vintage data				
$h = 1$	1.012	**0.901**	**0.908**	**0.935**
$h = 2$	**1.113**	0.964	0.966	**0.904**
$h = 4$	**1.063**	**1.132**	**1.178**	**1.149**
Benchmark: AR				
$h = 1$		**0.906**	**0.892**	0.965
$h = 2$		0.979	0.995	0.957
$h = 4$		1.189	1.254	1.245

For both Tables II(A) and II(B) the entries in the first panel are ratios of the RMSFE for the indicated use of real-time data (either RTVint, which denotes use of real-time data vintages, or EndVint, the use of end-of-sample vintages) to that for when final data are used. For Table II(C) the entries are ratios using real-time vintage data as in Koenig *et al*. (2003) to RMSFEs for when final data are used. In the second panel, entries are ratios of the RMSFE to that of an AR when both models are based on the indicated use of real-time data. Entries with a difference in RMSFE in excess of 5% are emboldened. The forecasting period is 1992Q2 to 2003Q4 (40 observations for each h) and the aim is to forecast final vintage (2005Q2) output growth. Data are available from 1959Q1 and data vintages on output growth are available from 1965Q1. The maximum number of lags of the multivariate models in quarters is 8

DOI: 10.1002/jae

when RTVint is used (second panel) at $h = 1$. However, when forecasting at long horizons ($h = 4$), forecasts with RTVint are worse than with EndVint, which in turn are less accurate than forecasts from the benchmark AR model. The RTVint strategy of Koenig *et al*. (2003) (see Table II(C)) is generally less successful than the first-release RTVint approach recorded in Table II(B) for the models with LIs, but is still preferable to EndVint for the shorter horizons. There is virtually no difference in the case of the AR model (compare the relevant columns in Tables II(B) and II(C)).

Based on the superior results from using RTVint, as evident from Table II, to save space subsequent results are normally only reported for models estimated with real-time vintage data (using first-release data, as in Table II(B)).

3.2. Measuring the Predictive Content of LIs Using MIDAS

Of interest is whether the use of monthly indicator data to predict output growth affects the results of tests of predictive ability compared to when quarterly data are used. Tests of out-of-sample predictability typically compare forecasts that use the information in the indicators against forecasts that do not use this information (generated by an AR model, for example). In this section, we test whether the differences in predictive accuracy between the regressions and the AR model are statistically significant. The tests take into account the uncertainty introduced by parameter estimation in addition to the underlying uncertainty in forecasting. West (1996) is a seminal paper on the impact of estimation uncertainty on tests of predictive accuracy, while West (2006) provides a general review of the literature.

The comparisons we make involve nested models. Compared to an AR, M-MIDAS-AR models may improve forecasting by including explanatory variables. M-MIDAS-AR also nests ADL models by allowing temporal disaggregation of the explanatory variables. The investigation of whether a model that excludes a particular variable is as good as the unrestricted model (which includes that variable) is the usual form of nested model comparison in the literature (i.e., that an ADL nests an AR). The testing approach that has been developed to enable tests of predictive ability for standard nested structures can also be applied to the case of nesting due to temporal disaggregation.

As our evaluation is of multi-step forecasts computed by direct estimation, as well as involving nested models, we follow the approach of Clark and McCracken (2005), who consider testing for equal forecast accuracy in this context. Clark and McCracken (2005) take the test of equal MSE of McCracken (2004) (labelled MSE-F) and show that the limiting distribution depends on unknown nuisance parameters. They find that a bootstrap implementation of these tests (similar to that of Kilian (1999)) appears to work well.

Letting $\text{MSE}_{h,i}$ denote the MSFE for model i at horizon h, then the test of equal accuracy (on MSFE) of the AR and either the MIDAS or ADL model (M) at horizon h is calculated as

$$\text{MSE-}F_h = n\left(\frac{\text{MSE}_{h,AR} - \text{MSE}_{h,M}}{\text{MSE}_{h,M}}\right)$$

when there are n forecasts (see McCracken, 2004; and Clark and McCracken, 2001, 2005).

However, testing for equal predictive ability is more complicated in real-time forecasting exercises when there are data revisions across the vintages, as discussed by Clark and McCracken (2007), who show that data revisions may affect the asymptotic behaviour of tests of equal forecast accuracy. However, their analysis is limited to real-time forecasting exercises based on EndVint, whereas our key interest is in RTVint. Our approach is simply to condition on the data made up

DOI: 10.1002/jae

from the first releases, and treat this as 'the' dataset in much the same way as the works cited above ignore data-vintage effects by using only the final vintage of data. This means that we can only test for equal predictive ability when we use either (i) real-time data in estimation and forecasting (i.e., RTVint) and calculate forecast errors using first-release data, or (ii) use final-vintage data in estimation and forecasting, and forecast evaluation. In this way we avoid the need to specify the (unknown) revisions process.

Hence, the bootstrap procedure to calculate p-values proceeds as follows (for RTVint). An AR model is estimated for the whole sample of first-release output growth data. We simulate data from the estimated model (so that the data are simulated under the null that the nesting MIDAS is no more accurate than the AR). For each sample of simulated data, we apply the same recursive estimation and forecasting procedures as applied to the actual data, to calculate the MSE-F statistic for that replication. (Note that the MIDAS model is estimated using data on the LIs that are held fixed over replications—at their actual first-release values—to avoid having to model the monthly LIs). The p-value is then calculated by comparing the MSE-F statistic for the actual data to the empirical distribution generated by repeated application of the process just described. A similar procedure is applied to compute p-values in the case of final-vintage data.

Table III(A) assesses the predictive content of the leading indicators using final-vintage data (for model estimation, forecasting and evaluation) for the sets of 10 and five indicators separately. Table III(B) reports results based on first-release data for the set of five real-time indicators. Finally, Table III(C) reports results using real-time vintage data to estimate and forecast the models, but final data to compute forecast errors. As noted above, tests of equal predictive ability can be calculated for the first two sets of results but not for the last of the three. Using final-vintage data (Table III(A)), the RMSFE ratios of the models with indicators relative to the AR generally exceed one, indicating no benefit to using LIs. Using first-release data (Table III(B)), some of the RMSFE ratios to the AR are less than unity for $h = 1$, but the p-values indicate that the evidence against the null of equal accuracy is only on the borderline of being significant. In marked contrast to these findings, Table III(C) shows reductions in RMSFE of up to 15% using M-MIDAS to forecast next quarter's output growth. Although we are unable to demonstrate the statistical significance of these reductions in RMSFE, their magnitude suggests they are likely to be of value.

Some insight into the superior performance of the MIDAS model can be gleaned from an examination of the estimated weights on the lagged monthly indicators for the M-MIDAS-AR model (see Figure 1). Figure 1 presents the estimates of the weights for each leading indicator using the full-sample real-time vintage data (RTVint). The estimated weights are generally far from uniform on the sets of adjacent 3 months, which is the restriction imposed by the ADL (but note in general the lag selection procedure for the ADL results in different lags being included compared to those illustrated in the figure for the MIDAS). The weights also vary over the two horizons, $h = 1$, 4, as well as the individual indicators. The estimated weighting functions illustrate the flexibility and variety of distributed lag patterns that can be obtained in a parsimonious way when the M-MIDAS model is used to draw on the information in a small group of leading indicators.

3.3. Combining Indicators versus Combining Forecasts

There is a vast literature on the combination or pooling of forecasts from different models or sources: see, for example, the reviews by Diebold and Lopez (1996), Newbold and Harvey (2002) and Timmermann (2006). Clements and Galvão (2006) consider the relative usefulness

DOI: 10.1002/jae

Table III. Measuring predictive content of leading indicators with ADL and MIDAS

Table III(A). Using only final vintage data

	M-MIDAS		M-MIDAS-AR		ADL	
	$K = 12$	$K = 24$	$K = 12$	$K = 24$	$p = 4$	$p = 8$
With 10 indicators						
$h = 1$	1.024	0.985	1.035	1.042	**0.944**	1.212
$h = 2$	1.084	1.081	1.094	1.063	1.329	1.442
$h = 4$	1.435	1.530	1.434	1.486	1.126	1.386
			MSE-F		MSE-F	
$h = 1$			−2.66[0.25]	−3.17[0.32]	4.94[0.01]	−12.8[0.99]
$h = 2$			−6.56[0.59]	−4.58[0.43]	−17.4[0.99]	−20.8[0.99]
$h = 4$			−20.6[0.99]	−21.9[0.99]	−8.44[0.90]	−19.2[0.99]
With 5 indicators						
$h = 1$	1.033	0.995	1.047	1.019	1.046	1.094
$h = 2$	1.160	1.149	1.156	1.133	1.180	1.184
$h = 4$	1.117	1.133	1.112	1.118	1.230	1.215
			MSE-F		MSE-F	
			−3.48[0.60]	−1.46[0.26]	−5.90[0.93]	−3.44[0.60]
			−10.1[0.99]	−8.84[0.85]	−13.4[0.99]	−11.3[0.97]
			−7.68[0.84]	−8.02[0.84]	−14.3[0.99]	−13.6[0.98]

Table III(B). Using only first-released data

	M-MIDAS		M-MIDAS-AR		ADL	
	$K = 12$	$K = 24$	$K = 12$	$K = 24$	$p = 4$	$p = 8$
With 5 indicators (full RT)						
$h = 1$	0.964	0.996	**0.979**	1.005	**0.978**	1.117
$h = 2$	1.077	1.100	1.109	1.085	**0.969**	1.011
$h = 4$	1.310	1.382	1.214	1.370	1.358	1.368
			MSE-F		MSE-F	
$h = 1$			1.73[0.09]	−0.42[0.18]	1.81[0.08]	−7.94[0.92]
$h = 2$			−7.45[0.96]	−6.00[0.83]	2.56[0.05]	−0.83[0.27]
$h = 4$			−12.9[0.99]	−18.7[0.99]	−18.3[0.99]	−18.6[0.99]

of combining information in modelling versus combing forecasts for quarterly-frequency models of output growth. In this section we compare the forecast performance of the MIDAS regression with multiple LIs to combining the forecasts from single-indicator MIDAS regressions. The form of combination is the simple arithmetic mean (denoted in the table as 'Mean MIDAS'). The results for two values of K and the two groups of LIs considered earlier are presented in Table IV. Using first-release actual values, the results strongly favour the pooling of forecasts whether we consider the 10 or five indicator groups. Using final vintage actuals the multiple-indicator MIDAS model is more than competitive with pooling at the shorter horizons ($h = 1, 2$). In general, though, our results serve to underpin the value of forecast pooling.

Although forecast averaging is often found to be beneficial using measures of accuracy such as RMSFE, this need not be the case when assessing forecasts of the direction of

Table III. (*Continued*)
Table III(C). Using real-time vintage data and forecast errors computed with final vintage data

	M-MIDAS		M-MIDAS-AR		ADL	
	$K = 12$	$K = 24$	$K = 12$	$K = 24$	$p = 4$	$p = 8$
With 10 indicators						
$h = 1$	**0.946**	**0.918**	**0.926**	**0.886**	**0.953**	1.033
$h = 2$	**0.940**	**0.891**	0.961	**0.930**	1.057	1.298
$h = 4$	1.271	1.305	1.196	1.252	1.184	1.367
With 5 indicators (full RT)						
$h = 1$	**0.874**	**0.847**	**0.882**	**0.859**	**0.950**	1.048
$h = 2$	0.968	0.974	0.993	0.972	0.990	1.004
$h = 4$	1.139	1.269	1.065	1.247	1.195	1.215

The entries are ratios of the RMSFE of the indicated models against an AR(1) using real-time vintage data. Emboldened entries indicate that the null of equal forecast accuracy is rejected against the alternative that the indicators have predictive content using the MSE-F statistic (bootstrapped p-values are given in brackets) or (in Table III(C)) that the RMSFE is 5% smaller than the AR benchmark. The forecasting period is 1992Q2 to 2003Q4 ($n = 40$ observations). The data begin in 1959Q1, and data vintages on output growth are available from 1965Q1. In the first panel, the multivariate models are estimated with 10 LIs. In the second panel, the models are estimated with five indicators (chosen either because they are not revised or because monthly data vintages are available)

change. Consequently, we consider the relative performance of MIDAS model forecasts, and pooled individual-indicator MIDAS models' forecasts, versus the AR model benchmark. Suppose interest is in the event that there is a decrease in output growth h quarters ahead, that is, $\pi_{t,h} = I((y_{t+h} - y_t) < 0)$, where $I()$ is the indicator function. We measure the performance of both M-MIDAS and mean MIDAS in predicting this event using the Kuipers score (see Granger and Pesaran, 2000, for a discussion of the relationship between the Kuipers score and economic measures of forecasting performance). A prediction of a decrease in output growth at h steps ahead occurs when $\hat{\pi}_{t,h} = I((\hat{y}_{t+h} - y_t) < 0)$ takes the value 1, where $\hat{y}_{t+h}$ is a forecast from a specific model. The Kuipers score compares the proportion of 'hits', $H(h)$ (forecasts of 1's when decreases occurred) with the proportion of false alarms, $F(h)$ (forecasts of 1's when decreases did not materialize):

$$H(h) = \frac{\sum_{t=1}^{n} \pi_{t,h}\hat{\pi}_{t,h}}{n\overline{\pi}_h}, \; F(h) = \frac{\sum_{t=1}^{n}(1 - \pi_{t,h})\hat{\pi}_{t,h}}{n(1 - \overline{\pi}_h)}$$

where $\overline{\pi}_h = n^{-1}\sum_{t=1}^{n} \pi_{t,h}$ and n is the number of forecast observations. The Kuipers score is then

$$\text{KS}(h) = H(h) - F(h)$$

and lies between -1 and 1 by construction, where a score of 1 is 'ideal'.

The Kuipers scores for M-MIDAS, mean MIDAS and the AR are presented in Table V. The scores indicate that the LIs have predictive ability for the direction of change, as the KS values for both M-MIDAS and mean MIDAS are superior to that for the AR. In this regard, the results are in agreement with the tests of equal forecast accuracy on RMSFE of the previous section. It is also apparent that the M-MIDAS is better than the mean MIDAS when the actuals are final-vintage numbers, which is at odds with the usual finding in favour of pooling in the literature. However,

Table IV. Comparing combining indicators in modelling with combining forecasts

Table IV(A). Using 2005Q2 vintage to compute forecast errors

	M-MIDAS versus mean MIDAS		M-MIDAS-AR versus mean MIDAS-AR	
	$K = 12$	$K = 24$	$K = 12$	$K = 24$
With 10 indicators				
$h = 1$	**1.054**	1.009	0.987	0.952
$h = 2$	1.009	0.967	1.012	0.982
$h = 4$	**1.264**	**1.319**	**1.200**	**1.250**
With 5 indicators (full RT)				
$h = 1$	1.023	0.997	0.988	0.983
$h = 2$	0.982	0.991	**0.940**	0.988
$h = 4$	1.030	1.021	1.044	**1.119**

Table IV(B). Using first-released data to compute forecast errors

	M-MIDAS versus mean MIDAS		M-MIDAS-AR versus mean MIDAS-AR	
	$K = 12$	$K = 24$	$K = 12$	$K = 24$
With 10 indicators				
$h = 1$	**1.070**	**1.085**	1.002	1.031
$h = 2$	**1.124**	**1.054**	**1.097**	**1.085**
$h = 4$	**1.353**	**1.355**	**1.270**	**1.253**
With 5 indicators (full RT)				
$h = 1$	**1.092**	**1.098**	1.035	1.049
$h = 2$	**1.148**	**1.166**	**1.137**	**1.105**
$h = 4$	**1.300**	**1.363**	**1.216**	**1.344**

The entries are ratios of the multiple regressor MIDAS (M-MIDAS) to the combination (mean) of the forecasts from single-regressor MIDAS specifications. In the first panel, the multivariate models are estimated with 10 LIs. In the second panel, the models are estimated with the five indicators indicated in Table I. Entries with a difference of RMSFE larger than 5% are emboldened. All forecasting models are estimated with real-time vintage data (including for the indicators when using five LIs). The forecasting period is from 1992Q2 to 2003Q4

the relative merits of pooling (mean MIDAS) and multiple regression MIDAS (M-MIDAS) are overturned if we consider the five-indicator group and first-release actuals. Nevertheless, which ever data vintage we use to calculate the actual direction of change, the AR model is dominated by the other two.

As well as simply comparing scores, we can use the Kuipers score as the basis of a direction-of-change test—whether the forecast sign is independent of the sign of the actual movement (see Pesaran and Timmermann, 1992; Granger and Pesaran, 2000; and the review by Clements, 2005). However, we found that the test rejected the null of 'no value' at the 10% level for all models and forecast horizons.

3.4. The MIDAS Performance Using Information on the Current Quarter

A potentially important property of the MIDAS approach is that it is able to utilize available monthly information on the current quarter, as described in Section 2.4. Clements and Galvão (2008) show that the use of current-quarter monthly information leads to significant

DOI: 10.1002/jae

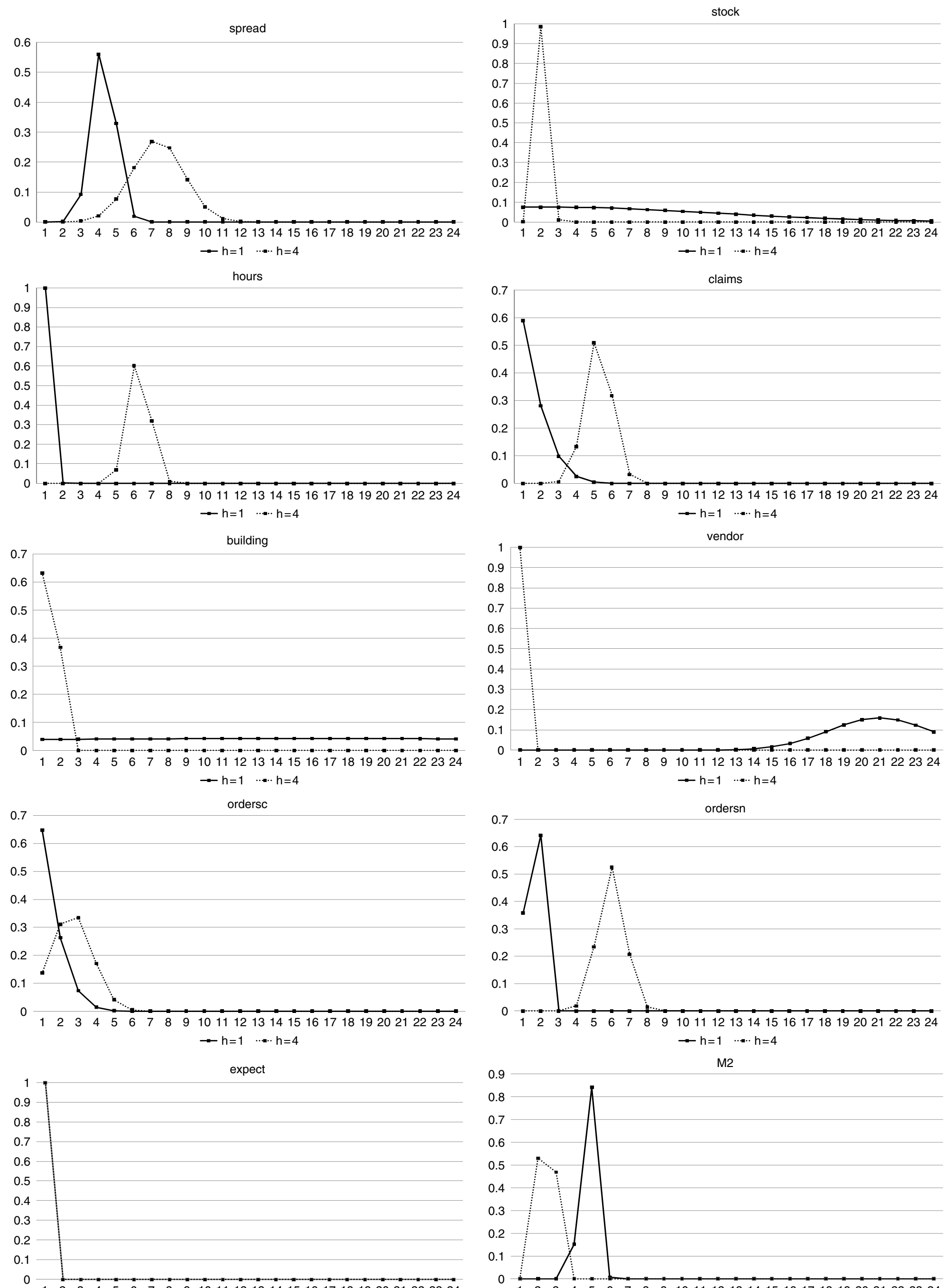

Figure 1. Estimates of M-MIDAS-AR weighting functions of each leading indicator to forecast output growth at $h = 1$ and $h = 4$

Table V. Evaluating the forecasts of direction of change: Kuipers score

Table V(A). Using 2005Q2 vintage to compute forecast errors

	AR	M-MIDAS	Mean Midas
With 10 indicators			
$h = 1$	0.361	0.499	0.499
$h = 2$	0.284	0.535	0.284
$h = 4$	0.323	0.222	0.369
Average	0.323	0.419	0.384
With 5 indicators (full RT)			
$h = 1$		0.594	0.556
$h = 2$		0.506	0.460
$h = 4$		0.374	0.369
Average		0.491	0.462

Table V(B). Using first-released data to compute forecast errors

	AR	M-MIDAS	Mean Midas
With 10 indicators			
$h = 1$	0.450	0.650	0.600
$h = 2$	0.390	0.350	0.346
$h = 4$	0.389	0.404	0.389
Average	0.410	0.468	0.445
With 5 indicators (full RT)			
$h = 1$		0.450	0.650
$h = 2$		0.409	0.399
$h = 4$		0.458	0.417
Average		0.439	0.489

The entries are Kuipers scores for the event that $(y_{t+h} - y_t) < 0$. The Kuipers score is the difference between hit and false alarm rates (see main text). In the first panel, the multivariate models are estimated with 10 LIs. In the second panel, the models are estimated with the five indicators indicated in Table I. All forecasting models are estimated with real-time vintage data (including for the indicators when using five LIs). The forecasting period is from 1992Q2 to 2003Q4

improvements in forecasts based on coincident indicators. We consider whether the use of current-quarter LI information improves output growth forecasts of two quarters ahead and 1 year ahead.

Table VI(A) presents the results of such a comparison for our usual choices of K and groups of LIs, when the actuals are taken to be final vintage data. We compare the RMSFEs of MIDAS models that use information on the first two months of the current quarter to those using only past-quarter information. There are sizeable gains when forecasting next year's growth rate ($h = 4$) for both the 10-indicator information set and the five-indicator real-time set. The second panel of Table VI(A) reports a comparison of the MIDAS using current-quarter information to the AR. With current-quarter information the LIs deliver gains in excess of 10% for two-quarter-ahead forecasts. If instead we use first-release actuals, the use of 'current-quarter' information improves the performance of the MIDAS, but it is only a par with the AR at a two-quarter forecast horizon (see Table VII(B)).

DOI: 10.1002/jae

Table VI. Forecasting using current-quarter monthly indicator information

Table VI(A). Using 2005Q2 vintage to compute forecast errors

	M-MIDAS 10 indicators		M-MIDAS 5 indicators (full RT)	
	$K = 12$	$K = 24$	$K = 12$	$K = 24$
Benchmark: M-MIDAS using only past quarter information				
$h = 2$	**0.936**	0.979	**0.742**	**0.892**
$h = 4$	**0.861**	**0.856**	**0.865**	**0.810**
Benchmark: AR				
$h = 2$	**0.879**	**0.873**	**0.853**	**0.870**
$h = 4$	1.092	1.117	0.985	1.030

Table VI(B). Using first-released data to compute forecast errors

	M-MIDAS 10 indicators		M-MIDAS 5 indicators (full RT)	
	$K = 12$	$K = 24$	$K = 12$	$K = 24$
Benchmark: M-MIDAS using only past quarter information				
$h = 2$	**0.927**	1.003	**0.918**	**0.926**
$h = 4$	**0.840**	**0.835**	**0.828**	**0.746**
Benchmark: AR				
$h = 2$	0.979	0.993	0.988	1.019
$h = 4$	1.136	1.206	1.015	1.033

The MIDAS model forecasts are of two and four quarters ahead, but using monthly indicator information on the first 2 months, so that the effective horizon (in terms of indicator information) is 4 months. The benchmarks are a MIDAS model that does not use the first 2 months of the quarter information (first panel) and an AR model (second panel). The entries are the ratios of the RMSFE of the M-MIDAS (using current-quarter information) to the two benchmarks. Entries with a difference in RMSFE larger than 5% are emboldened. All forecasting models are estimated with real-time vintage data (including for the indicators when using 5 LIs). The forecasting period is from 1992Q2 to 2003Q4

3.5. Measuring the Predictive Power of the Leading Indicators during the 2001 Recession

Stock and Watson (2003) found that the erratic behaviour of individual LI models could be improved upon by combining the forecasts of the individual models. In terms of predicting output growth around the time of the 2001 recession, they found gains from forecast combination of around 5% in terms of MSFE, relative to an AR, when forecasting output growth at horizons $h = 2, 4$ (Stock and Watson, 2003, Table IV). One of the principal motivations of our paper is to see whether better use can be made of monthly indicator data, so of interest is whether our multiple-indicator MIDAS forecasts more accurately around the time of the 2001 recession than the pooled single-indicator ADL model forecasts.

Hence, in this section we compare the forecasting performance of combining forecasts of single ADL regressions against that of MIDAS-AR models for the Stock–Watson out-of-sample period of 1999Q1–2002Q3. Furthermore, by conducting the exercise in real-time we can gauge whether the gains to combination they identified would have been attainable in reality. As described in Section 3.1, we use two ways of exploiting real-time data on output growth (RTVint and EndVint),

DOI: 10.1002/jae

Table VII. Forecasting performance of leading indicators for the period 1999Q1–2002Q3
Table VII(A). Using 2003Q1 data to compute forecast errors

	M-MIDAS-AR		Mean ADL	
	$K=12$	$K=24$	$K=12$	$K=24$
Using final-vintage data				
With 10 indicators				
$h=1$	**0.839**	0.849	0.926	0.929
$h=2$	0.921	**0.908**	0.989	0.994
$h=4$	1.348	1.286	**0.991**	0.993
With 5 indicators				
$h=1$	0.932	0.924	**0.918**	0.995
$h=2$	0.922	**0.899**	0.975	1.028
$h=4$	0.881	**0.872**	0.928	1.134
Using end-of-sample vintage data				
With 10 indicators				
$h=1$	**0.856**	0.935	0.944	0.948
$h=2$	**0.879**	0.974	0.962	0.979
$h=4$	1.230	1.243	**0.992**	0.993
With 5 indicators (full RT)				
$h=1$	0.973	0.950	**0.931**	0.925
$h=2$	0.886	**0.865**	0.937	0.950
$h=4$	0.865	**0.863**	0.925	0.924
Using real-time vintage data				
With 10 indicators				
$h=1$	0.873	**0.791**	0.956	0.966
$h=2$	0.821	**0.807**	0.939	0.944
$h=4$	1.340	1.267	**0.986**	0.986
With 5 indicators (full RT)				
$h=1$	0.788	**0.718**	0.939	0.939
$h=2$	0.853	**0.815**	0.925	0.941
$h=4$	1.148	1.494	**0.976**	1.018

and also use real-time data on the indicators in the same two ways in the case of the subset of five indicators. Finally, we use $K=12$ and 24, and $p=4$ and $p=8$, and the method of Birchenhall *et al.* (1999) to select lags in each single ADL regression. Stock and Watson report results based on $p=4$.

The results for horizons $h=1,2$ and 4 and presented in Table VII. Firstly, consider Table VII(A), where the forecast errors are constructed using final vintage data. The performance of the mean ADL forecasts confirm the results in Stock and Watson (2003): the pooled forecast improves upon the AR at all horizons for the period around the 2001 recession. The use of real-time data (either by EndVint or RTVint) does not have a large impact on the performance of the combined ADL forecast—on a positive note, the gains identified by Stock and Watson are achievable in real time. However, the results for the MIDAS model indicate that much larger gains are possible at the shorter horizons. That is, combining the forecasts of individual ADL model forecasts substantially underestimates the potential predictive power of the monthly leading indicators for output growth during this period. For example, using real-time vintage data, the MIDAS RMSFE is almost 30% lower than the AR at $h=1$ using the five-indicator set, compared to a gain of only around 5% using the ADL. Markedly more accurate forecasts are also obtained

DOI: 10.1002/jae

Table VII. (*Continued*)

Table VII(B). Using first-released data to compute forecast errors

	M-MIDAS-AR		Mean ADL	
	$K = 12$	$K = 24$	$K = 12$	$K = 24$
Using end-of-sample vintage data				
With 10 indicators				
$h = 1$	**0.815**	0.838	0.917	0.922
$h = 2$	**0.831**	1.042	0.934	0.950
$h = 4$	1.253	1.256	0.983	**0.978**
With 5 indicators (full RT)				
$h = 1$	0.947	1.093	**0.917**	0.923
$h = 2$	0.945	0.942	**0.928**	0.939
$h = 4$	0.984	0.991	0.935	**0.933**
Using real-time vintage data				
With 10 indicators				
$h = 1$	0.950	**0.864**	0.965	0.979
$h = 2$	**0.769**	0.792	0.916	0.918
$h = 4$	1.345	1.256	0.977	**0.973**
With 5 indicators (full RT)				
$h = 1$	0.796	**0.739**	0.908	0.919
$h = 2$	0.841	**0.786**	0.908	0.918
$h = 4$	1.209	1.558	**0.961**	1.005

The entries are RMSFE ratios of the M-MIDAS-AR and the mean of single ADL regressions to an AR. Results are presented using recursive forecasts with data from the 2003Q1 vintage (final vintage) and two ways of organizing real-time data: end-of-sample (EndVint) and real-time vintage (RTVint). In the first panel, the multivariate models are estimated with 10 LIs. In the second panel, the models are estimated with the five indicators indicated in Table I. The smallest ratios in each row are emboldened. The number of forecasting errors for each horizon is 11, in contrast to 40 of the previous tables

from the MIDAS relative to the combination of ADL model forecasts if instead all 10 indicators are used. Using end-of-vintage data the MIDAS forecasts are relatively less accurate, in agreement with our earlier findings for the whole forecast period.

Finally, Table VII(B) is based on forecast errors calculated using first-release data. Provided we use real-time vintage data to estimate the models and generate forecasts, the MIDAS model forecasts are again markedly more accurate.

4. CONCLUSIONS

One of the problems with using LIs to forecast output growth is that recessions in the USA over the last few years do not appear to have had common causes, in the sense that different indicators have worked well for some recessions but not others (see, for example, Stock and Watson, 2003). As recessions account for the some of the major fluctuations, this suggests single-indicator models are unlikely to fare well at predicting output growth over a reasonable historical period. An obvious solution is to combine a number of LIs, either in terms of a single forecasting model, or by combination of the forecasts of the individual models.

We consider the MIDAS regression model as a way of combining the information in multiple LIs in order to predict output growth up to 1 year ahead. There are a number of advantages to using

DOI: 10.1002/jae

MIDAS models in this context: we can exploit the fact that standard LIs are available at a monthly frequency, and this information can be used directly (without aggregating it up to the quarterly frequency); that monthly information on the current quarter will sometimes be available; and that MIDAS allows a number of indicators to be included in the forecasting model without many more parameters needing to be estimated. Our findings indicate that these factors do contribute to more accurate forecasts of output growth.

We also found that the use of real-time vintage data on output growth led to more accurate forecasts using LIs than the traditional end-of-sample real-time forecasting approach. The gains to MIDAS over an AR were generally diminished if first-release data were used to assess predictive ability rather than final vintage data. In agreement with much of the literature, forecast combination fares well when forecasts are assessed using standard evaluation measures such as RMSFE, although the multiple-indicator MIDAS performed well relative to forecast combination when the output growth forecasts were evaluated in terms of their ability to predict the direction of change of output growth and when only the period around the 2001 recession was considered.

Our appraisal concludes that using MIDAS to combine indicators in modelling and exploiting current-quarter monthly information delivers the result that LIs do outperform an autoregression.

ACKNOWLEDGEMENTS

We are grateful to the Associate Editor, Steven Durlauf, and to three anonymous referees for helpful comments.

REFERENCES

Bhansali RJ. 2002. Multi-step forecasting. In *A Companion to Economic Forecasting*, Clements MP, Hendry DF (eds). Oxford: Blackwell; 206–221.

Birchenhall CR, Jessen H, Osborn D, Simpson P. 1999. Predicting US business cycle regimes. *Journal of Business and Economic statistics* **17**(3): 313–323.

Clark TE, McCracken MW. 2001. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* **105**: 85–110.

Clark TE, McCracken MW. 2005. Evaluating direct multi-step forecasts. *Econometric Reviews* **24**: 369–404.

Clark TE, McCracken MW. 2007. Tests of equal predictive ability with real-time data. Discussion paper, Economic Research Department, Federal Reserve Bank of Kansas City.

Clements MP. 2005. *Evaluating Econometric Forecasts of Economic and Financial Variables*. Palgrave Texts in Econometrics. Palgrave Macmillan: Basingstoke, UK.

Clements MP, Galvão AB. 2006. Combining predictors and combining information in modelling: forecasting US recession probabilities and output growth. In *Nonlinear Time Series Analysis of Business Cycles*, Milas C, Rothman P, van Dijk D (eds). *Contributions to Economic Analysis Series*, Elsevier: Amsterdam; 55–73.

Clements MP, Galvão AB. 2008. Macroeconomic forecasting with mixed-frequency data: forecasting output growth in the United States. *Journal of Business and Economic Statistics* **26**: 546–554.

Clements MP, Galvão AB, Kim JH. 2008. Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance* **15**: 729–750.

Croushore D, Stark T. 2001. A real-time data set for macroeconomists. *Journal of Econometrics* **105**: 111–130.

Croushore D, Stark T. 2003. A real-time data set for macroeconomists: does the data vintage matter? *Review of Economics and Statistics* **85**: 605–617.

Diebold FX, Lopez JA. 1996. Forecast evaluation and combination. In *Handbook of Statistics*, Vol. 14, Maddala GS, Rao CR (eds). North-Holland: Amsterdam; 241–268.

Diebold FX, Rudebusch GD. 1991. Forecasting output with the composite leading index: a real-time analysis. *Journal of the American Statistical Association* **86**: 603–610.
Galvão AB. 2007. Changes in predictive ability with mixed frequency data. Working Paper No. 595, Queen Mary, University of London.
Ghysels E, Wright J. 2008. Forecasting professional forecasters. *Journal of Business and Economic Statistics* (forthcoming).
Ghysels E, Santa-Clara P, Valkanov R. 2004. The MIDAS touch: mixed data sampling regression models. Mimeo, Chapel Hill, NC.
Ghysels E, Santa-Clara P, Valkanov R. 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* **131**: 59–95.
Ghysels E, Sinko A, Valkanov R. 2007. MIDAS regressions: further results and new directions. *Econometric Reviews* **26**: 53–90.
Giannone D, Reichlin L, Small D. 2008. Nowcasting GDP and inflation: the real-time informational content of macroeconomic data releases. *Journal of Monetary Economics* **55**: 665–676.
Golinelli R, Parigi G. 2005. Short-run Italian GDP forecasting and real-time data. CEPR Discussion Paper No. 5302.
Granger CWJ, Pesaran MH. 2000. Economic and statistical measures of forecast accuracy. *Journal of Forecasting* **19**: 537–560.
Hendry DF, Mizon GE. 1978. Serial correlation as a convenient simplification, not a nuisance: a comment on a study of the demand for money by the Bank of England. *Economic Journal* **88**: 549–563. (Reprinted in Hendry DF. *Econometrics: Alchemy or Science?* Blackwell: Oxford: 1993, and Oxford University Press, 2000).
Kilian L. 1999. Exchange rates and monetary fundamentals: what do we learn from long-horizon regressions. *Journal of Applied Econometrics* **14**: 491–510.
Koenig EF, Dolmas S, Piger J. 2003. The use and abuse of real-time data in economic forecasting. *Review of Economics and Statistics* **85**(3): 618–628.
Marcellino M, Stock JH, Watson MW. 2006. Comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of Econometrics* **135**: 499–526.
McCracken MW. 2004. Asymptotics for out-of-sample tests of causality. Discussion paper, University of Missouri.
Newbold P, Harvey DI. 2002. Forecasting combination and encompassing. In *A Companion to Economic Forecasting*, Clements MP, Hendry DF (eds). Blackwell: Oxford; 268–283.
Orphanides A. 2001. Monetary policy rules based on real-time data. *American Economic Review* **91**(4): 964–985.
Pesaran MH, Timmermann A. 1992. A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics* **10**: 461–465.
Robertson JC, Tallman EW. 1998. Data vintages and measuring forecast model performance. *Federal Reserve Bank of Atlanta Economic Review* Fourth Quarter: 4–20.
Rünstler G, Sédillot F. 2003. Short-term estimates of euro area real GDP by means of monthly data. Working Paper Series, No. 276, European Central Bank.
Schumacher C, Breitung J. 2008. Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting* **24**: 386–398.
Stock JH, Watson MW. 2003. How did leading indicator forecasts perform during the 2001 recession? *Federal Reserve Bank of Richmond, Economic Quarterly* **89**(3): 71–90.
Timmermann A. 2006. Forecast combinations. In *Handbook of Economic Forecasting*. Vol. 1: *Handbook of Economics 24*, Elliott G, Granger C, Timmermann A (eds). Elsevier: Amsterdam; 135–196.
West KD. 1996. Asymptotic inference about predictive ability. *Econometrica* **64**: 1067–1084.
West KD. 2006. Forecasting evaluation. In *Handbook of Economic Forecasting*. Vol. 1: *Handbook of Economics 24*, Elliott G, Granger C, Timmermann A (eds). Elsevier: Amsterdam; 99–134.
Zheng IY, Rossiter JN. 2006. Using monthly indicators to predict quarterly GDP. Working Paper Series, No. 2006–26, Bank of Canada.

DOI: 10.1002/jae