BANCO CENTRAL DE RESERVA DEL PERÚ

Predicting quarterly aggregates with monthly indicators

Diego Winkelried*

* Central Reserve Bank of Peru

DT. N° 2012-023
Serie de Documentos de Trabajo
Working Paper series
Diciembre 2012

Predicting quarterly aggregates with monthly indicators[‡]

Diego Winkelried

December 12, 2012

Abstract

Many important macroeconomic variables measuring the state of the economy are sampled quarterly and with publication lags, although potentially useful predictors are observed at a higher frequency almost in real time. This situation poses the challenge of how to best use the available data to infer the state of the economy. This paper explores the merits of the so-called Mixed Data Sampling (MIDAS) approach that directly exploits the information content of monthly indicators to predict quarterly Peruvian macroeconomic aggregates. To this end, we propose a simple extension, based on the notion of smoothness priors in a distributed lag model, that weakens the restrictions the traditional MIDAS approach imposes on the data to achieve parsimony. We also discuss the workings of an averaging scheme that combines predictions coming from non-nested specifications. It is found that the MIDAS approach is able to timely identify, from monthly information, important signals of the dynamics of the quarterly aggregates. Thus, it can deliver significant gains in prediction accuracy, compared to the performance of competing models that use exclusively quarterly information.

JEL Classification : C22, C53, E27.
Keywords : Mixed-frequency data, MIDAS, model averaging, nowcasting, backcasting.

[*]Macroeconomic Models Department, Central Reserve Bank of Peru (email: diego.winkelried@bcrp.gob.pe, diegowq@cantab.net).

[†]I would like to thank participants at the XXX Encuentro de Economistas and at research seminars at the Central Reserve Bank of Peru for useful suggestions. I am also indebted to César Carrera for valuable comments. The opinions herein are those of the author and do not necessarily reflect those of the Central Reserve Bank of Peru.

1 Introduction

Policymakers routinely face the dilemma of making decisions in an environment where data are not all sampled at the same frequency. Whereas many macroeconomic variables are sampled monthly (e.g., inflation, monetary aggregates and manufacturing production), others – remarkably real aggregates from national accounts (e.g., GDP, investment and consumption) – are available only quarterly. Moreover, it is often the case that due to publication lags and revisions, at a given period the realizations of important variables may be missing. A leading example is GDP growth, a key variable for monitoring the state of the economy. In the Peruvian case, quarterly GDP figures are released usually 45 days after the end of the quarter, though it is not uncommon to experience further delays, whereas in the US and the Eurozone the delay is even greater, up to two quarters (see, respectively, Clements and Galvão, 2008, 2009; Kuzin et al., 2011). Nonetheless, before the release of the quarterly figure, monthly data that contain useful information on the evolution of GDP (sales, imports, business expectations, among others) become increasingly available. The challenge is, thus, how to best use the available data to predict in real time the state of the economy. By "predicting", we mean either *nowcasting* (in a particular calendar month we predict GDP for the current quarter) and *backcasting* (we predict the, yet unavailable, GDP figure for the previous quarter).

The differences in sampling frequencies pose difficulties for time series analysis. On the one hand, the variables that are available at high frequency surely contain valuable information. On the other hand, such data may not be directly used if some of the variables are available at a lower frequency, as standard time series methods involve data sampled at the same interval. Prima facie solutions are to aggregate or interpolate the data so that all variables are expressed at the same frequency. Nonetheless, matching the sampling frequencies implies that potentially useful information might be discarded (upon aggregation) and may rely on (usually strong) untested assumptions (upon interpolation), rendering the relation between the variables difficult to detect. However, the mixed sampling dilemma has attracted considerable attention, and two relevant strands of the literature propose more satisfactory approaches.

The first one is to formulate a state space model that treats the underlying monthly values of variables only available quarterly as latent variables, and use the Kalman filter as a computational device to tackle missing data. State space models consist of a system of two equations, a measurement equation which links observed series to the latent processes, and a state equation which describes the latent processes dynamics (see, for instance, Evans, 2005; Proietti, 2006; Giannone et al., 2008). In order to account for the mixed-frequency nature of the data, low-frequency variables are interpolated according to their stock-flow nature, implying specific time aggregation schemes (see Harvey, 1989, sec. 6.3). Even though the approach is elegant and has many benefits (e.g., it handles nowcasting and backcasting easily), it can be quite involved, as it requires to fully specify a linear dynamic structure, possibly with many parameters.

An alternative, that has proven useful for various forecasting purposes, is the so-called Mixed Data Sampling (MIDAS) approach developed in Ghysels et al. (2004), Ghysels et al. (2007) and Andreou et al. (2010), and thoroughly surveyed in Armesto et al. (2010). Its main advantages are simplicity, as it relies on standard regression methods, and flexibility to deal with nowcasting and backcasting situations. MIDAS can be regarded as a time-series regression that directly accommodates regressand and regressors sampled at different frequencies, and where distributed lag polynominals are used to ensure parsimonious specifications. In contrast to state space methods, the MIDAS setup formulates a single (reduced form) equation that serves as a direct forecast tool. It is worth mentioning that it is difficult to rank the MIDAS and other approaches based purely on theoretical considerations since their relative merits depend heavily on the DGP. Their performances are better assessed in specific economic applications, and Kuzin et al. (2011) and Bai et al. (2012) conclude that in many circumstances of interest the added complexity of the unobservable variables approach does not provide further forecasting gains.

The purpose of this paper is to evaluate the performance of the MIDAS approach in nowcasting and backcasting quarterly Peruvian macroeconomic aggregates using a set of monthly indicators of economic activity. To this end, we develop a version of the MIDAS regression based on priors to a distributed lag model, and evaluate their predictions recursively, in a pseudo-real time way. The theoretical grounds of our modified MIDAS approach are well-known and we only focus on its workings in our particular application. On the other hand, our MIDAS regressions include only the target quarterly variable and a single monthly indicator (and its lags and leads); hence, in order to move to a multivariate setup, the MIDAS predictions are suitably combined to arrive at a representative figure. Our main finding is that the flexibility of the MIDAS approach in accommodating intra-quarter information can lead to important gains in nowcasting and backasting accuracy, mainly because it is able to timely identify, from monthly information, important signals of the dynamics of the quarterly aggregates.

The remainder of the paper is organized as follows. In section 2 we discuss methodological issues. In particular, we follow Foroni et al. (2012) and argue that the usual MIDAS approach, as described for instance in Ghysels et al. (2007), may impose strong restrictions to the data. Thus, we consider much weaker restrictions in the form of smoothness priors (Shiller, 1973). The method, however, involve the comparison of non-nested models, an issue which is dealt with by averaging predictions using information criteria (cf. Kapetanios et al., 2008). Section 3 describes the data used and the design of our evaluation exercise, introduces benchmark models that use exclusively quarterly data, and reports the results. Finally, section 4 gives closing remarks and avenues for further research.

2 Methodology

This section describes the MIDAS regression approach and illustrates its flexibility to deal with various data structures emerging from the different sampling frequencies of regressors and regressand. Then, some refinements that suit our empirical application below are suggested. Finally, we also propose a prediction combination scheme as a useful device to aggregate predictions coming form non-nested models, and to obtain an average prediction from many individual indicators.

2.1 MIDAS and data structure

We refer to the low frequency time unit as *quarters* and the high frequency unit as *months*. This is for expositional convenience only, as the formulation of the problem assumes that each low-frequency period is composed by an arbitrary number m of high-frequency observations. To keep the notation simple, we use a time index t at the higher (monthly) frequency and regard the low frequency (quarterly) data as observable only if t is a multiple of m, and missing otherwise. More precisely, let y_t be a target variable which is observed every m periods; i.e., in a sample containing n quarters, y_t is observable only for $t = m, 2m, 3m, \ldots, nm$. On the other hand, let x_t be a monthly indicator, observable for $t = 1, 2, 3, \ldots, nm$.

The MIDAS regression approach consists in formulating a simple distributed lag model (DLM) that relates y_t with x_t and its lags. To set up ideas, we describe the approach with the simplest model in this class, namely

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots + \beta_q x_{t-q} + \varepsilon_t \qquad \text{for} \qquad t = m, 2m, 3m, \ldots, nm\,, \tag{1}$$

where ε_t is an error term assumed to be serially uncorrelated and homoscedastic. The parameter $q \geq 0$ is the lag length in the dynamic relationship between y_t and x_t. We discuss more sophisticated specifications in section 2.4.

Equation (1) may be thought of as a standard regression model, after suitably accommodating the regressors to account for different sampling frequencies. Hence, it can be straightforwardly written in matrix form as $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ are $n\times 1$ vectors and $\boldsymbol{X}$ is an $n\times(q+1)$ matrix of regressors. Each row in these matrices corresponds to periods t that are multiples of m, and so the coefficients in $\boldsymbol{\beta}$ can be estimated using the skip sample with n available low-frequency observations of the regressand.

Table 1 illustrates the data structure of the regression (1), i.e. the pattern followed by the rows of $\boldsymbol{y}$ and $\boldsymbol{X}$, for $q = 6$. Given the convention that t represents months and that quarterly data are observed every $m = 3$ periods, in each row of the table t corresponds to the last month of each quarter (Mar, Jun, Sep, Dec), $t - 1$ corresponds to the middle month of each quarter (Feb, May, Aug, Nov), $t - 2$ to the first month of each quarter (Jan, Apr, Jul, Oct), $t - 3$ to the last month of the previous quarter (Dec, Mar, Jun, Sep), and so on. As in the case of a DLM with a single sampling frequency, the $(s-1)$-th element of $\boldsymbol{y}$ equals the lag of the s-th element; in contrast, the $(s-1)$-th row of $\boldsymbol{X}$ is not to the usual (monthly) lagged version of the s-th row, but it equals instead the s-th row shifted m periods (i.e., a *quarterly* lag). A mechanical way to arrive to such a data structure is to first write the regression model at the highest frequency, and then to remove the rows corresponding to missing values of y_t (t not a multiple of m).

Often, the purpose of (1) is to predict y_t using information in x_t. Suppose that the s-th row in $\boldsymbol{X}$, $\mathbf{x}_s$, is available, but

y_s is not. This situation arises naturally because we may access a sequence of monthly data until a new quarterly figure becomes available. Furthermore, it is usually the case that y_t is subject to publication lags and revisions whereas x_t is not. In any event, the setup is such that prediction amounts to *direct forecasting*: if $\boldsymbol{b}$ denotes a vector of estimates of $\boldsymbol{\beta}$, then the direct forecast is given by $\hat{y}_s = \mathbf{x}_s\boldsymbol{b}$. In terms of the example in Table 1, the observation of y_t corresponding to the first quarter of year 2 can be predicted with data of x_t from September of year 1 to March of year 2; in this case, the estimate $\boldsymbol{b}$ is obtained using quarterly information up to the fourth quarter of year 1, and monthly information up to December of year 1.

2.2 Smoothness priors

A potential problem with (1) is parameter proliferation. Since the regressors are observed at a higher frequency than the dependent variable, the adequate modeling of the dynamic relationship between y_t and x_t may require the inclusion of many lags (q large, relative to n), which can easily lead to overparametrization if (1) is left unrestricted. For instance, if quarterly observations of y_t are affected by two quarters worth of lagged *daily* x_t, we would need $q = 2 \times 3 \times 20 = 120$ lags at the higher frequency (assuming 20 daily observations per month). The usual MIDAS approach (Ghysels et al., 2004) deals with this issue by specifying the coefficients β_r ($r = 0, 1, 2, \ldots, q$) as flexible smooth functions of r that depend upon few parameters (a nonlinear, deterministic version of equation (2) below), thereby allowing for long lags parsimoniously (see Ghysels et al., 2007, for a detailed discussion). This formulation introduces nonlinearities in the parameters space, and a restricted version of (1) can be finally estimated by nonlinear least squares (see Andreou et al., 2010).

As illustrated, overparameterization is likely to be a major limitation when the difference between sampling frequencies is high (common in financial applications). Nonetheless, Foroni et al. (2012) argue that if the difference between sampling frequencies is narrow (say, quarterly and monthly, common in macroeconomic applications) the nonlinear MIDAS approach may be too restrictive. For instance, if quarterly observations of y_t are affected by two quarters worth of lagged *monthly* x_t, we would need $q = 2 \times 3 = 6$ lags at the higher frequency, and so the unrestricted DLM (1) may suffice for estimation, even when n is moderate.

However, the results in Foroni et al. are based on simulations with $n = 100$, which almost doubles the sample size in our application below. Thus, constraining the estimation of (1) may still be desirable to attenuate high variances due to a small sample size. In particular, we take an intermediate route between the nonlinear MIDAS of Ghysels et al. (which imposes probably too strong restrictions) and the unrestricted MIDAS, based on the notion of *smoothness priors*, first advanced in a classical paper by Shiller (1973) and then refined in Ullah and Raj (1979).

The purpose of imposing smoothness priors is to aid the estimation of (1) by complementing the limited information in a small sample with restrictions that seem a priori reasonable, in order to economize on the effective degrees of freedom used in the estimation. The imposition of such priors almost surely lead to biased estimators, but with lower variances, therefore opening the possibility of achieving gains in terms of mean squared errors (see Rao et al., 2008, ch. 5). Here, each coefficient in the DLM is required to *approximately* lie on an "Almon" polynomial of degree d,

$$\beta_r = \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \ldots + \alpha_p r^d - \omega_r \qquad \text{for} \qquad r = 0, 1, 2, \ldots, q\,, \tag{2}$$

where ω_r is a stochastic term with zero mean, and $\alpha_0, \alpha_1, \ldots, \alpha_d$ are unknown coefficients to be estimated. If $q > d$, there will be fewer parameters α than β, thus (2) imposes $q - d$ restrictions on $\boldsymbol{\beta}$. Note however that these restrictions are stochastic (not expected to hold exactly), which is advantageous since d can be relatively small without imposing overly strong restrictions on the data.

A convenient way to obtain an estimator $\boldsymbol{b}$ subject to (2) is by means of the so-called mixed estimation technique (see Rao et al., 2008, sec. 5.10), that can be regarded as an approximation to a full-fledged Bayesian procedure (as originally proposed by Shiller). Let $\boldsymbol{\alpha}$ be the $(d+1) \times 1$ vector that collects the coefficients $\alpha_0, \alpha_1, \ldots, \alpha_d$. Then, (2) can be written in matrix form as

$$\boldsymbol{\beta} = \boldsymbol{A}\boldsymbol{\alpha} - \boldsymbol{\omega}\,, \tag{3}$$

where $\boldsymbol{A}$ is the $(q+1)\times(d+1)$ Almon transformation matrix, $A_{ij} = (i-1)^{j-1}$ for $i = 1, 2, \ldots, q+1$ and $j = 1, 2, \ldots, d+1$, and $\boldsymbol{\omega}$ is a $(q+1)\times 1$ vector of uncorrelated and homoscedastic disturbances, $\mathrm{var}(\boldsymbol{\omega}) = V_\omega \boldsymbol{I}_{q+1}$. In (3), since $\boldsymbol{A}$ is a $(q+1)\times(d+1)$ matrix of full column rank, one can find a $(q-d)\times(q+1)$ matrix $\boldsymbol{R}$ such that $\boldsymbol{RA} = \boldsymbol{0}$. It is not difficult to verify that $\boldsymbol{R}$ is the full row rank matrix of $(d+1)$ differences of the form

$$\boldsymbol{R} = \begin{bmatrix} C_1 & C_2 & C_3 & \cdots & C_{d+2} & 0 & 0 & \cdots & 0 \\ 0 & C_1 & C_2 & \cdots & C_{d+1} & C_{d+2} & 0 & \cdots & 0 \\ 0 & 0 & C_1 & \cdots & C_d & C_{d+1} & C_{d+2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & C_1 & C_2 & C_3 & \cdots & C_{d+2} \end{bmatrix} \quad \text{where} \quad C_i = (-1)^{i-1}\binom{d+1}{i-1}. \tag{4}$$

Then, it follows from (3) that the stochastic restrictions amount to $\boldsymbol{R\beta} + \boldsymbol{R\omega} = \boldsymbol{0}$ (see Ullah and Raj, 1979).

Let $\lambda = V_\varepsilon / V_\omega$, a parameter that can be interpreted as the amount of information embodied in the stochastic restrictions. The estimation of $\boldsymbol{\beta}$ is based on a stacked model that combines sample with stochastic prior information,

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{R} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{R\omega} \end{bmatrix} \qquad \text{with} \qquad \mathrm{var}\left(\begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{R\omega} \end{bmatrix} \right) = V_\varepsilon \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & (1/\lambda)\boldsymbol{RR}' \end{bmatrix}. \tag{5}$$

and the mixed estimator equals the generalized least squares estimator of $\boldsymbol{\beta}$ in (5),

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{R}'(\boldsymbol{RR}')^{-1}\boldsymbol{R})^{-1}\boldsymbol{X}'\boldsymbol{y}, \tag{6}$$

Given $\boldsymbol{R}$ and λ, this estimator is very easy to compute. Note that if $V_\omega \to \infty$, the restrictions (2) are basically redundant as they effectively require β_r to take almost any real value. As a consequence $\lambda \to 0$ and $\boldsymbol{b}$ approaches the ordinary least squares estimator of the unrestricted model. Thus, the unrestricted MIDAS estimator of Foroni et al. is encompassed by (6). In contrast, when $V_\omega \to 0$, the restrictions (2) become binding, and $\lambda \to \infty$. Simple but tedious algebra shows that in this case $\boldsymbol{b}$ boils down to a constrained least squares estimator subject to the exact restrictions $\boldsymbol{R\beta} = \boldsymbol{0}$ (the so-called Almon distributed lag estimator).

2.3 Multimodel inference and prediction combinations

The estimator $\boldsymbol{b}$ depends on the lag length q, the degree of the Almon polynominal d, and the ratio λ, all quantities that need to be specified to complete the estimation. Nonetheless, a difficulty is that estimators associated with different triplets (q, d, λ) may be *non-nested*. For instance, we cannot contrast in a standard fashion an estimator that uses (q, d, λ) against one that uses $(q+1, d, \lambda)$. Moreover, a practical limitation of mixed estimation is the determination of λ, since there is no entirely satisfactory procedure to do so (in practice, λ is "guesstimated").

A convenient way around these model selection issues is to use of information criteria, in order to allocate a meaningful score to each of many possible specifications in terms of fit and model size (or complexity). Thus, suppose that we have estimated $\mathcal{M}$ different versions of (6); model i ($i = 1, 2, \ldots, \mathcal{M}$) is characterized by a matrix of regressors $\boldsymbol{X}_i$, a matrix of restrictions $\boldsymbol{R}_i$, and a parameter λ_i. In the context of linear parametric dynamic regression models estimated via maximum likelihood (or ordinary least squares), Hurvich and Tsai (1989) propose using the extended Akaike information criterion (AIC),

$$A_i = \log\left(\frac{\boldsymbol{e}_i{}'\boldsymbol{e}_i}{n}\right) + \frac{n + K_i}{n - K_i - 2}, \tag{7}$$

where $\boldsymbol{e}_i = \boldsymbol{y} - \boldsymbol{X}_i\boldsymbol{b}_i$ is the vector of residuals in model i and K_i is the number of regressors. We arrive at expression (7) by adding to the standard AIC a further penalization term that is increasing in K_i: A_i incorporates a second-order bias correction into the standard AIC and is preferable for relatively small sample sizes (for instance, $n/K_i < 40$).

Hurvich et al. (1998) discuss the applicability of (7) for a wider class of methods, including nonparametric and biased estimators such as (6). They propose a refinement to better approximate the "effective number of degrees" in biased estimation, which simply amounts to set K_i equal to the trace of the so-called "hat" matrix,

$$\boldsymbol{H}_i = \boldsymbol{X}_i(\, \boldsymbol{X}_i{}'\boldsymbol{X}_i + \lambda_i \boldsymbol{R}_i{}'(\boldsymbol{R}_i\boldsymbol{R}_i{}')^{-1}\boldsymbol{R}_i\,)^{-1}\boldsymbol{X}_i{}' \quad \text{such that} \quad K_i = \operatorname{tr}(\boldsymbol{H}_i)\,. \tag{8}$$

Note that if $\lambda_i = 0$, then $K_i = q + 1$ is equal to the number of columns (the rank) of $\boldsymbol{X}_i$, as in a standard regression, whereas it can be shown that if $\lambda_i \to \infty$ then $K_i \to \operatorname{rank}(\boldsymbol{X}_i) - \operatorname{rank}(\boldsymbol{R}_i) = d + 1$, which is clearly lower than the rank of $\boldsymbol{X}_i$. Thus, the penalization term (increasing in K_i) reduces in more restricted models, as the complexity of the model is limited by the imposition of a priori restrictions. This effect, of course, is likely to be compensated by an increase in the sum of squared residuals $\boldsymbol{e}_i{}'\boldsymbol{e}_i$ induced by the restrictions, precisely the trade-off dealt with by A_i.

Given $A_1, A_2, \ldots, A_{\mathcal{M}}$, a usual choice is to select the model with the minimum A. However, a large literature points out that model averaging may be a better practice, especially for the purpose of prediction (see, for instance, Timmermann, 2006). One particularly promising weighting scheme relies on the so-called *Akaike weights*

$$w_i = \frac{\exp\left(-\frac{1}{2}A_i\right)}{\sum_{k=1}^{\mathcal{M}} \exp\left(-\frac{1}{2}A_k\right)}\,, \tag{9}$$

such that if model i gives a prediction $\hat{y}_{is}$, improved inference can be based on the combined prediction[1]

$$\tilde{y}_s = \sum_{i=1}^{\mathcal{M}} w_i \hat{y}_{is}\,. \tag{10}$$

It is beyond the scope of this paper to provide a more detailed discussion on the merits Akaike weights. Excellent references are Burnham and Anderson (2002, ch. 4) and, for forecasting, Kapetanios et al. (2008).

Even though the generalization of the MIDAS regression (1) to multiple predictors is straightforward, model averaging can also be used to make the analysis multivariate. Indeed, an additional advantage is that since $\boldsymbol{X}_i$ is allowed to depend on i, we are able to find not only a combined prediction from many specifications of a given indicator x_t, but also (as done below) a combined prediction *among different indicators*.

2.4 Nowcasting and backcasting

The MIDAS approach can be easily adapted to different data structures, in order to form a prediction with all available information at a given period. Suppose that we are interested in predicting the value of y_t, and that its latest available (quarterly) observation is $y_{t-(\kappa+1)m}$. The parameter κ, the *lag in publication*, controls the span between the latest available datum and the prediction target: if $\kappa = 0$, then the last observation is from the previous quarter y_{t-m}, if $\kappa = 1$ it is from the quarter before y_{t-2m}, and so on. Also, consider that the latest available observation of the monthly indicator is x_{t+h}. The parameter h defines the nature of the prediction. If $-m < h \leq 0$, then x_{t+h} and its lags contain information within quarter t before the quarter ends; thus, prediction amounts to *nowcasting*. On the other hand, if $h > 0$, then x_{t+h} and its lags have information beyond quarter t, and prediction amounts to *backcasting*.

With these considerations in mind, the basic DLM (1) can be extended to an autoregressive distributed lag model (ARDLM) of the form

$$y_t = \beta_0 x_{t+h} + \beta_1 x_{t+h-1} + \beta_2 x_{t+h-2} + \ldots + \beta_q x_{t+h-q} + \gamma_1 y_{t-(\kappa+1)m} + \gamma_2 y_{t-(\kappa+2)m} + \ldots + \gamma_p y_{t-(\kappa+p)m} + \varepsilon_t\,, \tag{11}$$

where $q \geq 0$ is the lag length of the monthly indicator, and $p \geq 0$ is the number of (quarterly) autoregressive

[1] It is easy to show that the estimated variance of $\hat{y}_{is} = \mathbf{x}_{is}\boldsymbol{b}_i$ (conditional on the historical information in $\boldsymbol{X}_i$) is equal to $v_i = \hat{V}_0 \mathbf{r}_{is}\mathbf{r}_{is}{}'$, where $\mathbf{r}_{is} = \mathbf{x}_{is}(\, \boldsymbol{X}_i{}'\boldsymbol{X}_i + \lambda_i \boldsymbol{R}_i{}'(\boldsymbol{R}_i\boldsymbol{R}_i{}')^{-1}\boldsymbol{R}_i\,)^{-1}\boldsymbol{X}_i{}'$ and $\hat{V}_0$ an unbiased estimator of the error variance (see footnote 4). On the other hand, Burnham and Anderson (2002, p. 162) propose

$$v_y = \left(\sum_{i=1}^{\mathcal{M}} w_i \sqrt{v_i + (\hat{y}_{is} - \tilde{y}_s)^2} \right)^2$$

as an estimator of the variance of $\tilde{y}_s$ that accounts for both sampling variability and model uncertainty.

terms.[2] Strictly speaking, the coefficients β and γ depend on h and κ, but we leave this dependence implicit to avoid cluttering the notation. Yet, it is important to bear in mind that (11) needs to be reestimated as new statistical information becomes available, which in turns implies different (h, κ) patterns. This model can be also put in matrix format straightforwardly, and so the mixed estimation technique described in section 2.2 and the combination approach of 2.3 apply trivially.

The example in Table 1 corresponds to the case $h = \kappa = p = 0$. Table 2 displays alternative structures for $p = 1$, $q = 6$ and various values of h and κ. In panels (A) to (C) we set $\kappa = 0$ and so, in estimating the model and predicting the first quarter of year 2, we may use quarterly data up to the last quarter of year 1 (y_{t-m}). In (A), monthly information is available up to January, year 2 ($h = -2$), i.e. the first observation of the quarter, for nowcasting. In view of direct forecasting, the coefficients can be estimated with information up to October, year 1, i.e. the first observation of the latest available quarter. In (B) an additional monthly datum becomes available ($h = -1$), and the data structure is adequately updated: nowcasting is now performed with information up to February, year 2, and estimation up to November, year 1 (the second observations of the quarters). Panel (C) illustrates how backcasting can be implemented ($h = 2$): the latest datum corresponds to the second month of the *next* quarter. On the other hand, panels (D) to (F) have $\kappa = 1$ and so quarterly data runs to the third quarter of year 1. Note that the monthly blocks are the same than in cases (A), (B) and (C) for $h = -2, -1, 2$, respectively, but an adjustment is incorporated for the relevant lag of y_t, which now becomes y_{t-2m}.

3 Evaluation exercise

This section presents the results of a recursive exercise aimed at assessing the pseudo-real time performance of various monthly indicators in nowcasting and backcasting important Peruvian macroeconomic aggregates. First, the data and design of the evaluation exercise are described. Second, we present useful benchmarks that use solely quarterly data, in order to evaluate the performance of the MIDAS models in relative terms. Finally, the results are reported. It is found that although performance depends heavily on the indicator used, the combined prediction can achieve important gains in terms of reducing the mean-squared forecast error. Furthermore, the flexibility of the MIDAS regression in incorporating monthly data within a quarter proves helpful to improve upon models using exclusively quarterly information.

3.1 Data and design

The dataset contains quarterly real gross domestic product (GDP), private investment (gross fixed capital formation) and private consumption from 1993Q1 to 2012Q2. The target variables y_t are the year-on-year (4-quarter) growth rates of these aggregates, and consequently run from 1994Q1. The dataset also includes 11 monthly variables until September 2012, classified into 5 groups: Import volumes (total, capital goods, intermediate goods and durable consumption goods), real monetary aggregates (private credit and currency holdings), economic activity (electricity generation, local cement dispatches and non-primary manufacturing), business expectations from surveys, and fiscal (real added-value tax collection). When required, nominal values are deflated using the Consumer Price Index of Metropolitan Lima. The monthly indicators x_t correspond to the year-on-year (12-month) growth rate of the aforementioned variables.[3]

We have chosen this particular set of indicators, besides their economic significance, because of their timing of publication: many of them are available during the first few days of the following month (say, a week) and most are published within two weeks. An exception is non-primary manufacturing that, under special circumstances, may take about 20 days to be released (though in these cases, it is normal to have an accurate preliminary figure at the

[2] Clements and Galvão (2008) impose common factor restrictions in the estimation of the γ coefficients in order to rule out discontinuities (every m periods) in the impulse response function of x_t on y_t. We do not purse this route here (and leave the γ's unrestricted), but reckon it is an interesting avenue for further research.

[3] Most of the data come from the Weekly Report of the Central Reserve Bank of Peru (www.bcrp.gob.pe/statistics.html): currency (Table 1), credit (Table 5), consumer price index (Table 49), imports (Table 57), non-primary manufacturing (Table 63), tax collection (Table 76), and the quarterly aggregates (Table 82). Business expectations are taken from the Central Bank's business confidence survey (www.bcrp.gob.pe/docs/Estadisticas/Encuestas/Series-de-indices.xls). Electricity generation data are from the Interconnected System Operating Committee (www.coes.org.pe) and cement dispatches come from the National Statistical Office (www.inei.gob.pe).

beginning of the month). Thus, for all practical purposes we ignore the intricacies of differing availability within a month, and assume that all the monthly data become available at once, by the middle of the month. Also, it is worth mentioning that the quarterly aggregates are often subject to revisions (even though the monthly indicators are not) and that our dataset is *final* (as of October 2012), not a real-time one. It does not contain vintages of data, so that we cannot discuss the influence of revisions on the relative forecasting accuracy here (see Clements and Galvão, 2008; Kuzin et al., 2011, for interesting applications exploring this issue).

We use a recursive estimation scheme to evaluate the prediction capabilities of the MIDAS regressions for various patterns of missing values at the end of the sample, see equation (11). In particular, we consider three cases for nowcasting $h = \{-2, -1, 0\}$ and three for backcasting $h = \{1, 2, 3\}$, for situations where the previous quarterly observation of the regressand is available ($\kappa = 0$) and when there is a one-quarter publication lag ($\kappa = 1$). The evaluation sample consists on the 42-quarter window running from 2002Q1 to 2012Q2 where for each combination (h, κ) a direct prediction error $\hat{y}_s - y_s$ can be computed for all indicators and for a combined prediction (using Akaike weights) among indicators. We use the square root of the average of $(\hat{y}_s - y_s)^2$ across the evaluation window, i.e. the root mean squared prediction error (RMSPE), as our metric of prediction accuracy.

Regarding the triplet (q, d, λ), for each indicator we consider the values $q = \{3, 6, 9, 12\}$ and $d = \{1, 2, 3, 4\}$ (recall that only models with $q > d$ are valid). Furthermore, we set $\lambda = \hat{V}_0 \times \delta$, where $\hat{V}_0$ is the usual error variance estimate of the unrestricted model[4] and $\delta = \{0, 1, 5, 10, 50, 100, 500, 1000\}$. Note that the case $\delta = 0$ corresponds to the unresticted model, so the choice of d becomes irrelevant. All in all, the combinations of these parameter amount to a total of $\mathcal{M} = 102$ predictors of y_s for each indicator, that are then suitably combined as in (10) to produce a representative prediction.

3.2 Benchmark models

In order to have a better grasp of the information content of the monthly indicators, we compare the performance of the MIDAS regressions to that of prediction models that use only quarterly information.

Our first benchmark is the so-called *Airline model*, ubiquitous in the univariate time-series literature (see, for instance, Harvey, 1989, sec. 2.5), that has proven successful in fitting and forecasting series with the features typically encountered in macroeconomic data, namely stochastic trend and seasonality. In terms of the year-on-year growth rate y_t, the Airline model is an ARIMA(0,1,5) with restrictions on its moving average component,

$$y_t - y_{t-1} = \zeta_t + \theta\zeta_{t-1} + \Theta\zeta_{t-4} + \theta\Theta\zeta_{t-5} \tag{12}$$

where ζ_t is an *iid* innovation. Since (12) uses only information on y_t, the model's prediction in the case $\kappa = 0$ corresponds to its one-step-ahead forecast, whereas the prediction in the case $\kappa = 1$ is its two-step-ahead forecast.

The remaining benchmark models are ARDLM formulated at a quarterly frequency, based on the fact that an alternative solution to the problem of mixed sampling frequencies is simply to convert higher-frequency data to match the sampling rate of the lower frequency data. The simplest method is to compute the average of the observations of x_t that occur between samples of the lower-frequency variable. If such an average is denoted by X_t, then (11) is replaced by

$$y_t = \phi_0 X_{t+h^*} + \phi_1 X_{t+h^*-m} + \phi_2 X_{t+h^*-2m} + \ldots + \phi_{q^*} X_{t+h^*-q^*m} + \gamma_1 y_{t-(\kappa+1)m} + \gamma_2 y_{t-(\kappa+2)m} + \ldots + \gamma_p y_{t-(\kappa+p)m} + \varepsilon_t\,, \tag{13}$$

where q^* is the quarterly lag length and h^* has a similar interpretation than κ (an "averaged" h). Note that since $X_t = (x_t + x_{t-1} + \ldots + x_{t-m+1})/m$, equation (13) turns out to be a constrained version of (11), where averaging imposes restrictions on the shape of the transfer function from x_t to y_t: $\beta_0 = \beta_1 = \cdots = \beta_{m-1} = \phi_0/m$, $\beta_m = \beta_{m+1} = \cdots = \beta_{2m-1} = \phi_1/m$, and so on. This time-averaging model is parsimonious but discards any information about the timing of innovations to higher-frequency data (i.e., it discretizes the transfer function $\beta_0, \beta_1, \ldots, \beta_q$). As with many features of the dynamic models in this paper, whether these restrictions hold is an empirical matter.

[4] That is to say, $\hat{V}_0 = \boldsymbol{y}'\boldsymbol{M}\boldsymbol{y}/\mathrm{tr}(\boldsymbol{M})$ for $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$.

Given κ, the first benchmark model uses $h^* = -m = -3$ in (13), so that the latest available observations from x_t are from quarter $t - m$. This model is to be compared with nowcasting (monthly) models that use $h = -2, -1, 0$ in order to have a sense of the marginal gains coming from the availability of monthly information within quarter t. By the same token, the second benchmark model sets $h^* = 0$, and is to be compared with backcasting models with $h = 1, 2, 3$. Note that, in principle, the quarterly model (13) with $h^* = 0$ uses the same information set as the monthly model (11) with $h = 0$. Thus, comparing both models provides an informal procedure to evaluate the effects of the "averaging restrictions" on the predicting performance of the ARDLM.

3.3 Results

The results for GDP growth are displayed in Table 3. At first glance, some expected patterns arise. First, in most cases the MIDAS regressions are better than their quarterly benchmarks (either with $h^* = -3$ or $h^* = 0$), because they exploit within quarter information. Second, and related to the previous point, the gains in prediction (i.e., the reduction in the RMSPE) are increasing in h. Third, the performance of all models deteriorates as we move form $\kappa = 0$ to $\kappa = 1$, but the univariate model suffers more (recall that this is a two-step-ahead forecast), and consequently further significant relative gains are found for many of the MIDAS regressions.

Be that as it may, we also note that performance varies widely with the indicator and, mainly, with the data structure (h, κ). No single regressor seems to outperform its rivals in all situations, and it is difficult to establish a robust ranking. For instance, the outstanding performance of survey expectations in nowcasting and backcasting GDP growth with $\kappa = 0$, is shadowed by the performance of currency when introducing a publication lag, $\kappa = 1$. Furthermore, the comparison between the nowcasting MIDAS regressions with $h = 0$ and the quarterly ARLDM with $h^* = 0$ is somehow inconclusive, though there is some indication that the former is better, hence the "averaging restrictions" of the ARDLM are not supported by the data. All in all, these results illustrate how the properties of the MIDAS approach are quite dependent on the DGP.

Nonetheless, the combined predictions display a much more stable behavior, along with a remarkable performance. The prediction capacity of the quarterly benchmarks is significantly better than that of the univariate models (with reductions in the RMSPE ranging from 10%-20% for nowcasting to 30%-40% for backcasting), and in turn significantly worse than that of the monthly MIDAS models (with reductions in the RMSPE also from 10% to 40%). It is then apparent that the proposed model averaging scheme is an effective tool to "separate the wheat from the chaff", especially in a setup where individual results are scattered.

The implications of our findings for real-time analysis are important. As mentioned, in the Peruvian case quarterly data are released about 45 days after the end of the quarter, so the data structures that suit this pattern of publication are backcasting with $\kappa = 1$ and $h = 1$ (say, predicting 2nd quarter with data up to July or 3rd quarter with data up to October), nowcasting with $\kappa = 1$ and $h = -2$ (predicting 3rd quarter with data up to July) and nowcasting with $\kappa = 0$ and $h = \{-1, 0\}$ (predicting 3rd quarter with data up to August or September). Thus, in practice the combined prediction can achieve substantial gains in prediction accuracy, that amount to 20% to 30% of the quarterly RMSPE and 30% to 40% of the univariate RMSPE in the cases with $\kappa = 0$, and to around 20% of the quarterly RMSPE and 40% to 50% of the univariate RMSPE in the cases with $\kappa = 1$.

Figure 1 helps understanding the sources of prediction errors across models. During the first part of the evaluation window, especially from 2004 to 2008, the Peruvian economy grew at high and increasing rates. The univariate models show difficulties in tracking the upward trend timely, resembling random walk forecasts. In contrast, MIDAS regressions benefit from monthly information already signalling that a momentum is developing, which is reflected in more accurate predictions. On the other hand, the economy's dynamism was abruptly interrupted as the international financial crisis began (late 2008). A few quarters later, the economy experienced a fast and strong recovery. Such rare events are difficult to forecast with past information only, and so the performance of univariate models is poor during the crisis and its immediate aftermath. Again, by incorporating monthly information of indicators that have already suffered the impacts of the crisis, and that later on bounced to pre-crisis values, the MIDAS approach nowcasts both the sharp decrease and increase timely and closely. Even though the performance of the univariate models seems superior by the very end of the evaluation window, it is apparent that the combined predictions of the MIDAS models prove quite useful for the real time monitoring of the economy.

The above discussion suggests that the RMPSE gains of the MIDAS models are due to bias reduction. Although we do not perform more formal RMSPE decompositions, it is interesting to note that reductions in the prediction error variance may also be important. This effect can be assessed by the width of the prediction intervals in Figure 1 (the shaded areas). For $\kappa = 0$ the variance of the MIDAS predictions appear to be roughly the same as that of the univariate forecasts, but it is visibly smaller for $\kappa = 1$. This is, of course, an additional advantage of using intra-quarter information. It is worth-mentioning that whereas the variance of univariate forecasts includes sampling variability, the variance of the MIDAS predictions incorporates *also* model uncertainty (see footnote 1).

Table 4 and Table 5 display the results for investment and consumption growth, respectively. We reach roughly the same qualitative conclusions as in the case of GDP growth, though the are some important points to highlight. First, investment is considerably more volatile (and hence more difficult to predict) than GDP, and thus many more individual predictions outperform the univariate forecasts. In contrast, consumption is less volatile and so the performance of univariate models is fairer. Consequently, the gains of the combined MIDAS predictions, relative to the univariate models, are generally higher for investment growth and smaller for consumption growth, with GDP growth being an intermediate case. On the other hand, the results on the "averaging restrictions" are mixed, depending – again – on (h, κ): in some cases, simply averaging the higher-frequency data produces no discernible disadvantage, whereas in other cases explicitly modeling the flow of data render clearer benefits. Finally, the finding that the combined prediction represents a useful predicting tool in real time remains robust.

4 Concluding remarks

Policymakers and researchers are often confronted with the problem of mixing data frequencies, in particular when inferring about the state of the economy in real time. We have explored the workings of a modified version of the so-called MIDAS approach to nowcast and backcast important quarterly Peruvian macroeconomic aggregates, subject to publication lags, using monthly indicators of economic activity, available almost in real time.

MIDAS can be regarded as a useful and simple tool to accommodate different sampling frequencies in a single parsimonious regression model. The properties of the MIDAS setup depend on the underlying unknown DGP, and thus there is no theoretical argument to expect the MIDAS regressions *per se* to deliver good predictions. Nevertheless, in our evaluation exercise the MIDAS regressions do capture valuable within-quarter information from the monthly data. Unfortunately, it is also found that the performance of the various monthly indicators vary widely with the pattern of missing values at the end of the sample, rendering robust rankings difficult to establish. To deal with this issue, we combine predictions using Akaike weights, and find that the combined prediction is stable and delivers the expected outcome of more accurate predictions as more information is employed. Thus, it is the complementarity between the flexibility of the MIDAS approach to handle different prediction situations, and a device to sort out underperformers (in our case, the combination of predictions), what provides a valuable prediction tool for real time analysis.

We reckon that a promising extension to our work is to explore different averaging schemes, possibly based on combining information rather than predictions. For instance, common factors may be extracted from the monthly indicators, as done in Evans (2005) or Giannone et al. (2008), and then entered the MIDAS regression (11) directly. On the other hand, Clements and Galvão (2008, 2009) and Kuzin et al. (2011) have documented that, respectively for the US and the Eurozone, MIDAS models perform well when using real-time vintage data (instead of final data) of y_t. Thus, an interesting task for future research would be to inquire whether Clements and Galvão's findings feature in other datasets, as the Peruvian one used in this paper.

References

Andreou E., E. Ghysels and A. Kourtellos (2010), “Regression models with mixed sampling frequencies”, *Journal of Econometrics*, 158(2), 246-261.

Armesto, M. T., K. M. Engemann and M. T. Owyang (2010), “Forecasting with mixed frequencies”, Federal Reserve Bank of St. Louis, *Review*, 92(6), 521-536.

Bai, J., E. Ghysels and J. Wright (2012), “State space models and MIDAS regressions”, *Econometric Reviews* (forthcoming).

Burnham, K. P. and D. R. Anderson (2002), *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*, 2nd edition, New York: Springer-Verlag.

Clements, M. P. and A. B. Galvão (2008), “Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States”, *Journal of Business & Economic Statistics*, 26(4), 546-554.

Clements, M. P. and A. B. Galvão (2009), “Forecasting US output growth using leading indicators: An appraisal using MIDAS models”, *Journal of Applied Econometrics*, 24(7), 1187-1206.

Evans, M. D. D. (2005), “Where are we now? Real-time estimates of the macroeconomy”, *International Journal of Central Banking*, 1(2), 127-175.

Foroni, C., M. Marcellino and C. Schumacher (2012), “U-MIDAS: MIDAS regressions with unrestricted lag polynomials”, CEPR Discussion Papers 8828.

Ghysels, E., P. Santa-Clara and R. Valkanov (2004), “The MIDAS touch: Mixed data sampling regression models”, CIRANO Working Papers 2004s-20.

Ghysels E., A. Sinko and R. Valkanov (2007), “MIDAS regressions: Further results and new directions”, *Econometric Reviews*, 26(1), 53-90.

Giannone, D., L. Reichlin and D. Small (2008), “Nowcasting: The real-time informational content of macroeconomic data”, *Journal of Monetary Economics*, 55(4), 665-676.

Harvey, A. C. (1989), *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge University Press.

Hurvich, C. M. and C. Tsai (1989), “Regression and time series model selection in small samples”, *Biometrika*, 76(2), 297-307.

Hurvich, C. M., J. S. Simonoff and C. L. Tsai (1998), “Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion”, *Journal of the Royal Statistical Society (Series B)*, 60(2), 271-293.

Kapetanios G., V. Labhard and S. Price (2008), “Forecasting using Bayesian and information-theoretic model averaging: An application to U.K. inflation”, *Journal of Business & Economic Statistics*, 26(1), 33-41.

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.

Proietti, T. (2006), “Temporal disaggregation by state space methods: Dynamic regression methods revisited”, *Econometrics Journal*, 9(3), 357-372.

Rao, C. Radhakrishna, H. Toutenburg, Shalabh and C. Heumann (2008), *Linear Models and Generalizations: Least Squares and Alternatives*, Third extended edition, Springer Series in Statistics, Springer-Verlag.

Shiller, R. J. (1973), “A distributed lag estimator derived from smoothness priors”, *Econometrica*, 41(4), 775-788.

Timmermann, A. (2006), “Forecast combinations”, in Elliott, G., C. W. J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Elsevier, vol. 1, ch. 4, 135-196.

Ullah, A. and B. Raj (1979), “A distributed lag estimator derived from Shiller’s smoothness priors: An extension”, *Economics Letters*, 2(3), 219–223.

Table 1. *Data structure in the simple MIDAS regression (1)*

y_t	x_t	x_{t-1}	x_{t-2}	x_{t-3}	x_{t-4}	x_{t-5}	x_{t-6}
1st 1	Mar 1	Feb 1	Jan 1	Dec 0	Nov 0	Oct 0	Sep 0
2nd 1	Jun 1	May 1	Apr 1	Mar 1	Feb 1	Jan 1	Dec 0
3rd 1	Sep 1	Aug 1	Jul 1	Jun 1	May 1	Apr 1	Mar 1
4th 1	Dec 1	Nov 1	Oct 1	Sep 1	Aug 1	Jul 1	Jun 1
1st 2	Mar 2	Feb 2	Jan 2	Dec 1	Nov 1	Oct 1	Sep 1

Note: 1st T, 2nd T, 3rd T and 4th T denote quarters of year T, whereas Jan T, Feb T, . . ., Nov T, Dec T denote months of year T.

Table 2. *Data structure in the extended MIDAS regression (11)*

y_t	$y_{t-(\kappa+1)m}$	x_{t+h}	x_{t+h-1}	x_{t+h-2}	x_{t+h-3}	x_{t+h-4}	x_{t+h-5}	x_{t+h-6}
(A) Nowcast ($h = -2$ and $\kappa = 0$). Last quarterly datum: 4th 1; last monthly datum: Jan 2								
1st 1	4th 0	Jan 1	Dec 0	Nov 0	Oct 0	Sep 0	Aug 0	Jul 0
2nd 1	1st 1	Apr 1	Mar 1	Feb 1	Jan 1	Dec 0	Nov 0	Oct 0
3rd 1	2nd 1	Jul 1	Jun 1	May 1	Apr 1	Mar 1	Feb 1	Jan 1
4th 1	3rd 1	Oct 1	Sep 1	Aug 1	Jul 1	Jun 1	May 1	Apr 1
1st 2	4th 1	Jan 2	Dec 1	Nov 1	Oct 1	Sep 1	Aug 1	Jul 1
(B) Nowcast ($h = -1$ and $\kappa = 0$). Last quarterly datum: 4th 1; last monthly datum: Feb 2								
1st 1	4th 0	Feb 1	Jan 1	Dec 0	Nov 0	Oct 0	Sep 0	Aug 0
2nd 1	1st 1	May 1	Apr 1	Mar 1	Feb 1	Jan 1	Dec 0	Nov 0
3rd 1	2nd 1	Aug 1	Jul 1	Jun 1	May 1	Apr 1	Mar 1	Feb 1
4th 1	3rd 1	Nov 1	Oct 1	Sep 1	Aug 1	Jul 1	Jun 1	May 1
1st 2	4th 1	Feb 2	Jan 2	Dec 1	Nov 1	Oct 1	Sep 1	Aug 1
(C) Backcast ($h = 2$ and $\kappa = 0$). Last quarterly datum: 4th 1; last monthly datum: May 2								
1st 1	4th 0	May 1	Apr 1	Mar 1	Feb 1	Jan 1	Dec 0	Nov 0
2nd 1	1st 1	Aug 1	Jul 1	Jun 1	May 1	Apr 1	Mar 1	Feb 1
3rd 1	2nd 1	Nov 1	Oct 1	Sep 1	Aug 1	Jul 1	Jun 1	May 1
4th 1	3rd 1	Feb 2	Jan 2	Dec 1	Nov 1	Oct 1	Sep 1	Aug 1
1st 2	4th 1	May 2	Apr 2	Mar 2	Feb 2	Jan 2	Dec 1	Nov 1
(D) Nowcast ($h = -2$ and $\kappa = 1$). Last quarterly datum: 3rd 1; last monthly datum: Jan 2								
4th 0	2nd 0	Jan 1	Dec 0	Nov 0	Oct 0	Sep 0	Aug 0	Jul 0
1st 1	3rd 0	Apr 1	Mar 1	Feb 1	Jan 1	Dec 0	Nov 0	Oct 0
2nd 1	4th 0	Jul 1	Jun 1	May 1	Apr 1	Mar 1	Feb 1	Jan 1
3rd 1	1st 1	Oct 1	Sep 1	Aug 1	Jul 1	Jun 1	May 1	Apr 1
1st 2	3rd 1	Jan 2	Dec 1	Nov 1	Oct 1	Sep 1	Aug 1	Jul 1
(E) Nowcast ($h = -1$ and $\kappa = 1$). Last quarterly datum: 3rd 1; last monthly datum: Feb 2								
4th 0	2nd 0	Feb 1	Jan 1	Dec 0	Nov 0	Oct 0	Sep 0	Aug 0
1st 1	3rd 0	May 1	Apr 1	Mar 1	Feb 1	Jan 1	Dec 0	Nov 0
2nd 1	4th 0	Aug 1	Jul 1	Jun 1	May 1	Apr 1	Mar 1	Feb 1
3rd 1	1st 1	Nov 1	Oct 1	Sep 1	Aug 1	Jul 1	Jun 1	May 1
1st 2	3rd 1	Feb 2	Jan 2	Dec 1	Nov 1	Oct 1	Sep 1	Aug 1
(F) Backcast ($h = 2$ and $\kappa = 1$). Last quarterly datum: 3rd 1; last monthly datum: May 2								
4th 0	2nd 0	May 1	Apr 1	Mar 1	Feb 1	Jan 1	Dec 0	Nov 0
1st 1	3rd 0	Aug 1	Jul 1	Jun 1	May 1	Apr 1	Mar 1	Feb 1
2nd 1	4th 0	Nov 1	Oct 1	Sep 1	Aug 1	Jul 1	Jun 1	May 1
3rd 1	1st 1	Feb 2	Jan 2	Dec 1	Nov 1	Oct 1	Sep 1	Aug 1
1st 2	3rd 1	May 2	Apr 2	Mar 2	Feb 2	Jan 2	Dec 1	Nov 1

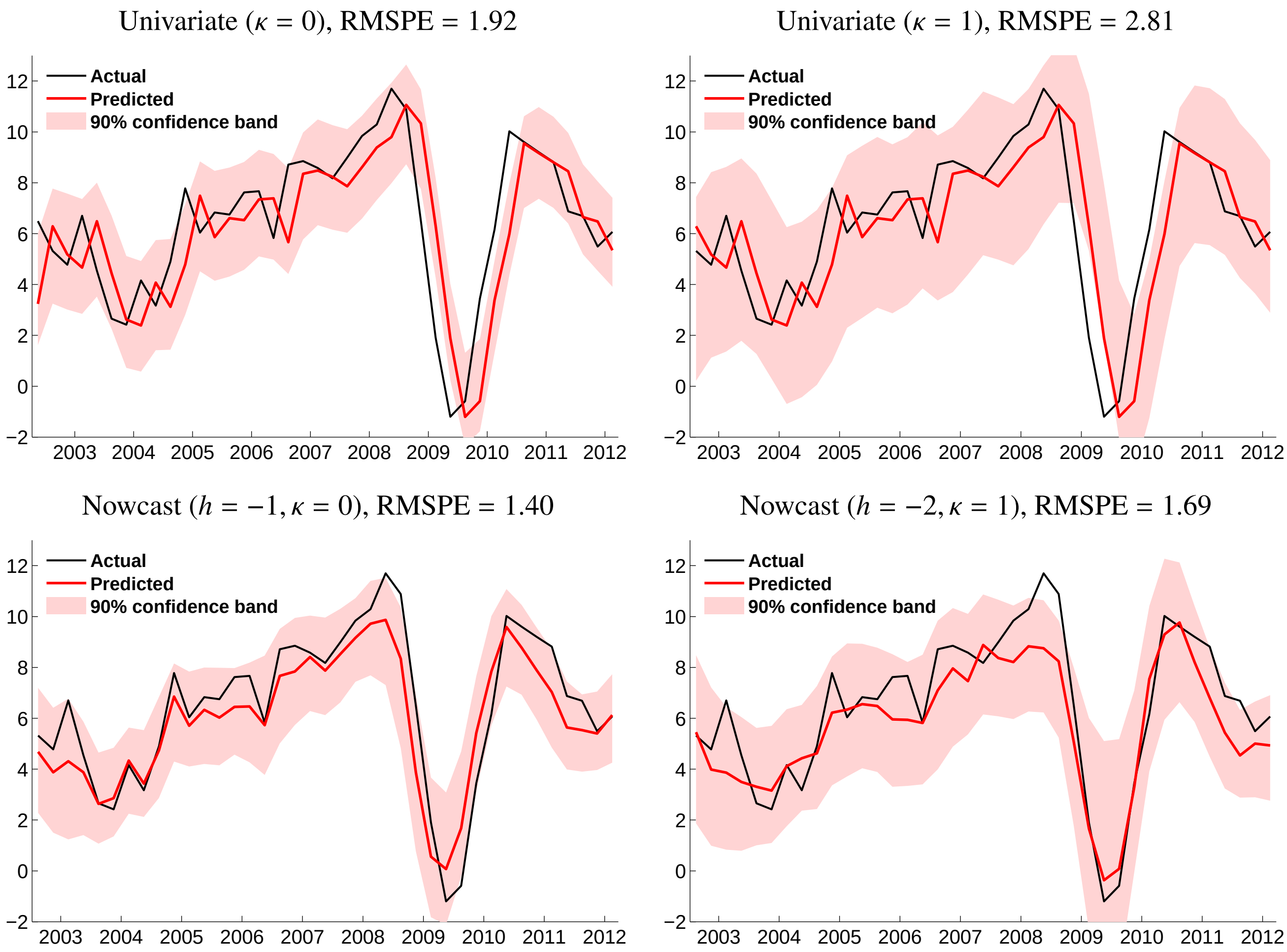
Note: 1st T, 2nd T, 3rd T and 4th T denote quarters of year T, whereas Jan T, Feb T, . . ., Nov T, Dec T denote months of year T.

Table 3. *Root mean squared prediction errors for GDP growth (2002 - 2012)*

	Nowcasting				*Backcasting*			
	Quarterly	*Monthly*			*Quarterly*	*Monthly*		
	$h^* = -3$	$h = -2$	$h = -1$	$h = 0$	$h^* = 0$	$h = 1$	$h = 2$	$h = 3$
				No lag in publication ($\kappa = 0$)				
Univariate	1.92							
Imports, total	**1.76 {0.92}**	1.86 [1.04]	**1.75 [0.98]**	**1.67 [0.93]**	**1.44 {0.75}**	**1.39 [0.93]**	**1.47 [0.99]**	**1.25 [0.84]**
Imports, capital	2.06 {1.07}	2.05 [0.99]	2.02 [0.97]	1.96 [0.96]	**1.85 {0.96}**	**1.82 [0.98]**	**1.62 [0.89]**	**1.41 [0.76]**
Imports, durable consumer	1.94 {1.01}	2.04 [1.05]	2.08 [1.06]	1.99 [1.01]	**1.79 {0.93}**	1.85 [1.01]	**1.88 [1.00]**	**1.74 [0.92]**
Imports, intermediate	2.18 {1.14}	2.33 [1.04]	2.11 [0.94]	2.05 [0.90]	1.92 {1.00}	2.09 [1.02]	2.15 [1.04]	**2.02 [0.98]**
Credit	**1.91 {0.99}**	**1.85 [0.98]**	**1.85 [0.98]**	**1.74 [0.92]**	**1.79 {0.93}**	**1.59 [0.90]**	**1.59 [0.89]**	**1.58 [0.90]**
Currency	**1.78 {0.93}**	1.79 [1.00]	**1.75 [0.99]**	**1.57 [0.90]**	**1.70 {0.89}**	**1.54 [0.90]**	**1.54 [0.90]**	**1.56 [0.91]**
Electricity	**1.90 {0.99}**	2.02 [1.04]	1.96 [1.02]	**1.91 [0.99]**	**1.86 {0.97}**	2.01 [1.08]	2.05 [1.10]	1.99 [1.07]
Cement	2.02 {1.05}	2.02 [0.95]	**1.70 [0.81]**	**1.76 [0.85]**	**1.77 {0.92}**	**1.71 [0.97]**	**1.70 [0.96]**	**1.77 [0.99]**
Non-primary manufacturing	1.98 {1.03}	**1.68 [0.81]**	**1.36 [0.67]**	**1.24 [0.61]**	**1.31 {0.69}**	1.31 [1.01]	1.33 [1.01]	**1.29 [1.00]**
Survey expectations	**1.64 {0.86}**	**1.39 [0.85]**	**1.40 [0.86]**	**1.36 [0.81]**	**1.41 {0.74}**	**1.35 [0.95]**	**1.38 [0.99]**	**1.34 [0.94]**
Added-value tax collection	2.19 {1.14}	2.17 [0.97]	2.23 [0.99]	2.38 [1.07]	2.30 {1.20}	2.31 [0.99]	2.28 [0.98]	2.31 [0.98]
Combined	**1.71 {0.89}**	**1.60 [0.94]**	**1.40 [0.82]**	**1.29 [0.76]**	**1.35 {0.70}**	**1.18 [0.87]**	**1.20 [0.89]**	**1.14 [0.84]**
				Lag in publication ($\kappa = 1$)				
Univariate	2.81							
Imports, total	**2.29 {0.81}**	**2.16 [0.92]**	**2.07 [0.87]**	**1.68 [0.72]**	**1.63 {0.58}**	**1.37 [0.82]**	**1.35 [0.84]**	**1.39 [0.85]**
Imports, capital	3.03 {1.08}	2.97 [0.98]	**2.64 [0.91]**	**2.29 [0.77]**	**2.32 {0.83}**	**2.21 [0.96]**	**1.96 [0.88]**	**1.82 [0.82]**
Imports, durable consumer	2.87 {1.02}	**2.56 [0.86]**	**2.67 [0.90]**	**2.45 [0.81]**	**2.36 {0.84}**	**2.20 [0.90]**	**2.19 [0.88]**	**2.08 [0.82]**
Imports, intermediate	3.21 {1.14}	**2.70 [0.84]**	**2.66 [0.82]**	**2.44 [0.77]**	2.94 {1.05}	**2.52 [0.85]**	**2.55 [0.85]**	**2.52 [0.85]**
Credit	**2.65 {0.95}**	**2.30 [0.87]**	**2.26 [0.86]**	**2.25 [0.85]**	**2.24 {0.80}**	**2.17 [0.96]**	**2.12 [0.95]**	**2.07 [0.93]**
Currency	**2.48 {0.88}**	**2.15 [0.86]**	**2.13 [0.86]**	**2.09 [0.83]**	**2.22 {0.79}**	**2.05 [0.92]**	**2.06 [0.92]**	**2.06 [0.92]**
Electricity	3.11 {1.11}	2.94 [0.96]	2.99 [0.96]	2.91 [0.94]	3.10 {1.10}	**2.69 [0.87]**	**2.81 [0.90]**	2.88 [0.92]
Cement	**2.77 {0.99}**	**2.16 [0.78]**	**2.10 [0.76]**	**2.14 [0.76]**	**2.52 {0.90}**	**2.02 [0.80]**	**2.02 [0.80]**	**2.10 [0.83]**
Non-primary manufacturing	**2.60 {0.93}**	**2.23 [0.87]**	**2.24 [0.86]**	**2.09 [0.82]**	**2.03 {0.72}**	2.01 [1.04]	2.03 [1.04]	**1.90 [1.00]**
Survey expectations	**2.32 {0.83}**	2.39 [1.03]	2.33 [1.01]	2.37 [1.03]	**2.15 {0.77}**	2.33 [1.09]	2.58 [1.19]	2.64 [1.22]
Added-value tax collection	3.53 {1.26}	3.12 [0.87]	3.14 [0.89]	3.15 [0.89]	3.42 {1.22}	2.86 [0.84]	2.91 [0.87]	2.82 [0.85]
Combined	**2.27 {0.81}**	**1.69 [0.75]**	**1.64 [0.72]**	**1.48 [0.65]**	**1.65 {0.59}**	**1.36 [0.82]**	**1.35 [0.82]**	**1.37 [0.83]**

Notes: Figures in curly braces {.} are RMSPE relative to the univariate models, whereas figures in squared braces [.] are relative to the quarterly ARDLM ($h^* = -3$ or $h^* = 0$). Figures in **boldface** correspond to quarterly ARDLMs with better performance than the univariate models, and monthly models with better performance than the univariate models *and* their quarterly benchmarks. The evaluation sample runs from 2002Q1 to 2012Q2 (42 quarters) for all indicators but "Survey expectations" where it runs from 2007Q2 to 2012Q2 (20 quarters).

Figure 1. *Actual and predicted GDP growth (2002 - 2012)*



Notes: Univariate predictions are the one-step-ahead ($\kappa = 0$) and two-step-ahead ($\kappa = 1$) forecasts of the Airline model (12). Nowcasts are the combined prediction of the MIDAS regressions described in the main text. Confidence intervals for the combined predictions are constructed using the variance estimator described in footnote 1, which accounts for model uncertainty.

Table 4. *Root mean squared prediction errors for private investment growth (2002 - 2012)*

	Nowcasting				*Backcasting*			
	Quarterly	*Monthly*			*Quarterly*	*Monthly*		
	$h^* = -3$	$h = -2$	$h = -1$	$h = 0$	$h^* = 0$	$h = 1$	$h = 2$	$h = 3$
	No lag in publication ($\kappa = 0$)							
Univariate	8.40							
Imports, total	**7.28 {0.87}**	**6.17 [0.84]**	**5.32 [0.72]**	**4.23 [0.58]**	**4.17 {0.50}**	4.29 [1.05]	4.36 [1.07]	4.83 [1.17]
Imports, capital	**7.75 {0.92}**	**6.77 [0.88]**	**5.45 [0.70]**	**3.91 [0.51]**	**4.12 {0.49}**	**3.90 [0.96]**	**3.78 [0.93]**	**3.59 [0.89]**
Imports, durable consumer	**7.32 {0.87}**	**7.28 [0.98]**	**7.32 [0.99]**	**6.40 [0.87]**	**6.48 {0.77}**	**6.18 [0.95]**	**6.41 [0.99]**	**6.33 [0.96]**
Imports, intermediate	**7.78 {0.93}**	7.70 [1.00]	**7.20 [0.92]**	**6.54 [0.86]**	**6.66 {0.79}**	6.84 [1.03]	7.15 [1.07]	7.08 [1.07]
Credit	**7.95 {0.95}**	**7.55 [0.96]**	**7.43 [0.94]**	**7.27 [0.92]**	**7.29 {0.87}**	**6.99 [0.95]**	**6.90 [0.94]**	**6.80 [0.93]**
Currency	**7.29 {0.87}**	**6.99 [0.95]**	**6.85 [0.95]**	**6.76 [0.94]**	**6.73 {0.80}**	6.71 [1.00]	6.76 [1.01]	6.75 [1.01]
Electricity	**7.56 {0.90}**	**7.40 [0.98]**	**7.54 [0.99]**	**7.43 [0.98]**	**7.45 {0.89}**	**7.51 [1.00]**	7.77 [1.03]	7.74 [1.03]
Cement	**7.70 {0.92}**	**6.95 [0.90]**	**6.95 [0.90]**	**6.80 [0.89]**	**6.74 {0.80}**	7.13 [1.05]	7.17 [1.06]	7.13 [1.06]
Non-primary manufacturing	**6.57 {0.78}**	**6.19 [0.93]**	**5.99 [0.91]**	**5.39 [0.81]**	**5.30 {0.63}**	**5.22 [0.98]**	5.36 [1.01]	5.39 [1.01]
Survey expectations	**7.29 {0.87}**	7.73 [1.04]	7.77 [1.06]	7.62 [1.04]	**7.48 {0.89}**	7.67 [1.03]	8.23 [1.10]	8.35 [1.11]
Added-value tax collection	**8.32 {0.99}**	**8.12 [0.96]**	**8.22 [0.98]**	**8.24 [0.98]**	**7.94 {0.94}**	**7.46 [0.93]**	**7.61 [0.96]**	**7.46 [0.94]**
Combined	**7.03 {0.84}**	**6.05 [0.86]**	**5.23 [0.74]**	**3.74 [0.53]**	**3.94 {0.47}**	**3.68 [0.93]**	**3.78 [0.96]**	**3.83 [0.97]**
	Lag in publication ($\kappa = 1$)							
Univariate	11.98							
Imports, total	**8.25 {0.69}**	**7.63 [0.92]**	**6.92 [0.84]**	**6.11 [0.75]**	**5.17 {0.43}**	6.13 [1.21]	5.93 [1.18]	5.99 [1.21]
Imports, capital	**10.33 {0.86}**	11.29 [1.09]	10.75 [1.06]	**9.70 [0.96]**	**5.17 {0.43}**	9.78 [1.94]	8.88 [1.91]	8.62 [1.88]
Imports, durable consumer	**10.64 {0.89}**	**9.21 [0.87]**	**9.32 [0.87]**	**8.86 [0.82]**	**8.29 {0.69}**	8.47 [1.01]	8.65 [1.03]	8.63 [1.04]
Imports, intermediate	12.72 {1.06}	**9.62 [0.76]**	**9.10 [0.72]**	**8.63 [0.68]**	**11.21 {0.94}**	**9.47 [0.83]**	**9.52 [0.84]**	**9.61 [0.85]**
Credit	**11.06 {0.92}**	**9.99 [0.91]**	**9.87 [0.90]**	**9.74 [0.88]**	**9.36 {0.78}**	9.60 [1.02]	9.59 [1.02]	9.52 [1.01]
Currency	**10.22 {0.85}**	**9.70 [0.96]**	**9.66 [0.95]**	**9.67 [0.95]**	**9.54 {0.80}**	9.80 [1.04]	9.80 [1.03]	9.82 [1.03]
Electricity	**11.69 {0.98}**	11.70 [1.00]	11.84 [1.01]	**11.70 [1.00]**	**11.58 {0.97}**	**11.25 [0.97]**	**11.50 [0.99]**	**11.42 [0.99]**
Cement	**10.15 {0.85}**	**9.09 [0.89]**	**9.05 [0.89]**	**9.06 [0.89]**	**9.66 {0.81}**	**8.99 [0.92]**	**8.99 [0.92]**	**8.99 [0.92]**
Non-primary manufacturing	**7.81 {0.65}**	9.51 [1.22]	9.46 [1.23]	9.03 [1.18]	**6.32 {0.53}**	9.12 [1.47]	9.09 [1.49]	8.97 [1.49]
Survey expectations	**11.10 {0.93}**	15.03 [1.36]	15.55 [1.40]	15.45 [1.38]	**11.50 {0.96}**	15.38 [1.35]	15.88 [1.38]	15.77 [1.37]
Added-value tax collection	12.50 {1.04}	**10.76 [0.86]**	**10.92 [0.88]**	**10.85 [0.87]**	**11.36 {0.95}**	**10.38 [0.91]**	**10.64 [0.94]**	**10.57 [0.93]**
Combined	**9.16 {0.76}**	**7.68 [0.84]**	**7.16 [0.78]**	**6.46 [0.71]**	**5.40 {0.45}**	6.43 [1.19]	6.36 [1.18]	6.40 [1.19]

Note: See notes to Table 3.

Table 5. *Root mean squared prediction errors for private consumption growth (2002 - 2012)*

	Nowcasting				*Backcasting*			
	Quarterly	*Monthly*			*Quarterly*	*Monthly*		
	$h^* = -3$	$h = -2$	$h = -1$	$h = 0$	$h^* = 0$	$h = 1$	$h = 2$	$h = 3$
	No lag in publication ($\kappa = 0$)							
Univariate	1.10							
Imports, total	**1.04** {**0.95**}	1.18 [1.13]	1.14 [1.09]	1.11 [1.11]	**0.88** {**0.80**}	1.07 [1.24]	1.09 [1.26]	1.02 [1.16]
Imports, capital	1.26 {1.15}	1.20 [0.92]	**1.10** [**0.85**]	**1.07** [**0.82**]	**1.05** {**0.96**}	**1.01** [**0.96**]	**0.97** [**0.89**]	**0.93** [**0.88**]
Imports, durable consumer	**1.09** {**0.99**}	**0.99** [**0.88**]	**0.98** [**0.86**]	**0.93** [**0.82**]	**0.89** {**0.81**}	0.89 [1.01]	1.04 [1.16]	0.99 [1.10]
Imports, intermediate	1.52 {1.38}	1.66 [1.06]	1.52 [0.98]	1.49 [0.94]	1.39 {1.27}	1.46 [1.02]	1.41 [0.97]	1.37 [0.95]
Credit	1.11 {1.01}	**1.09** [**0.98**]	1.10 [1.00]	**1.07** [**0.96**]	**1.06** {**0.96**}	**1.02** [**0.96**]	**1.02** [**0.96**]	**1.02** [**0.96**]
Currency	**0.96** {**0.88**}	0.98 [1.03]	0.97 [1.02]	**0.93** [**0.97**]	**0.92** {**0.84**}	0.93 [1.00]	**0.93** [**1.00**]	**0.93** [**0.99**]
Electricity	1.41 {1.28}	1.39 [0.98]	1.37 [0.97]	1.35 [0.95]	1.36 {1.24}	1.28 [0.93]	1.27 [0.92]	1.46 [1.05]
Cement	1.26 {1.14}	1.23 [0.98]	1.15 [0.92]	1.18 [0.95]	1.18 {1.08}	**1.11** [**0.92**]	**1.07** [**0.90**]	**1.09** [**0.90**]
Non-primary manufacturing	1.32 {1.20}	1.31 [0.94]	1.24 [0.89]	1.26 [0.90]	1.16 {1.05}	1.21 [1.00]	1.17 [0.98]	1.16 [0.96]
Survey expectations	1.21 {1.10}	1.23 [1.03]	1.20 [1.01]	1.18 [1.01]	1.15 {1.04}	1.13 [1.09]	1.11 [1.03]	1.13 [1.03]
Added-value tax collection	1.41 {1.29}	1.34 [0.88]	1.35 [0.91]	1.34 [0.92]	1.32 {1.20}	**1.09** [**0.80**]	**1.09** [**0.82**]	**1.09** [**0.82**]
Combined	**1.02** {**0.93**}	**0.87** [**0.85**]	**0.82** [**0.80**]	**0.78** [**0.77**]	**0.80** {**0.73**}	**0.76** [**0.95**]	**0.75** [**0.94**]	**0.75** [**0.93**]
	Lag in publication ($\kappa = 1$)							
Univariate	1.66							
Imports, total	**1.55** {**0.93**}	1.60 [1.03]	1.55 [1.01]	**1.41** [**0.89**]	**1.35** {**0.81**}	**1.38** [**0.97**]	1.36 [1.03]	1.45 [1.04]
Imports, capital	2.01 {1.21}	2.04 [1.01]	1.85 [0.90]	1.75 [0.83]	**1.62** {**0.97**}	1.74 [1.02]	**1.54** [**0.91**]	**1.60** [**0.93**]
Imports, durable consumer	1.79 {1.07}	1.72 [0.95]	1.84 [1.02]	**1.64** [**0.90**]	**1.38** {**0.83**}	1.56 [1.11]	1.61 [1.13]	1.51 [1.05]
Imports, intermediate	2.69 {1.61}	1.84 [0.67]	1.80 [0.65]	**1.66** [**0.61**]	2.45 {1.47}	1.79 [0.72]	1.90 [0.76]	1.89 [0.77]
Credit	1.67 {1.00}	**1.54** [**0.93**]	**1.51** [**0.92**]	**1.50** [**0.91**]	**1.50** {**0.90**}	**1.48** [**0.99**]	**1.45** [**0.98**]	**1.43** [**0.97**]
Currency	**1.47** {**0.88**}	**1.46** [**0.98**]	**1.44** [**0.97**]	**1.46** [**0.98**]	**1.42** {**0.85**}	1.45 [1.02]	**1.41** [**0.99**]	**1.39** [**0.99**]
Electricity	2.27 {1.37}	1.99 [0.87]	1.99 [0.87]	2.06 [0.88]	2.15 {1.29}	1.92 [0.87]	1.96 [0.87]	1.96 [0.87]
Cement	2.03 {1.22}	**1.35** [**0.67**]	**1.31** [**0.65**]	**1.34** [**0.67**]	1.78 {1.07}	**1.22** [**0.70**]	**1.35** [**0.76**]	**1.37** [**0.77**]
Non-primary manufacturing	2.07 {1.24}	**1.70** [**0.80**]	**1.65** [**0.78**]	**1.57** [**0.74**]	1.74 {1.05}	**1.56** [**0.91**]	**1.60** [**0.94**]	**1.47** [**0.88**]
Survey expectations	2.28 {1.37}	1.84 [0.82]	1.81 [0.83]	1.81 [0.84]	2.33 {1.40}	1.82 [0.83]	1.88 [0.88]	1.97 [0.90]
Added-value tax collection	2.35 {1.41}	1.88 [0.80]	1.87 [0.82]	1.86 [0.79]	2.06 {1.24}	1.73 [0.82]	1.80 [0.87]	1.73 [0.82]
Combined	**1.50** {**0.90**}	**1.17** [**0.78**]	**1.13** [**0.75**]	**1.06** [**0.71**]	**1.12** {**0.67**}	**1.02** [**0.91**]	**1.05** [**0.93**]	**1.04** [**0.93**]

Note: See notes to Table 3.